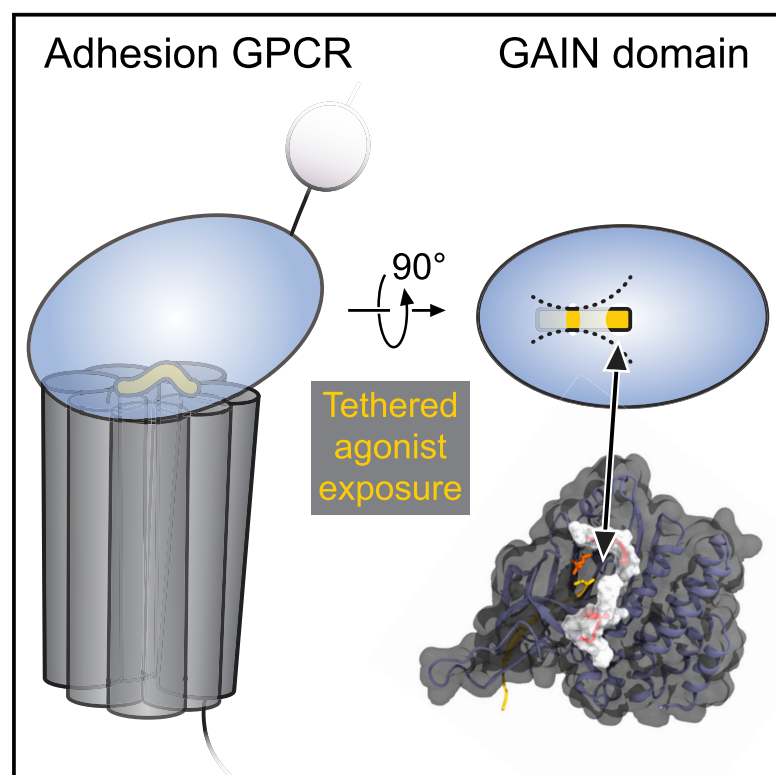# Tethered agonist exposure in intact adhesion/class B2 GPCRs through intrinsic structural flexibility of the GAIN domain

Graphical Abstract

Authors

Gerti Beliu, Steffen Altrichter, Ramon Guixà-González, ..., Peter W. Hildebrand, Markus Sauer, Tobias Langenhan

Correspondence

peter.hildebrand@medizin.uni-leipzig.de (P.W.H.),
m.sauer@uni-wuerzburg.de (M.S.),
tobias.langenhan@gmail.com (T.L.)

In Brief

Beliu et al. demonstrate how adhesion G protein-coupled receptors expose their hidden intramolecular trigger, the tethered agonist/Stachel, without breaking apart.

Highlights

- FRET imaging demonstrates intact adhesion GPCR heterodimers at the cell surface

- MD simulations delineate flexible regions in GAIN domains and TA exposure

- Bioorthogonal click labeling gauges tethered agonist accessibility in adhesion GPCRs

Article

Tethered agonist exposure in intact adhesion/class B2 GPCRs through intrinsic structural flexibility of the GAIN domain

Gerti Beliu,[1,8] Steffen Altrichter,[2,8] Ramon Guixà-González,[3,4,5,8] Mareike Hemberger,[2] Ina Brauer,[2] Anne-Kristin Dahse,[2] Nicole Scholz,[2] Robert Wieduwild,[2] Alexander Kuhlemann,[1] Hossein Batebi,[3] Florian Seufert,[3] Guillermo Pérez-Hernández,[6] Peter W. Hildebrand,[3,6,7,*] Markus Sauer,[1,*] and Tobias Langenhan[2,9,*]

[1]Department of Biotechnology and Biophysics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany
[2]Rudolf Schönheimer Institute of Biochemistry, Division of General Biochemistry, Medical Faculty, Leipzig University, Johannisallee 30, 04103 Leipzig, Germany
[3]Institute for Medical Physics and Biophysics, Medical Faculty, Leipzig University, Härtelstrasse 16–18, 04107 Leipzig, Germany
[4]Laboratory of Biomolecular Research, Paul Scherrer Institute (PSI), 5232 Villigen PSI, Switzerland
[5]Condensed Matter Theory Group, PSI, 5232 Villigen PSI, Switzerland
[6]Institute of Medical Physics and Biophysics, Charité - Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany
[7]Berlin Institute of Health, 10178 Berlin, Germany
[8]These authors contributed equally
[9]Lead contact
*Correspondence: peter.hildebrand@medizin.uni-leipzig.de (P.W.H.), m.sauer@uni-wuerzburg.de (M.S.), tobias.langenhan@gmail.com (T.L.)
https://doi.org/10.1016/j.molcel.2020.12.042

SUMMARY

Adhesion G protein-coupled receptors (aGPCRs)/family B2 GPCRs execute critical tasks during development and the operation of organs, and their genetic lesions are associated with human disorders, including cancers. Exceptional structural aGPCR features are the presence of a tethered agonist (TA) concealed within a GPCR autoproteolysis-inducing (GAIN) domain and their non-covalent heteromeric two-subunit layout. How the TA is poised for activation while maintaining this delicate receptor architecture is central to conflicting signaling paradigms that either involve or exclude aGPCR heterodimer separation. We investigated this matter in five mammalian aGPCR homologs (ADGRB3, ADGRE2, ADGRE5, ADGRG1, and ADGRL1) and demonstrate that intact aGPCR heterodimers exist at the cell surface, that the core TA region becomes unmasked in the cleaved GAIN domain, and that intra-GAIN domain movements regulate the level of tethered agonist exposure, thereby likely controlling aGPCR activity. Collectively, these findings delineate a unifying mechanism for TA-dependent signaling of intact aGPCRs.

INTRODUCTION

Adhesion-type G protein-coupled receptors (aGPCRs) form a large group within the GPCR superfamily, which exerts profound roles in the development and homeostatic control of organ functions (Hamann et al., 2015). aGPCRs are increasingly recognized as target molecules for immune and cancer therapies (Aust et al., 2016; Purcell and Hall, 2018; Scholz, 2018) because genetic lesions of their encoding genes can cause immune defects (Boyden et al., 2016) and belong to the most frequently mutated loci in many cancer types (Kan et al., 2010; O'Hayre et al., 2013). This highlights the need to better understand the molecular events underlying the signal transduction of aGPCRs.

aGPCRs possess molecular features setting them apart from other GPCRs. An aGPCR-specific juxtamembrane GPCR autoproteolysis-inducing (GAIN) domain processes the nascent receptors upon biosynthesis through auto-proteolytic cleavage

at the GPCR proteolysis site (GPS) generating an N- (NTF) and C-terminal fragment (CTF) (Araç et al., 2012; Lin et al., 2004). After proteolysis, both aGPCR cleavage fragments remain non-covalently bound to each other, forming a heterodimer at the cell surface (Gray et al., 1996; Krasnoperov et al., 1997). The most C-terminal β strand of the GAIN domain, which is severed from the NTF through self-cleavage (Araç et al., 2012; Salzman et al., 2016), functions as a tethered agonist (TA, alternatively termed *Stachel*, stalk), engaging with the heptahelical transmembrane (7TM) unit of the receptors to activate intracellular second-messenger pathways (Liebscher et al., 2014; Stoveken et al., 2015). However, the events leading to TA exposure, which are key to harnessing aGPCRs pharmacologically, have remained unclear (Liebscher and Schöneberg, 2016; Purcell and Hall, 2018).

X-ray crystal structures of aGPCR GAIN domains of rat Latrophilin 1 (ADGRL1/L1, subfamily I/L) and human BAI3 (ADGRB3/B3;
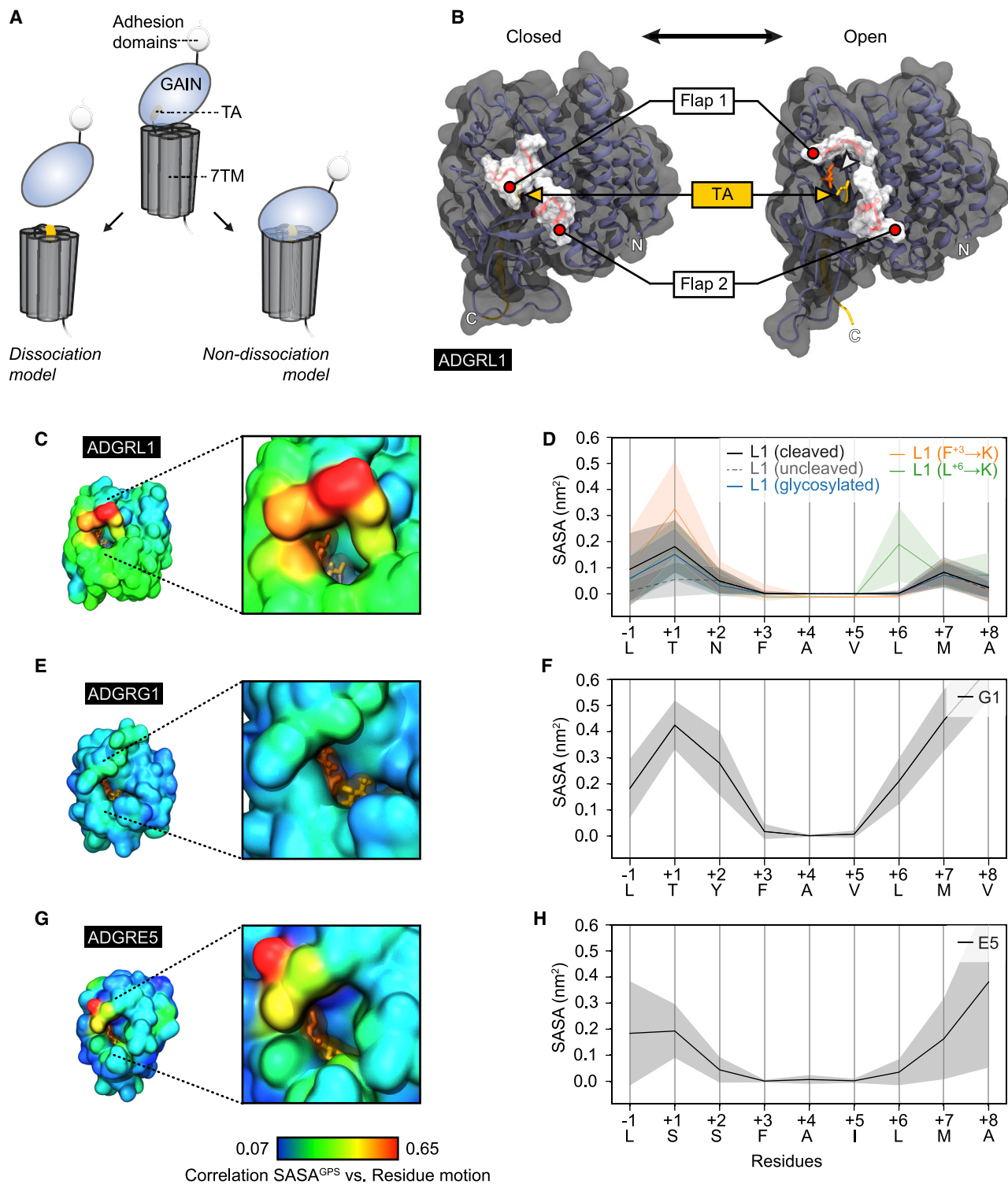
Figure 1. Molecular dynamics (MD) simulations reveal a conduit to the GPS, modulated by a lid at the protein surface

(A) Current signaling models on TA-dependent activation of aGPCRs. TA, tethered agonist; GAIN, GPCR autoproteolysis-inducing domain; 7TM, heptahelical transmembrane domain.

(B) Representative snapshots from MD simulations of L1 showing the exposure of the tethered agonist (yellow) at the level of the GPS (Leu^{-1} in orange) mediated by flaps 1 and 2.

(legend continued on next page)

subfamily VII/B), mouse GPR56 (ADGRG1/G1; subfamily VIII/G), and zebrafish GPR126 (ADGRG6/G6; subfamily VIII/G) have been solved to date (Araç et al., 2012; Leon et al., 2020; Salzman et al., 2016). All three structures show that subdomain B of the GAIN domain contains the proteolytic tripeptide, at which the domain is self-cleaved in cases of L1, G1, G6, and the TA (Figure S1A). Although the structures indicate that the TA is buried inside the GAIN domain, apparently inaccessible for 7TM domain engagement, mutagenesis studies in animal models and pharmacological interrogation of naturally cleavage-resistant aGPCR homologs have demonstrated that receptors with cleavage-deficient GAIN domains can function indistinguishably from cleavage-competent versions in a TA-dependent manner (Bohnekamp and Schöneberg, 2011; Prömel et al., 2012; Sando et al., 2019; Scholz et al., 2017; Wilde et al., 2016).

So far, two models for TA-dependent aGPCR activation have been discussed (Langenhan, 2020; Purcell and Hall, 2018; Vizurraga et al., 2020). In the dissociation model (hand grenade, shedding, or one-and-done model; Figure 1A), removal of the NTF, including the majority of the GAIN domain, uncaps the TA, which subsequently engages with the 7TM unit to stabilize an active receptor conformation (Liebscher et al., 2014; Stoveken et al., 2015). In contrast, the non-dissociation model (displacement or tunable model; Figure 1A) posits that TA-dependent receptor activation occurs through positional or conformational changes of the GAIN domain without its physical dissociation (Kishore et al., 2016; Salzman et al., 2017; Scholz et al., 2017). The models are not mutually exclusive and are compatible with the engagement of the extracellular region (ECR) with matricellular or cell-surface-anchored ligands in a mechanical force-dependent manner (Hamann et al., 2015; Purcell and Hall, 2018; Scholz et al., 2016; Zhu et al., 2019). Molecular structures of the 7TM domain or the GAIN-7TM domain pair of an aGPCR at atomic resolution that would aid in further defining the signaling mechanism of these receptors are not available yet.

In an effort to advance our knowledge of aGPCR activation, we have studied the dynamics and flexibility of GAIN domains from receptor homologs of four large aGPCR subfamilies (B, E, G, and L), which encompass 56% of all human aGPCRs. We have identified a common mechanism of TA exposure, which suggests that the TA becomes accessible for 7TM domain binding through the sideward opening of two flexible loops that shield the TA in the closed domain state. Our observations thus characterize the GAIN domain as a primary target structure for aGPCR pharmacotherapeutic approaches.

RESULTS

Molecular dynamics simulations uncover spontaneous, tethered agonist exposure in adhesion GPCR GAIN domains

To evaluate the intrinsic conformational flexibility of GAIN domains, and, in particular, of the TA, we conducted atomistic molecular dynamics (MD) simulations based on the crystal structures of the L1 GAIN/HormR domains (Araç et al., 2012) (PDB: 4DLQ). We set up this system, encompassing a standard protocol (see Method details), and ran nine independent simulations of 0.5 or 1.0 μs. In those simulations, the cleaved GAIN domain and the TA displayed an average root mean-squared deviation (RMSD) of 0.33 ± 0.10 nm and 0.19 ± 0.08, respectively. Those values are significantly less than the average RMSD of the preceding HormR domain (0.65 ± 0.21; Figures S2A and S2B) and are consistent with moderate changes in the relative position or structure of the GAIN domain.

However, upon visual inspection of the trajectories, two regions of the GAIN domain in close proximity to the TA appeared notably mobile. These regions were termed flap 1 (L1: Glu^{-64}-Arg^{-61}) and flap 2 (L1: Tyr^{-32}-Met^{-27}), respectively (Figure 1B; for consensus numbering, see Figure S1B). The analysis of protein motion per residue, assessed by root mean-squared fluctuation (RMSF), confirmed that both flaps 1 and 2 are significantly flexible (Figures S3A and S3B). Both flaps are located at the surface of subdomain B and form a double-winged lid above the TA that is buried inside the GAIN domain (Figures 1B, S1A, and S1C). In MD simulations, the lid repetitively opened through the lateral motion of both flaps (Video S1, left) corresponding to transitions between a closed and opened domain conformation (Figure 1B; links 1 and 10 at https://doi.org/10.5281/zenodo.4114651 provide fully interactive visualization of an L1 representative simulation via MDsrv [Tiemann et al., 2017]). In the open state, the displacement of the flaps revealed a conduit that penetrates from the domain surface down to the TA exposing its N-terminal residues to the solvent (Figure 1C), which are most highly evolutionarily conserved and essential for metabotropic receptor signaling (Liebscher et al., 2014; Stoveken et al., 2015). Further analysis showed that it is this motion of the flaps, particularly flap 1, that is most largely correlated with changes in GPS solvent-accessible surface area (SASA) in the L1 GAIN domain (Figures 1C, S3C, and S3D), hence, confirming that flap movement modulates solvent accessibility to the TA. Interestingly, cleavage at the GPS site additionally increases the

---

(C, E, and G) Correlation values between GPS exposure, i.e., solvent-accessible surface area (SASA) and residue motion mapped onto the GAIN domain surface. Yellow and red colors (i.e., higher correlations) clearly highlight the lid exposing the GPS in L1 (C) and E5 (G). The conduit spanning from the solvent into the domain's center holding the TA is apparent in each system (arrowheads). The disordered loop connecting beta sheets S5 and S6 (see Figure S1) was omitted for clarity in all figures.

(D, F, and H) Average SASA of the TA for cleaved L1, G1, and E5 GAIN domains (solid gray) and uncleaved L1 (dashed gray). Shaded areas represent the 95% confidence interval. Replacement of the +3 (orange) and +6 (green) residues with a lysine was additionally simulated for cleaved L1. As expected, the SASA is generally greater at the C-terminal end of the TA (exposed to solvent). In L1, the +3 mutation (orange) induced a slight gain of exposure at position +1, whereas these rearrangements did not have any effect on the exposure of any other TA residue. The +6 mutation (green) increased the mobility of the loop connecting H8 and S6 in subdomain B (positions −100 to −80; Figure S3A, top panel, green curve), which, in turn, increased the exposure of the +6 position itself. Importantly, none of the insertions affected the correlation between flap motion and TA exposure. Conversely, all systems, especially G1 (F), show an unexpectedly high SASA of residues −1 and +1 (i.e., the GPS; arrowheads), mostly buried in the crystal structures. Notably, GPS cleavage increases TA exposure in the L1 system, whereas glycosylation has no effect.

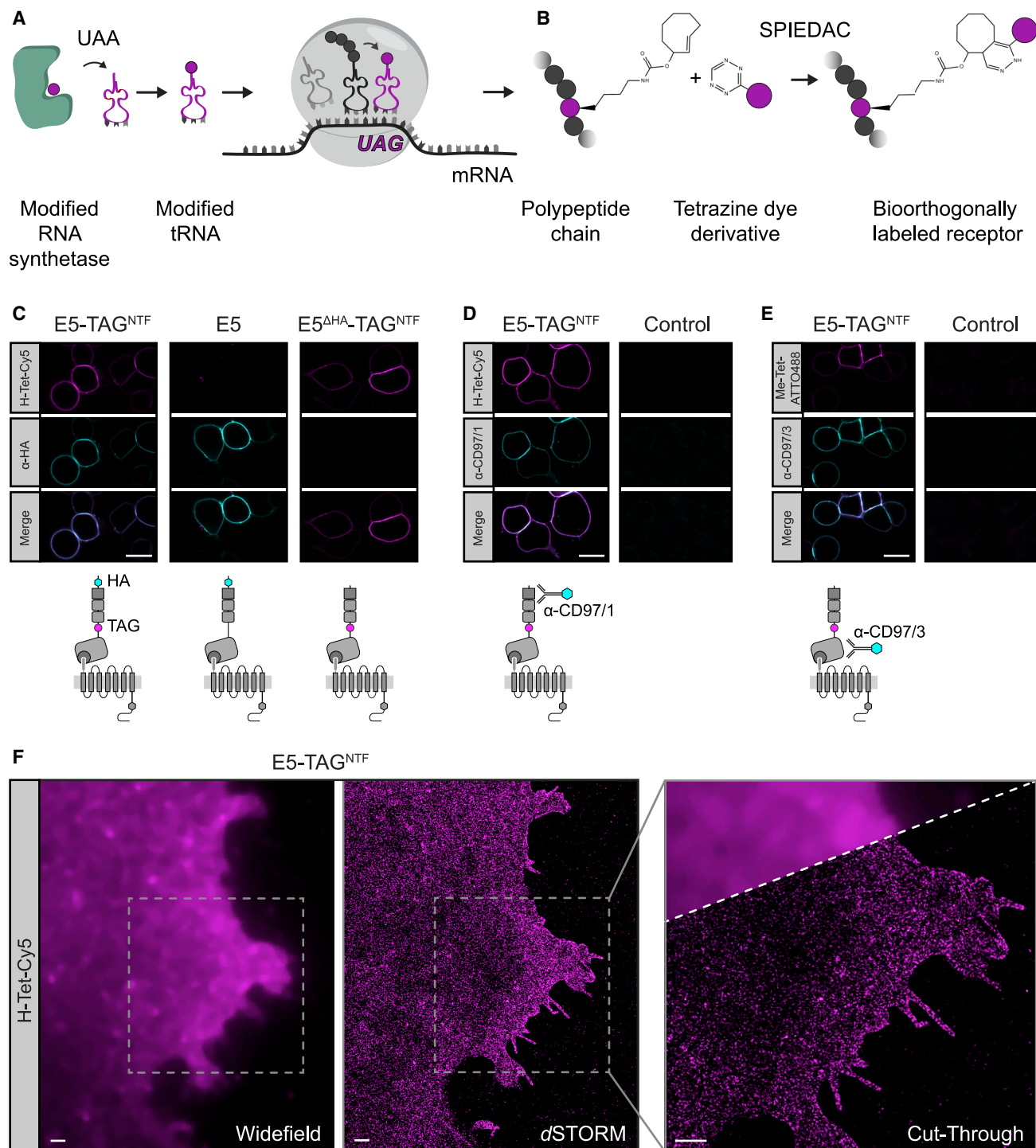See also Figures S1–S3 and Table S3.

Figure 2. Orthogonal click-labeling of E5 provides a minimally invasive, rapid, and specific means for attaching fluorescent labels to the receptor

(A) Expansion of the host genetic machinery with a modified RS/tRNA pair enables the insertion of UAA during protein biosynthesis.

(B) UAAs possess a reactive, strained alkene group, which poises the carrier protein for the formation of a covalent bond with a tetrazine group via a SPIEDAC click reaction.

(C) Confocal images of HEK293T cells expressing E5, E5$^{\Delta HA}$-TAG^{NTF}, and E5-TAG^{NTF} were click-labeled with H-Tet-Cy5 (magenta) and immunostained with α-HA-Alexa Fluor 488 antibody (cyan). Scale bar, 20 μm.

mobility of flap 1 (Figure S3A) and increases exposure of the terminal -1 and $+1$ residues generated upon cleavage (Figure 1D). This occurs mainly because of altered interactions between flap 1, L^{-1}, and T^{+1}, yet proteolysis does not affect the degree of exposure of the core TA residues (Figure 1D), whereas glycosylation of the L1 GAIN shows no effect on these properties (Figures 1D, S3C, and S3D). Likewise, we observed TA exposure at the most N-terminal residues along the N→C axis of the TA, which conforms to a lateral opening of the GAIN domain.

To study whether this mechanism is also present in other aGPCRs, we performed a novel set of MD simulations based on the crystal structure of the cleaved G1 GAIN/PLL domains (Salzman et al., 2016) (PDB: 5KVM). We followed a similar set-up protocol and simulated nine independent replicas (9×0.5 μs). Corresponding to L1, the cleaved G1 GAIN domain (particularly the TA) showed a low average RMSD of 0.45 ± 0.17 nm, consistent with a moderately high positional stability (Figure S2C). However, flap 1 (Asn^{-64}-Lys^{-62}) and flap 2 (Glu^{-32}-Thr^{-25}) also displayed high RMSFs, largely exposing the TA (Figures S3A and S3B; link 2 at https://doi.org/10.5281/zenodo.4114651). The cryptic conduit toward the TA displays, however, differences in size and surface morphology between L1 and G1. Specifically, G1 flaps 1 and 2 delimit a larger conduit that, in spite of flap movement, seems to remain continuously exposed during the simulations (Video S1, right). Monitoring the minimum distance between flaps 1 and 2 during the simulation confirms that the G1 flaps remain significantly separated from each other when compared with L1 (Figure S3G), and accordingly, our correlation analysis showed that, in the case of G1, GPS solvent accessibility does not depend on flap position (Figures 1E, 1F, S3E, and S3F).

To include a set of GAIN domains representing a third large aGPCR subfamily, subfamily II/E, we obtained a homology model for the cleaved GAIN domain of ADGRE5/E5/CD97. The model was based on the crystal structure of the brain-specific angiogenesis inhibitor 3 GAIN domain (PDB: 4DLO; see Method details) and represented the overall A–B subdomain structure yet with a distinctly shorter subdomain A than found in other aGPCRs because of the lack of several α helices (Araç et al., 2012) (Figure S1; link 3 at https://doi.org/10.5281/zenodo.4114651). MD simulations based on this homology model corroborated the results found for L1 and G1 and displayed similar flap movements (Figures S3A and S3B; link 4 at https://doi.org/10.5281/zenodo.4114651) and open-closed conformation transitions (Video S2, left) that modulate solvent accessibility to the TA (Figures 1G and 1H).

Bioorthogonal labeling of adhesion GPCRs

It is known that spontaneous fluctuations are representative of those conformational changes modulating protein function (Changeux and Edelstein, 2011; Sun et al., 2018), and we thus assumed that the motions in isolated GAIN domains also occur in full-length aGPCRs. To directly test that, we developed a protocol encompassing genetic code expansion (GCE) in combination with unnatural amino acid (UAA) incorporation and subsequent click labeling of tags within different aGPCRs. That approach offers a sensitive, yet minimally space-demanding, method to place single amino-acid-sized molecular tags in sterically confined protein regions, such as domains. As a prototype aGPCR, we started with E5, an aGPCR with roles in immune processes (Lin et al., 2017), angiogenesis (Wang et al., 2005), and cancer (Aust et al., 2016; Yin et al., 2018). E5 is autoproteolytically cleaved (Gray et al., 1996), shows advantageous heterologous expression behavior (Hamann et al., 1996), and has several monoclonal antibodies detecting different receptor epitopes available (Eichler et al., 1994; Kwakkenbos et al., 2000). For fluorescence labeling of E5 by bioorthogonal click chemistry, we employed a tetrapartite pyrrolysine Pyl^{RS}/$tRNA^{Pyl}$ system encompassing a UAA, an aminoacyl-tRNA synthetase (Pyl^{RS}), a cognate tRNA ($tRNA^{Pyl}$), and a genetic-expression template carrying a *UAG* amber stop codon in its open reading frame, which is interpreted as UAA encoding (Nikić et al., 2015) (Figure 2A). Trans-cyclooctene-L-lysine (TCO*-lysine) was used as the UAA in all assays, which reacts in a strain-promoted inverse-electron-demand Diels-Alder cycloaddition (SPIEDAC) with the tetrazine group of functionalized organic fluorophores (Beliu et al., 2019; Nikić et al., 2015) (Figure 2B). To assess E5 expression also through ELISA (enzyme-linked immunosorbent assay), immunoblotting, and immunohistochemistry/microscopy, we appended an N-terminal hemagglutinin (HA) tag after the signal peptide and a V5 epitope (from simian virus 5) tag within the CTF before the PDZ binding motif (Figure S4A).

First, we placed a bioorthogonal label in the NTF of E5 (Figure S4A). Based on previous aGPCR engineering approaches (Prömel et al., 2012; Scholz et al., 2017), the label position was chosen in an unstructured protein region between EGF5 and the start of the GAIN domain of E5 (based on an homology model calculated by the PSIPRED2 service [Jones, 1999]) to avoid deleterious effects to receptor folding and to ensure label exposure to the incubating medium (Figures S4A and S4B). To establish specificity of the bioorthogonal and α-HA immunolabels, we separately transfected E5 (HA-tag only), $E5^{\Delta HA}$-TAG^{NTF} (orthogonal tag only), and E5-TAG^{NTF} (both tags) into HEK293T and COS7 cells. All versions were co-expressed with a Pyl^{RS}/$tRNA^{Pyl}$ pair plasmid in the presence of TCO*-lysine and co-labeled with the cell-membrane-impermeable tetrazine-dye H-Tet-Cy5 and a monoclonal mouse-α-HA antibody directly conjugated with Alexa Fluor 488. Two-channel confocal fluorescence microscopy of transfected cells confirmed that both labels specifically marked their target tags (Figure 2C). Next, we used two monoclonal antibodies, α-CD97/1 (Eichler et al., 1994) and α-CD97/3 (Kwakkenbos et al., 2000) (Figures S4A and S4D), to compare the labeling efficiency of E5 by bioorthogonal click chemistry

---

(D and E) E5-TAG^{NTF} was click-labeled with (D) H-Tet-Cy5 (magenta) and immunostained with mouse-α-CD97/1-Alexa Fluor 532 antibody (cyan) or (E) click-labeled with Me-Tet-ATTO 488 (magenta) and immunostained with mouse-α-CD97/3-Alexa Fluor 647 antibody (cyan). Control panels: cells transfected with the modified RS/tRNA pair only. Scale bars, 20 μm.

(F) Click-labeled E5-TAG^{NTF}, imaged by *d*STORM, shows homogeneous distribution of aGPCRs in the plasma membrane without any indication of receptor cluster formation. (Right) Cut-through image showing the same sample imaged by widefield fluorescence and *d*STORM. Scale bars, 1 μm.

See also Figure S4.
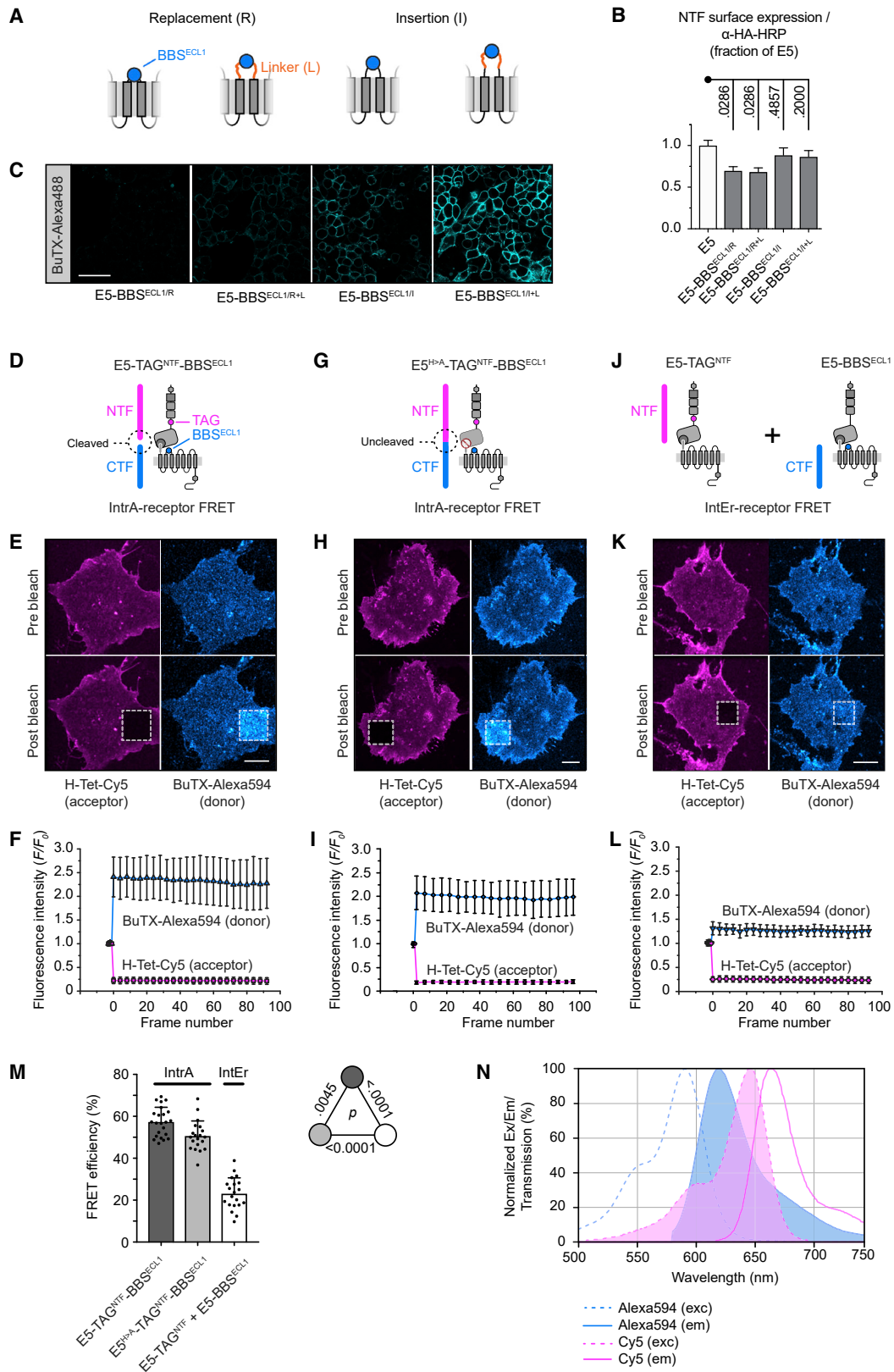
(legend on next page)

with classical immunolabeling approaches. Expression domains of E5-TAG^{NTF} in HEK293T cells for both antibody labels showed complete overlap with the bioorthogonal click label, indicating that the tags are present in the same target protein and that the bioorthogonal click chemistry approach matches conventional immunolabeling methods for aGPCRs (Figures 2D, 2E, S4D, and S4E). Expectedly (Ryu and Schultz, 2006; Schmied et al., 2014), inefficient amber stop-codon suppression and UAA incorporation reduced the total and surface pool of full-length receptors evaluated by ELISA (Figures S4F, S4G, S5C, and S5D) and western blot analyses of non-proteolyzable E5$^{H>A}$-TAG^{NTF} and E5$^{S>A}$-TAG^{NTF} receptor mutants (Figures S4H and S4I). However, comparable membrane immunolabeling of E5 suggests that bioorthogonal click label insertion into the receptor protein did not negatively affect receptor structure or membrane trafficking (Figures 2C–2E).

To assess the distribution of E5 in the plasma membrane, we used super-resolution fluorescence imaging by *direct* stochastic optical reconstruction microscopy (*d*STORM; Figure 2F) (Heilemann et al., 2008; van de Linde et al., 2011). Cells were fixed after bioorthogonal labeling and washing to minimize mobility of membrane receptors during imaging. The *d*STORM images showed a homogeneous distribution of E5 in the plasma membrane without signs of clustering (Figure 2F).

These results demonstrate that the bioorthogonal labeling approach is a sensitive and specific tool to attach chemical tags to aGPCRs, specifically to those ones without alternative methods for detection, e.g., specific antisera.

NTF and CTF of E5 form a heterodimer at the cell surface

Formation and maintenance of non-covalently associated NTF-CTF heterodimers after proteolytic processing through the GAIN domain is a core element in structural paradigms of aGPCRs and a prerequisite for the non-dissociation model of aGPCR signaling. Direct proof for such heterodimers is, however, lacking thus far. We developed a tagging strategy for direct assessment of aGPCR heterodimer formation through Förster resonance energy transfer (FRET), in which the bioorthogonal

TAG^{NTF} labels the NTF of E5. As a suitable CTF tag, we harnessed a 13-amino acid peptide derived from the nicotinic acetylcholine receptor that serves as a high-affinity binding site for the neurotoxic α-Bungarotoxin component (α-BuTX) of the many-banded krait *Bungarus multicinctus* venom (Harel et al., 2001). We introduced the bungarotoxin binding site (BBS) tag, flanked by flexible linkers in the first extracellular loop (ECL1) of E5 (Figures 3A and 3D). ELISA measurements of E5-BBS^{ECL1} confirmed that BBS-tag insertion did not affect surface delivery of receptors (Figure 3B). We then labeled HEK293T and COS7 cells expressing E5-BBS^{ECL1} with α-BuTX conjugated with Alexa Fluor 488 displaying strong and specific receptor labeling (Figure 3C).

Finally, we combined both NTF and CTF labels in E5-TAG^{NTF}-BBS^{ECL1}, allowing for separate visualization of both labels through confocal microscopy after adding H-Tet-Cy5 (NTF) and α-BuTX-Alexa Fluor 594 (CTF; Figure 3E), a fluorophore pair, whose emission spectra have sufficient spectral overlap to enable FRET measurements for donor-acceptor distance < 10 nm (Figure 3N). Subsequently, we assessed FRET effects occurring within the double-labeled E5-TAG^{NTF}-BBS^{ECL1} expressed in COS7 cells through dequenching of donor fluorescence by acceptor photobleaching (Figure 3F). To that end, we illuminated the transfected cells at 555 nm to detect pre-bleach donor intensity, bleached a selected region of interest (ROI) at 640 nm for ~20 s (100 iterations, 100% intensity) to achieve acceptor bleaching, and subsequently recorded postbleaching donor fluorescence intensities (Video S3). For the E5-TAG^{NTF}-BBS^{ECL1} receptor, we observed a FRET efficiency of 57% ± 7% (Figure 3M). FRET measurements of a GAIN-proteolysis-deficient receptor E5$^{H>A}$-TAG^{NTF}-BBS^{ECL1} returned similar FRET efficiencies (51% ± 7%) demonstrating that NTF and CTF of auto-cleaved and non-cleaved E5 variants are positioned at similar distance to each other, i.e., NTF and CTF of proteolyzed E5 variants are part of the same heterodimer (Figures 3G–3I and 3M).

To control for inter-receptor FRET events emerging from labeled aGPCR fragments in the plasma membrane, which are not part of the same receptor heterodimer, we co-transfected single-label E5-TAG^{NTF} and E5-BBS^{ECL1} receptors into COS7

Figure 3. NTF and CTF form receptor heterodimers at the cell surface

(A) A BBS was introduced into the ECL1 of E5 in different layouts: by replacing part of the loop without (R) or with flanking linkers (R + L) and by inserting the BBS-tag into ECL1 without replacement of loop residues, also without (I) and with linkers (I + L).

(B) ELISA measurements indicating surface delivery of different E5-BBS^{ECL1x} receptors. Data are shown as means ± SEM (N [independent sets] = 1, n [measurements per set] = 4).

(C) Staining of HEK293T cells expressing E5-BBS^{ECL1x} receptors with Alexa-Fluor-488-conjugated α-BuTX shows increasing receptor labeling with E5-BBS$^{ECL1/I+L}$ as most strongly labeled receptor version. This BBS-tagged E5 layout was chosen for subsequent FRET assays. Scale bar, 50 μm.

(D) Layout of auto-proteolyzed E5 receptor dually tagged with FRET-compatible fluorophores at the NTF (bioorthogonal click label and H-Tet-Cy5; magenta) and CTF (bungarotoxin-binding sequence and Alexa Fluor 594; blue).

(E, H, and K) Representative confocal fluorescence images of COS7 cells expressing different E5 receptor versions, as depicted above. Acceptor photobleaching was performed in the region marked with the dotted square. (F and H) Strong effects derived from FRET within a NTF-CTF heterodimer are visible as an increase in donor fluorescence, whereas only minimal FRET is apparent between fluorophores attached to different receptor heterodimers (K). Scale bars, 10 μm.

(F, I, and L) Representative time course of normalized donor (blue) and acceptor (magenta) fluorescence. Data are shown as means ± SD. N = 3; n = 23 for (F); n = 20 for (I); and n = 20 for (L).
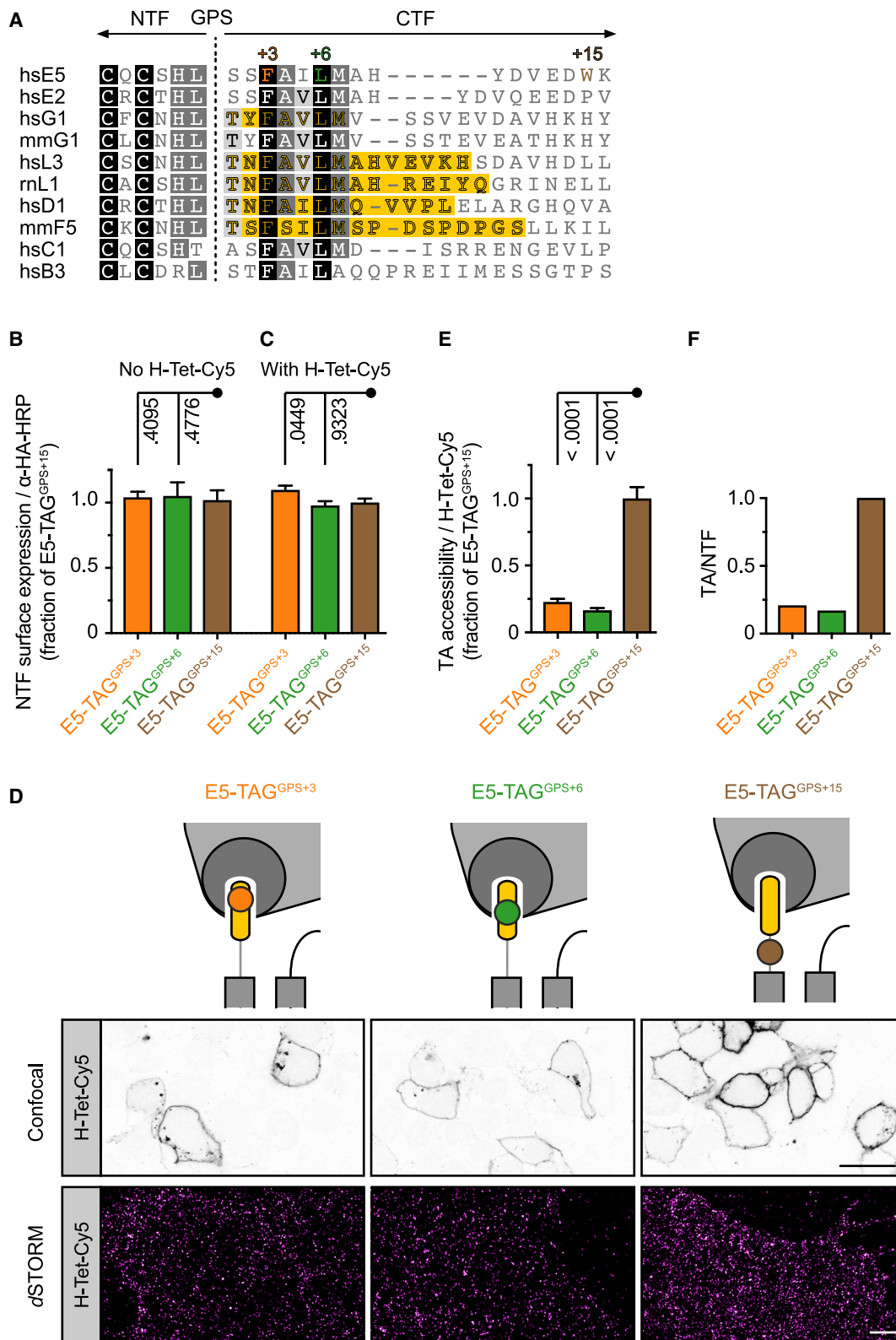
(G) GAIN domain autoproteolysis-deficient double tagged E5 layout.

(J) Layout of inter-receptor FRET control receptors.

(M) Comparison of FRET efficiencies occurring through intra-receptor and inter-receptor FRET experiments. Data are shown as means ± SD.

(N) Spectral overlap of FRET pair Alexa Fluor 594 and Cy5. The absorption (dotted line) and emission spectra (solid line) of the donor Alexa Fluor 594 (blue) and the acceptor Cy5 (magenta) are shown, respectively. Energy transfer occurs because of the spectral overlap between the emission of the donor and absorption of the acceptor (intersection).

See also Videos S1, S2, S3, and S4.

A

	NTF	GPS		+3	+6				+15	
hsE5	C Q C S H L		S S	F	A I	L	M A H - - - - - - - Y D V E D W K
hsE2	C R C T H L		S S	F	A V	L	M A H - - - - Y D V Q E E D P V
hsG1	C F C N H L		T Y	F	A V	L	M V - - S S V E V D A V H K H Y
mmG1	C L C N H L		T Y	F	A V	L	M V - - S S T E V E A T H K H Y
hsL3	C S C N H L		T N	F	A V	L	M A H V E V K H S D A V H D L L
rnL1	C A C S H L		T N	F	A V	L	M A H - R E I Y Q G R I N E L L
hsD1	C R C T H L		T N	F	A I	L	M Q - V V P L E L A R G H Q V A
mmF5	C K C N H L		T S	F	S I	L	M S P - D S P D P G S L L K I L
hsC1	C Q C S H T		A S	F	A V	L	M D - - - I S R R E N G E V L P
hsB3	C L C D R L		S T	F	A I	L	A Q Q P R E I I M E S S G T P S

B No H-Tet-Cy5

NTF surface expression / α-HA-HRP (fraction of E5-TAG^GPS+15)

.4095   .4776

E5-TAG^GPS+3   E5-TAG^GPS+6   E5-TAG^GPS+15

C With H-Tet-Cy5

.0449   .9323

E5-TAG^GPS+3   E5-TAG^GPS+6   E5-TAG^GPS+15

E

TA accessibility / H-Tet-Cy5 (fraction of E5-TAG^GPS+15)

<.0001   <.0001

E5-TAG^GPS+3   E5-TAG^GPS+6   E5-TAG^GPS+15

F

TA/NTF

E5-TAG^GPS+3   E5-TAG^GPS+6   E5-TAG^GPS+15

D

E5-TAG^GPS+3   E5-TAG^GPS+6   E5-TAG^GPS+15

Confocal H-Tet-Cy5

dSTORM H-Tet-Cy5

(legend on next page)

cells (Figures 3J–3L). FRET measurements of that experiment revealed FRET efficiencies of 23% ± 8% (Figure 3M; Video S4). This indicates that a very small amount of inter-receptor FRET contributes to the FRET signals of the double-labeled receptor, and that, consequently, FRET values obtained from E5-TAG^{NTF}-BBS^{ECL1} experiments represent intra-receptor FRET events. Thus, this set of experiments shows that NTF and CTF of E5 reside as a heterodimer at the cell membrane.

Demonstration of tethered agonist exposure employing bioorthogonal labels

To place bioorthogonal labels into the TA region of aGPCRs, we aligned the amino acid sequence of the E5 GAIN domain region containing the last β strand with other aGPCRs, for which the TA was experimentally characterized and/or crystal structures of the GAIN domain are known. The alignment showed a high level of residue conservation across different receptors (Figure 4A) (Liebscher and Schöneberg, 2016) and implied that the +3 and +6 (see Figure S1B for consensus numbering) positions relative to the GPS of E5 are located within the TA. In contrast, a control residue at position +15 is located in the linker region between the GAIN and 7TM domains, thus likely irrelevant for tethered agonism in E5 (Hilbig et al., 2018) and located outside the GAIN domain (Figure 4A). We reasoned that individual introduction of UAAs at these three positions would allow us to probe the accessibility of the TA of membrane-resident E5. A priori modeling of UAA insertion into the E5 GAIN domain confirmed that introduction of TCO*-lysine and linkage to H-Tet-Cy5 at position +3 (link 5 at https://doi.org/10.5281/zenodo.4114651) and +6 (link 6 at https://doi.org/10.5281/zenodo.4114651) of the E5 GAIN domain is sterically feasible (Figure S5B). Further, to validate that lysine insertion within the TA did not affect surface exposure on its own, we simulated molecular movements of the L1 GAIN domain upon UAA introduction at positions +3 and +6 using MD simulations. As expected, the hydrophobic environment of the +3 and +6 positions rearranged upon lysine incorporation. Although these changes did not affect the exposure of residue +3 to the solvent, the mutation caused an increased solvent exposure for residue +6 (Figures S3C and S3D), although not via the conduit revealed by flaps 1 and 2 but, rather, because of the increased mobility of the loop around positions −100 to −80 (Figure S3A, top panel). Importantly, none of the lysine-induced rearrangements affected the overall exposure of the TA or the correlation between flap motion and GPS exposure (Figures

S3C and S3D; links 8 and 9 at https://doi.org/10.5281/zenodo.4114651).

Accordingly, we generated receptor variants E5-TAG^{GPS+3}, E5-TAG^{GPS+6}, and E5-TAG^{GPS+15} (Figures S4A, S4C, and S5B). We first ascertained whether TCO*-lysine integration into the different E5-TAG^{GPS+X} variants affect receptor biosynthesis or surface delivery in the absence and presence of the dye because of potential structural effects on the GAIN domain. To that end, we co-transfected two batches of HEK293T cells with Pyl^{RS}/tRNA^{Pyl} and E5-TAG^{GPS+3}, E5-TAG^{GPS+6}, E5-TAG^{GPS+15}, or E5-TAG^{NTF} in the presence of TCO*-lysine. The first batch was fixed after 24 h without addition of H-Tet-Cy5, and surface expression was immediately determined by ELISA. We found that expression of E5-TAG^{GPS+3}, E5-TAG^{GPS+6}, and E5-TAG^{GPS+15} was not impaired compared with that of E5-TAG^{NTF} (Figure S5C). In addition, surface delivery of the E5-TAG^{GPS+X} receptor versions was indistinguishable from each other (Figure 4B). The second batch was supplemented with H-Tet-Cy5 for 30 min before fixation and ELISA reading to ensure incorporation of the dye at the surface-exposed E5-TAG^{GPS+X} receptors. Similarly, we found no effect of dye addition on surface expression of E5-TAG^{GPS+X} receptors (Figures 4C, S5C, and S5D).

To further control for GAIN domain integrity, we labeled the different E5-TAG^{GPS+X} versions with a polyclonal α-CD97 antiserum and assessed the surface labeling of E5-TAG^{GPS+3}, E5-TAG^{GPS+6}, and E5-TAG^{GPS+15} (Figures S5F and S5G). We also observed no overt differences in surface levels of receptors carrying UAA inserts inside the GAIN domain or the linker region (Figure S5H). This establishes that TCO*-lysine integration into the TA and H-Tet-Cy5 insertion does not impede GAIN domain formation and receptor trafficking. In addition, western blot analysis of E5-TAG^{GPS+3}, E5-TAG^{GPS+6}, E5-TAG^{GPS+15}, and a set of similar receptors carrying an autoproteolysis-disabling S > A mutation (E5$^{S>A}$-TAG^{GPS+3}, E5$^{S>A}$-TAG^{GPS+6}, and E5$^{S>A}$-TAG^{GPS+15}) showed that TCO*-lysine integration at positions +3 and +6 disables GAIN domain self-cleavage, whereas it is retained in the presence of the +15 label (Figure S5I). Hence, NTF and CTF labels reside in the same receptor molecule.

Next, we performed imaging experiments to analyze the labeling efficiency of the receptor variants by confocal microscopy and dSTORM. Intriguingly, we found that E5-TAG^{GPS+3} and E5-TAG^{GPS+6} exhibited weaker, but distinct, bioorthogonal click labeling from that of E5-TAG^{GPS+15}, indicating that E5-TAG^{GPS+15} is more-easily accessible for the click label (Figure 4D). We capitalized on that finding and measured

---

Figure 4. Bioorthogonal click-labeling reveals that the tethered agonist region of aGPCRs is accessible in the intact receptor heterodimer
(A) Amino acid sequence alignment of exemplary aGPCR regions encompassing the GPS and TA regions. Dimensions of experimentally confirmed tethered agonists, in yellow, were taken from Stoveken et al. (2015) (hsG1), Müller et al. (2015) (rnL1), Liebscher et al. (2014) (hsD1), and Brown et al. (2017) (mmF5). Label positions in the TA and linker region between the GAIN and 7TM domain are indicated in magenta; residue numbering with respect to GPS is indicated above the alignment. +3 and +6 are located in the center of the TA; +15 is outside the TA.
(B and C) Quantification of surface levels for bioorthogonally labeled E5 variants without (B) and with (C) addition of the dye by ELISA shows no detrimental effect on the amount of plasma membrane resident receptors. Dataset normalized to surface level of E5-TAG^{GPS+15}. Data are shown as means ± SEM (N = 3, n = 4). See also Figures S4, S5C, and S5D. Color scheme for click-label positions in all receptor designs shown in this work: +3 (orange), +6 (green), and +15/17 (brown).
(D) Confocal and dSTORM images of bioorthogonally marked E5 variants reveal a distinct label at all positions. Note that the TA +3 and +6 labels are reduced compared with the +15 label. Scale bars, 20 μm (confocal) and 1 μm (dSTORM).
(E) Quantification of intensity of all bioorthogonal labels. Data are shown as means ± SEM (N = 3, n = 20).
(F) Ratio of TA label versus NTF quantities for all bioorthogonal labels in E5 demonstrate accessibility of the TA to the solvent in intact E5 heterodimers.
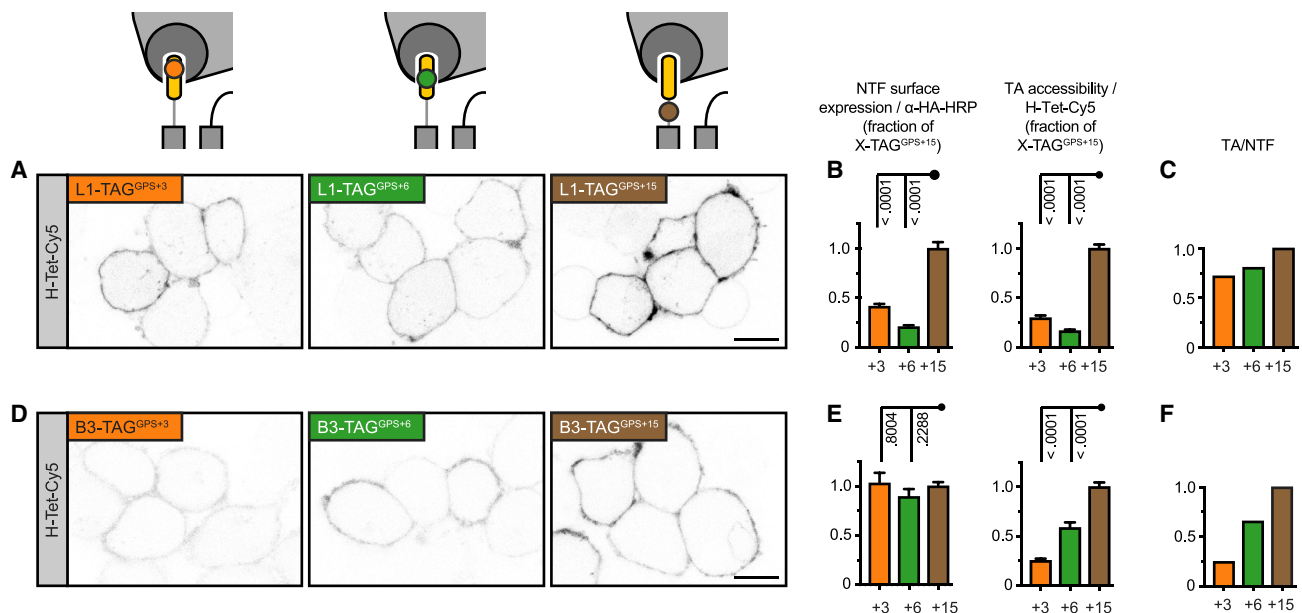See also Figures S4–S6.

Figure 5. Bioorthogonal labeling of L1 and B3 confirms the generality of TA exposure in intact GAIN domains of aGPCRs

(A) Confocal images of L1 variants bioorthogonally marked at the GPS region. Scale bar, 10 μm.

(B) Quantification of surface levels (data are shown as means ± SEM [N = 3, n = 4]) and TA accessibility for bioorthogonally labeled L1 variants (data are shown as means ± SEM [N = 3, n = 20]). See also Figure S6A.

(C) Ratio of TA label versus NTF quantities for all bioorthogonal labels in L1 showing the accessibility of the TA to the solvent in intact L1 heterodimers.

(D–F) Similar measurements as in (A)–(C) for the human B3 full-length receptor. Scale bar, 10 μm. See also Figure S6B.

bioorthogonal click-label counts at the different positions within the TA to estimate accessibility of the GAIN domain interior to the solvent. As a control, we compared E5-TAG^{GPS+3} and E5-TAG^{GPS+6} expression to cells only transfected with Pyl^{RS}/tRNA^{Pyl} and exposed to TCO*-lysine and H-Tet-Cy5 and cells that expressed the E5-TAG^{GPS+15} receptor. We found similar counts for E5-TAG^{GPS+3} and E5-TAG^{GPS+6} (Figures 4D and 4E). Further, E5-TAG^{GPS+15} exhibited an approximately 3-fold labeling intensity over the individual E5-TAG^{GPS+3} and E5-TAG^{GPS+6} receptors (Figures 4D and 4E). These findings were confirmed after normalizing dye labeling to the NTF surface expression of all receptor layouts (Figure 4F). As an additional control, we repeated the TA accessibility analysis with E5$^{S>A}$-TAG^{GPS+3}, E5$^{S>A}$-TAG^{GPS+6}, and E5$^{S>A}$-TAG^{GPS+15} receptor variants and found no differences in TA label intensity to the receptors without the cleavage-disabling S > A mutation (Figures S5J and 5K). These sets of results corroborate the MD simulation findings and confirm that the TA inside the GAIN domain of E5 is solvent exposed in the intact fold.

To test for the generality of TA accessibility, we generated sets of the bioorthogonally taggable receptor version for L1 and B3 containing homologous *TAG* codons at positions +3, +6, and in the TA-TM1 linker (G1 was excluded from further analyses because of its poor expression in HEK293T cells). Quantification of receptor surface expression and TA label density in HEK293T cells returned results consistent with data on E5 showing that the TA in L1 (Figures 5A–5C and S6A) and B3 (Figures 5D–5F and S6B) is also solvent accessible in the intact GAIN domain of the full-length receptor.

A vibratory urticaria mutation in E2/EMR2 GAIN domain does not increase TA accessibility

With that knowledge, we finally investigated the effect of a human missense mutation in the GAIN domain of E2, which had been linked to a rare autosomal-dominant form of vibratory urticaria (VBU) (Boyden et al., 2016). E2^{C492Y} was suggested to destabilize the NTF-CTF heterodimer through reduction of the non-covalent interaction network of the auto-proteolyzed GAIN domain. We reasoned that this effect should lead to increased exposure of the TA of E2 and be readily detectable through bioorthogonal labeling. MD simulations of a homology model of the E2 GAIN domain (see STAR methods) showed, again, that the domain surface possesses a high degree of intrinsic flexibility above the GPS (Figures S3A and S3B; link 7 at https://doi.org/10.5281/zenodo.4114651). This is reflected by the presence of the two flaps at the domain surface, similar to the other GAIN domains, which eventually open to give access to the TA (Video S2, right). In addition, as shown in Figures 6A, 6B, S3B, and S3F, flaps 1 and 2 clearly modulate GPS accessibility because both protein regions correlate with an increase in SASA. Intriguingly, neither insertion of the VBU mutation nor the enforced cysteine bridging between Cys^{-26} and Cys^{-190} (i.e., Cys^{492}) had a marked effect on the domain's conformational flexibility (Figures S3A and S3B) and TA accessibility (Figures 6A, 6B, S3A, S3B, S3E, and S3F), indicating that the conformation of the domain is sculpted by other structural constraints.

To further test that finding, we engineered E2^{C492Y}-TAG^{+3}, E2^{C492Y}-TAG^{+6}, and E2^{C492Y}-TAG^{+17} receptor versions (Figure S6C) and co-labeled their HA and respective bioorthogonal
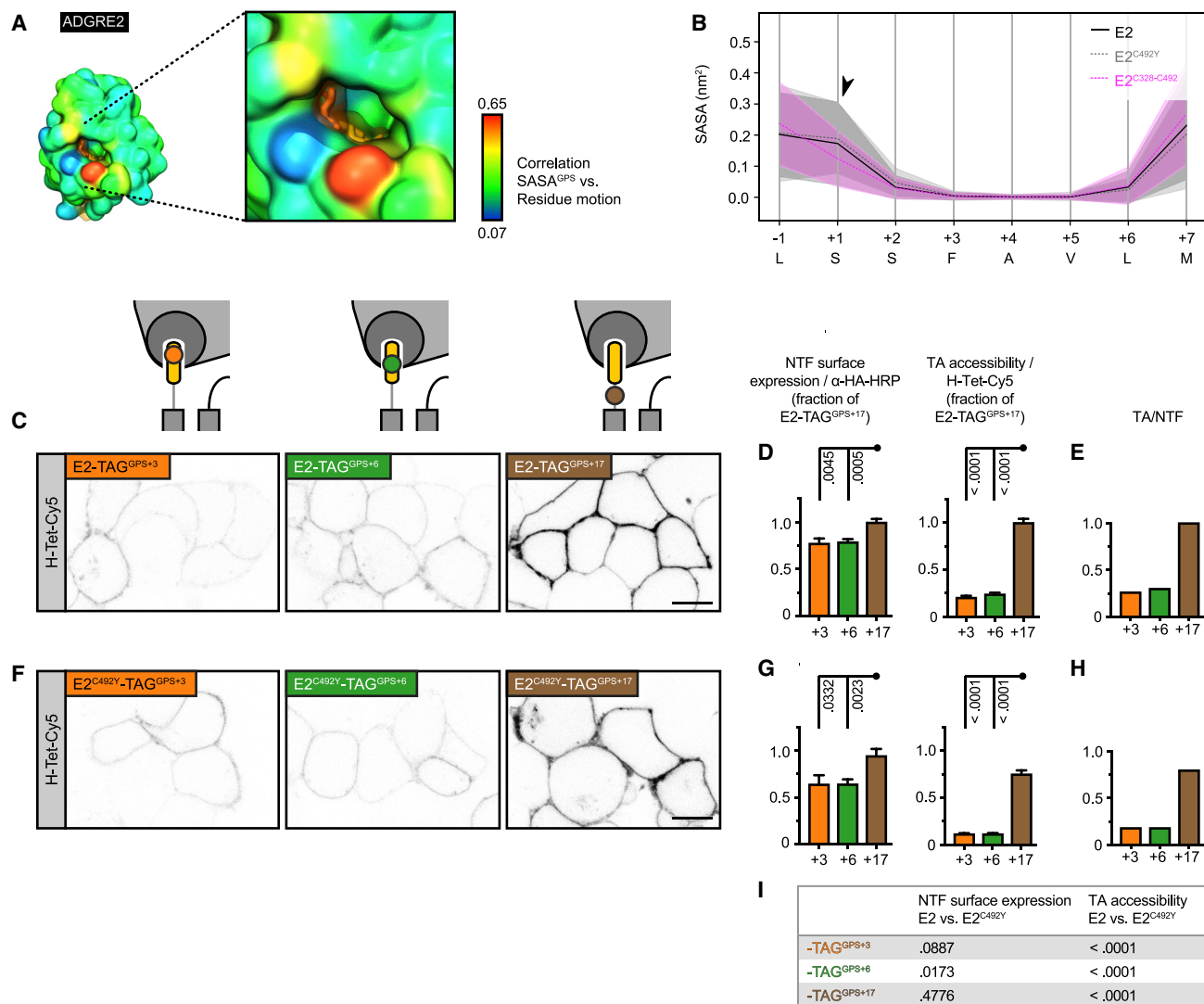
Figure 6. The vibratory urticaria mutation in the E2 GAIN domain does not affect TA accessibility

(A) Values of the correlation between GPS exposure, i.e., SASA, and residue motion of the E2 wild-type system mapped onto the GAIN domain surface. The surface exhibits an apparent conduit to the TA (arrowheads) as in the other systems. Importantly, flaps 1 and 2 are the most correlated regions of the protein. The disordered loop connecting beta sheets S5 and S6 (see Figure S1) was omitted for clarity in all figures.

(B) Average SASA of the tethered agonist for wild-type E2, E2^{C492Y}, and E2 with an artificially introduced cysteine bridge between Cys328 and Cys492 (E2$^{C328-C492Y}$). Similar to L1, G1, and E5 (Figures 1D, 1F, and 1H), the SASA of the TA is higher at the exposed C-terminal end and the GPS (arrowheads). The plot shows, however, no differences in solvent accessibility between the E2 GAIN domain variants. Shaded areas represent the 95% confidence interval.

(C) Confocal images of bioorthogonally marked E2 carrying a wild-type GAIN domain. The TA +3 and +6 labels are reduced compared with the +17 label.

(D) Quantification of surface delivery of the receptor (data are shown as means ± SEM [N = 3, n = 4]), and intensity of bioorthogonal labels within and C-terminal to the TA (data are shown as means ± SEM [N = 3, n = 20]).

(E) Ratio of TA label versus NTF quantities for all bioorthogonal labels in E2 demonstrate accessibility of the TA to the solvent in intact E2 heterodimers.

(F–H) Similar measurements as in (C)–(E) for an E2 full-length receptor with a GAIN domain carrying the human vibratory urticaria mutation C492Y. Datasets in (G) and (H) are normalized to the E2-TAG^{GPS+17}.

(I) p values of statistical comparison of surface expression and TA accessibility between E2 and E2^{C492Y} shows no increased TA accessibility because of the mutation.

Scale bars, 10 μm. See also Figures S6C and S7 and Table S3.

tags through co-expression with Pyl^{RS}/tRNA^{Pyl}, TCO*-lysine, and H-Tet-Cy5 (Figures 6C–6I). Although the VBU mutation slightly decreased surface expression of the E2^{C492Y}-TAG^{+6} variant, E2^{C492Y}-TAG^{+3} and E2^{C492Y}-TAG^{+17} were indistinguishable from their E2-TAG^{+X} counterparts (Figures 6D, 6E, 6G, and 6H). Interestingly, the TA of E2 was slightly, yet significantly, less solvent accessible when the GAIN domain carried the C492Y mutation (Figures 6D, 6G, and 6I). In addition, mechanical
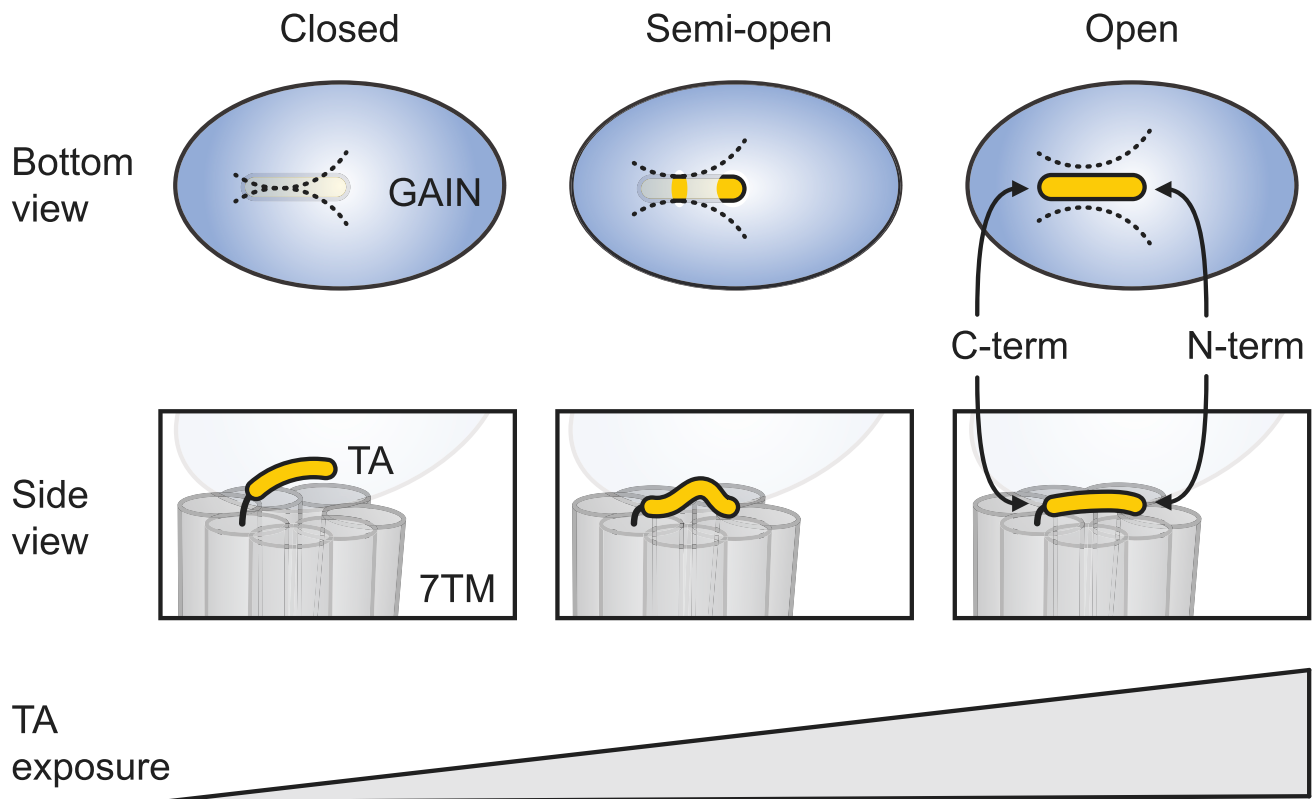
Figure 7. Model of the conformational flexibility and TA exposure in the intact GAIN domain

The GAIN domain of aGPCRs adopts conformations in which the TA is differentially exposed to the solvent and possibly also to the extracellular face of the 7TM domain through the opening of two flaps (dashed lines) in the surface of the intact, non-dissociated domain. Conformational transitions, e.g., through engagement with ligands and upon mechanical forces, gradually expose the TA and activate receptor signaling. The closed and semi-open states were observed in crystal structures, molecular models, and MD simulations. The open state, which would allow for engagement of the full TA sequence and 7TM domain, is speculative, but bioorthogonal labeling results indicate that lid opening can expose a larger TA segment or the entire TA.

stimulation through shaking of the cells before or during labeling did not affect TA accessibility (Figure S7), challenging the conclusion that the C492Y exchange in patients with VBU causes low-threshold mechanoactivation of E2 through GAIN domain destabilization and subsequent increased TA exposure (Boyden et al., 2016).

DISCUSSION

Thus far, the X-ray structures of the L1 and B3 GAIN domains implicate a full encasement of the TA inside the domain, which overtly renders the TA unavailable for interactions with the 7TM domain or other molecules (Araç et al., 2012). This framework has been at odds with a tier of results on aGPCR structure and activation. First, several aGPCRs, such as ADGRG5/Gpr114, are not autoproteolytically processed for lack of a consensus GPS tripeptide sequence yet can be activated through their TA (Wilde et al., 2016). The constitutive activity of cleavable aGPCRs (e.g., ADGRG6/Gpr126 and ADGRD1/Gpr133) is maintained even if GAIN domain-cleavage is suppressed (Liebscher et al., 2014). Further, our results provide an explanation for seemingly paradoxical findings, which show that aGPCR cleavage is

dispensable, whereas the correct sequence of the TA is absolutely necessary for receptor activity (Prömel et al., 2012; Scholz et al., 2017). Findings in *Caenorhabditis elegans* showed that cleavage-deficient variants of the aGPCR LAT-1 are able to fully complement a *lat-1* null mutation (Prömel et al., 2012). Similarly, a cleavage-deficient allele for the latrophilin homolog *dCirl* from *Drosophila melanogaster*, preventing GAIN domain-proteolysis by mutating the −2 residue of the GPS, shows a wild-type level of control over neuronal mechanosensitivity (Scholz et al., 2017). In contrast, a dCIRL mutant that disabled GAIN domain self-cleavage at the +1 site, and thus also the TA, completely abrogated receptor function. This implies that the TA, although not exposable any longer through simple NTF-CTF separation in accord with the dissociation model (Figure 1A), is accessible for receptor signaling in the intact GAIN domain of the unbroken receptor heterodimer.

Interestingly, the structure of the G1 GAIN domain (Salzman et al., 2016) contains a thus-far unrecognized opening, which provides accessibility of the TA to the solvent. In MD simulations, we observed that this TA exposure is a general structural feature achieved through the exceptional intrinsic flexibility of the GAIN domain, implicating a central mechanistic role in aGPCR

function (Figures 1A and 7). This observation was confirmed by subsequent bioorthogonal labeling of conserved positions within the TA of several aGPCRs, demonstrating that the TA also becomes accessible in expressed full-length receptors. Notably, the labeling experiments demonstrated that even larger portions of the TA can become solvent accessible. In the computational approach, in which the isolated GAIN domains were simulated, only the two N-terminal residues of the TA are exposed to the solvent, whereas the bioorthogonal labeling data of the expressed full-length receptors show that this opening can extend at least to the +6 residue of the TA (Figure 7). The additional exposure of the TA may be promoted through adequate stimuli of aGPCRs, such as engagement with ligands (Hamann et al., 2015), mechanical forces, or by interaction with the extracellular face of the 7TM domain in the full-length receptor (Petersen et al., 2015; Scholz et al., 2017; 2015; White et al., 2014). For the current set of simulations, we did not consider the application of stimuli because they have remained largely ill-defined to date.

Comparison with protease-activated receptors (PARs), class A GPCRs, which are also activated by exposure of a TA through proteolytic processing of their extracellular N-termini by enzymatic digestion, e.g., through thrombin (Vu et al., 1991), allows for speculations about the mechanism of aGPCR activation. The crystal structure of PAR1 shows that binding of an virtually irreversible antagonist, vorapaxar, to its 7TM domain possibly occurs at a superficial extracellular surface, which precludes TA binding. This indicates this patch as the orthosteric binding site of PAR1 (Zhang et al., 2012). The conformational flexibility of the GAIN domain of aGPCRs lends itself to a model in which its open conformation exposes the domain-bound TA to superficial interactions with the extracellular loops of the 7TM domain, similar to the PAR1 receptor (Figure 7). Because of the expected high-energy barrier, the complete opening, however, would require extensive structural rearrangements in the GAIN domain, effecting breaks in the network of non-covalent interactions that anchor the TA to the domain. This likely progresses through multiple sub-states and structural intermediates (Frauenfelder et al., 1991), which could gradually increase TA exposure and recognition by the 7TM domain and allow for a stepwise, graded transduction of the physiological signal. Ultimately, the fully exposed TA core region and additional C-terminal non-core residues that may confer specificity (Demberg et al., 2017) engage the elusive orthosteric binding pocket of the receptor, akin to the transition from pre- to a fully 7TM domain-bound TA state, as previously suggested (Demberg et al., 2015; Liebscher and Schöneberg, 2016).

Our analysis of TA exposure using complementary *in silico* and *in cellulo* techniques, indeed, suggests such a stepwise mechanism of TA recognition and binding. Rather than occurring abruptly, the MD simulations suggest that TAs are uncovered along their N→C axis, starting with the +1 residue. The polar residues (+1, +2, and +7) of the core region of the TA are exposed in the intact GAIN domain, whereas hydrophobic residues (+3 to +6) are initially shielded from the polar environment. We speculate that the polar residues form the initial recognition signature between TA and 7TM domain through long-range electrostatic steering and, eventually, by direct interaction (Schreiber, 2002) with the 7TM domain.

Through insertion and bioorthogonal labeling at the boundary residues of the hydrophobic +3-to-+6 stretch, we observed that even those are exposed to the solvent, indicating that this part of the TA core is gradually revealed to engage in a TA-7TM domain interaction (Figure 7). This work thus also demonstrates the utility of UAA incorporation through genetic code expansion and click labeling for imaging studies on aGPCRs as similarly recently shown for class A and B1 GPCRs (Serfling et al., 2019). Furthermore, this technique has proven particularly useful for obtaining measurements on the accessibility of intra-protein sites in spatially confined folds, such as the TA in the GAIN domain, which have thus far remained refractory to the placement of larger marker moieties, such as biochemical tags or genetically encoded chromophores.

Our work also paves the way for future studies that establish a direct link between flap dynamics and aGPCR activation. The MD simulations provide a solid rationale for experimentally perturbing TA accessibility via mutation of the flap regions. Specifically, we envisage that mutations that either alter the hydrogen-bonding network (e.g., for rat L1 D653A, E774A, Y806A, or E808A) or the electrostatics (D653K, E774R, E808R, or R777E) of the L1 flaps could serve as the starting point for these experiments. Alternatively, cross-linking experiments aimed at reducing the flexibility of the flaps, for instance, by adding a disulfide bridge between helix 6 and flap 1 (e.g., E774C-D653C) or flap 2 (Y806C/E808C-D653C), should shed light on whether TA accessibility via flaps/openings have a relevant role in aGPCR activation.

In summary, here, we provide evidence for a hitherto unknown general mechanism of spontaneous TA exposure to the 7TM domain of the receptor, which likely governs downstream signaling. Pharmacological targeting of GAIN domain conformations or their recognition sites in the 7TM domain may thus represent an unconventional strategy to control TA-dependent aGPCR signals and aid in the development of compounds to combat diseases that arise from dysfunctional aGPCR signals. Future investigations, including atomistic co-structures of the GAIN-7TM domain pair of aGPCRs, are required to further detail their steric relationship and delineate interaction contacts to instruct suitable targeting strategies. Thus far, however, the results in this work indicate that such an interface arises through the unexpected structural flexibility of the GAIN domain.

Limitations

aGPCR research faces conceptual and technical challenges because of the unusual heterodimeric structure, autoproteolytic activity, and versatile signaling behavior of aGPCRs. Here, we have used MD simulations, GCE-mediated bioorthogonal click labeling, and intra-receptor FRET measurements as new technical approaches to bypass those obstacles. Although the combination of all three approaches provides novel insights into the unexpected accessibility of the TA of aGPCRs, each carries intrinsic limitations.

MD simulations in this work are restricted to the GAIN domain because no structural information exists for the 7TM domain of any of the investigated aGPCRs. In addition, the inclusion of isolated 7TM structures would present modeling challenges in its own right, without necessarily increasing the reliability of the

results. Beyond that restriction, the general limitations of the MD approach apply: limited sampling of conformational space and possible systematic errors in the simulation parameters. However, these shortcomings do not significantly affect our results because full exploration of conformational space or precise structural and/or kinetic predictions were not the focus of this work. Rather, we exploited the time-resolved nature of the MD trajectories to establish a connection between the motion of two flexible segments of the GAIN domain (flap 1 and flap 2) and the solvent accessibility of the GPS and TA. Although the listed constraints of the MD approach may propagate to the reported numerical values (i.e., slightly different SASA, RMSF, or correlation values), it is highly unlikely that those errors invalidate the established connection between flaps 1 and 2 and the GPS or invert the exposure trends observed in our data.

A caveat of the GCE approach is its incomplete incorporation efficiency, even when optimized GCE systems are used (Serfling et al., 2018). Therefore, in most studies, strong overexpression systems are used, limiting amber-suppression experiments to cell lines with high protein expression. Additionally, considerate choice of UAA placement in the target protein and gauging of its experimental consequences in advance is important. Specifically for transmembrane proteins, such as aGPCRs, the repurposed *TAG* codon for a subsequent UAA insertion should be placed N-terminal to the first transmembrane helix (as applied to all expression construct designs in this study) because inefficient UAA incorporation will result in fragmentation of the target protein upon biosynthesis (Figures S4H and S4I). This will generate a mixed population of full-length and truncated proteins expressed in the same cell, which may affect quantitative assessment of protein abundance and molecular activity of the target (e.g., through immunolabeling, flow cytometry, ELISA, or pharmacological signaling assays). A further complication arises from the spatial properties of the UAA, e.g., TCO*lysine, and the bioorthogonal click labels, whose molecular size may affect local protein structure. Depending on the site of manipulation of the target protein, some of its functional characteristics may be affected by the UAA introduction (e.g., suppressed autoproteolytic activity in E5-TAG^{GPS+3} and E5-TAG^{GPS+6} receptors; Figure S5I). This requires heedful *a priori* assessment of the structural effects on the target protein, e.g., by modeling the insertion effects (Figures 1D, S3, and S5B) or surveying unimpeded recognition by antisera (Figures 2C–2E, S4D, S4E, S5F, and S5G). Label-free techniques may offer alternative approaches to confirm results obtained through bioorthogonal click labels but may require substantially more complex experimental systems.

Quantification of FRET via acceptor photobleaching is regularly used to analyze inter- and intramolecular protein interactions. However, control experiments are critical to determine non-specific FRET between donor and acceptor dye molecules. In particular, FRET effects resulting from non-specific protein-protein interactions are difficult to identify. Labeling of partner fragments within the aGPCR with a suitable FRET pair (separately and closely attached to NTF and CTF) allows for direct assessment of aGPCR heterodimer formation via *intra*-FRET signals (Figures 3E and 3H), whereas labeling of two individual receptor constructs bearing either one of the fluorophores enables

the recording of *inter*-FRET events (Figure 3K). Here, inter-FRET efficiencies can be prone to FRET resulting from non-specific protein-protein interactions (e.g., due to protein mobility in the plasma membrane) or the physiologically relevant interplay of the protein partners (e.g., the formation of multimers). Ideally, another donor-/acceptor-labeled biomolecule, which is different from the target (e.g., another non-E5 adhesion GPCR, a non-adhesion-GPCR, or a non-GPCR surface molecule), can be employed to determine the contribution of non-specific FRET to the measured values. However, because differences in surface expression, relative dye location within the protein structure, and fluorophore orientation to the FRET-partner fluorophore may render the faithful interpretation of control FRET results futile, a control biomolecule needs diligent development and assessment before use.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILIBILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availibility
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Cell culture
- METHOD DETAILS
  - ○ Molecular biology
  - ○ Transfection
  - ○ Surface ELISA
  - ○ Western blot
  - ○ Bioorthogonal labeling and immunolabeling
  - ○ Labeling of α-bungarotoxin binding site
  - ○ Cell vibration
  - ○ Confocal microscopy
  - ○ *d*STORM imaging
  - ○ Molecular modeling
  - ○ Molecular dynamics simulations
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Quantification of imaging data
  - ○ FRET experiments
  - ○ Statistical methods

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.12.042.

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25.

Araç, D., Boucard, A.A., Bolliger, M.F., Nguyen, J., Soltis, S.M., Südhof, T.C., and Brunger, A.T. (2012). A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. EMBO J. 31, 1364–1378.

Aust, G., Zhu, D., Van Meir, E.G., and Xu, L. (2016). Adhesion GPCRs in tumorigenesis. Handb. Exp. Pharmacol. 234, 369–396.

Beliu, G., Kurz, A.J., Kuhlemann, A.C., Behringer-Pliess, L., Meub, M., Wolf, N., Seibel, J., Shi, Z.-D., Schnermann, M., Grimm, J.B., et al. (2019). Bioorthogonal labeling with tetrazine-dyes for super-resolution microscopy. Commun. Biol. 2, 261.

Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343–350.

Bohnekamp, J., and Schöneberg, T. (2011). Cell adhesion receptor GPR133 couples to Gs protein. J Biol Chem 286, 41912–41916.

Boyden, S.E., Desai, A., Cruse, G., Young, M.L., Bolan, H.C., Scott, L.M., Eisch, A.R., Long, R.D., Lee, C.-C.R., Satorius, C.L., et al. (2016). Vibratory urticaria associated with a missense variant in ADGRE2. N. Engl. J. Med. 374, 656–663.

Brown, K., Filuta, A., Ludwig, M.-G., Seuwen, K., Jaros, J., Vidal, S., Arora, K., Naren, A.P., Kandasamy, K., Parthasarathi, K., et al. (2017). Epithelial Gpr116 regulates pulmonary alveolar homeostasis via Gq/11 signaling. JCI Insight 2, 18.

Changeux, J.-P., and Edelstein, S. (2011). Conformational selection or induced fit? 50 years of debate resolved. F1000 Biol. Rep. 3, 19.

Demberg, L.M., Rothemund, S., Schöneberg, T., and Liebscher, I. (2015). Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor GPR64/ADGRG2. Biochem. Biophys. Res. Commun. 464, 743–747.

Demberg, L.M., Winkler, J., Wilde, C., Simon, K.-U., Schön, J., Rothemund, S., Schöneberg, T., Prömel, S., and Liebscher, I. (2017). Activation of adhesion G protein-coupled receptors: agonist specificity of Stachel sequence-derived peptides. J. Biol. Chem. 292, 4383–4394.

Eichler, W., Aust, G., and Hamann, D. (1994). Characterization of an early activation-dependent antigen on lymphocytes defined by the monoclonal antibody BL-Ac(F2). Scand. J. Immunol. 39, 111–115.

Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. Science 254, 1598–1603.

Gray, J.X., Haino, M., Roth, M.J., Maguire, J.E., Jensen, P.N., Yarme, A., Stetler-Stevenson, M.A., Siebenlist, U., and Kelly, K. (1996). CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. J. Immunol. 157, 5438–5447.

Hamann, J., Vogel, B., van Schijndel, G.M., and van Lier, R.A. (1996). The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF). J. Exp. Med. 184, 1185–1189.

Hamann, J., Aust, G., Araç, D., Engel, F.B., Formstone, C., Fredriksson, R., Hall, R.A., Harty, B.L., Kirchhoff, C., Knapp, B., et al. (2015). International union of basic and clinical pharmacology, XCIV: adhesion G protein-coupled receptors. Pharmacol. Rev. 67, 338–367.

Harel, M., Kasher, R., Nicolas, A., Guss, J.M., Balass, M., Fridkin, M., Smit, A.B., Brejc, K., Sixma, T.K., Katchalski-Katzir, E., et al. (2001). The binding site of acetylcholine receptor as visualized in the x-ray structure of a complex between alpha-bungarotoxin and a mimotope peptide. Neuron 32, 265–275.

Heilemann, M., van de Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., and Sauer, M. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. Angew. Chem. Int. Ed. Engl. 47, 6172–6176.

Hilbig, D., Sittig, D., Hoffmann, F., Rothemund, S., Warmt, E., Quaas, M., Stürmer, J., Seiler, L., Liebscher, I., Hoang, N.A., et al. (2018). Mechano-dependent phosphorylation of the PDZ-binding motif of CD97/ADGRE5 modulates cellular detachment. Cell Rep. 24, 1986–1995.

Huang, J., and MacKerell, A.D., Jr. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J. Comput. Chem. 34, 2135–2145.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol, Graph. 14, 33–38, 27–28.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95.

Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. J. Comput. Chem. 29, 1859–1865.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202.

Kan, Z., Jaiswal, B.S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H.M., Yue, P., Haverty, P.M., Bourgon, R., Zheng, J., et al. (2010). Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 466, 869–873.

Kishore, A., Purcell, R.H., Nassiri-Toosi, Z., and Hall, R.A. (2016). Stalk-dependent and Stalk-independent signaling by the adhesion G protein-coupled receptors GPR56 (ADGRG1) and BAI1 (ADGRB1). J. Biol. Chem. 291, 3385–3394.

Krasnoperov, V.G., Bittner, M.A., Beavis, R., Kuang, Y., Salnikow, K.V., Chepurny, O.G., Little, A.R., Plotnikov, A.N., Wu, D., Holz, R.W., and

Petrenko, A.G. (1997). αa-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. Neuron 18, 925–937.

Kwakkenbos, M.J., van Lier, R.A.W., and Hamann, J. (2000). Characterization of EGF-TM7 family members by novel monoclonal antibodies. In Leucocyte Typing VII, D. Mason, ed. (Oxford University Press), pp. 381–383.

Langenhan, T. (2020). Adhesion G protein-coupled receptors-candidate metabotropic mechanosensors and novel drug targets. Basic Clin. Pharmacol. Toxicol. 126 (Suppl 6), 5–16.

Leon, K., Cunningham, R.L., Riback, J.A., Feldman, E., Li, J., Sosnick, T.R., Zhao, M., Monk, K.R., and Araç, D. (2020). Structural basis for adhesion G protein-coupled receptor Gpr126 function. Nat. Commun. 11, 194.

Liebscher, I., and Schöneberg, T. (2016). Tethered agonism: a common activation mechanism of adhesion GPCRs. Handb. Exp. Pharmacol. 234, 111–125.

Liebscher, I., Schön, J., Petersen, S.C., Fischer, L., Auerbach, N., Demberg, L.M., Mogha, A., Cöster, M., Simon, K.-U., Rothemund, S., et al. (2014). A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. Cell Rep. 9, 2018–2026.

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

Lin, H.-H., Hsiao, C.-C., Pabst, C., Hébert, J., Schöneberg, T., and Hamann, J. (2017). Adhesion GPCRs in Regulating Immune Responses and Inflammation. Adv. Immunol. 136, 163–201.

McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.-P., Lane, T.J., and Pande, V.S. (2015). MDTraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys. J. 109, 1528–1532.

Müller, A., Winkler, J., Fiedler, F., Sastradihardja, T., Binder, C., Schnabel, R., Kungel, J., Rothemund, S., Hennig, C., Schöneberg, T., and Prömel, S. (2015). Oriented cell division in the C. elegans embryo is coordinated by G-protein signaling dependent on the adhesion GPCR LAT-1. PLoS Genet. 11, e1005624.

Nikić, I., Kang, J.H., Girona, G.E., Aramburu, I.V., and Lemke, E.A. (2015). Labeling proteins on live mammalian cells using click chemistry. Nat. Protoc. 10, 780–791.

Nikić, I., Estrada Girona, G., Kang, J.H., Paci, G., Mikhaleva, S., Koehler, C., Shymanska, N.V., Ventura Santos, C., Spitz, D., and Lemke, E.A. (2016). Debugging eukaryotic genetic code expansion for site-specific click-PAINT super-resolution microscopy. Angew. Chem. Int. Ed. Engl. 55, 16172–16176.

O'Hayre, M., Vázquez-Prado, J., Kufareva, I., Stawiski, E.W., Handel, T.M., Seshagiri, S., and Gutkind, J.S. (2013). The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. Nat. Rev. Cancer 13, 412–424.

Petersen, S.C., Luo, R., Liebscher, I., Giera, S., Jeong, S.-J., Mogha, A., Ghidinelli, M., Feltri, M.L., Schöneberg, T., Piao, X., and Monk, K.R. (2015). The adhesion GPCR GPR126 has distinct, domain-dependent functions in Schwann cell development mediated by interaction with laminin-211. Neuron 85, 755–769.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Prömel, S., Frickenhaus, M., Hughes, S., Mestek, L., Staunton, D., Woollard, A., Vakonakis, I., Schöneberg, T., Schnabel, R., Russ, A.P., and Langenhan, T. (2012). The GPS motif is a molecular switch for bimodal activities of adhesion class G protein-coupled receptors. Cell Rep. 2, 321–331.

Purcell, R.H., and Hall, R.A. (2018). Adhesion G protein-coupled receptors as drug targets. Annu. Rev. Pharmacol. Toxicol. 58, 429–449.

Ryu, Y., and Schultz, P.G. (2006). Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. Nat. Methods 3, 263–265.

Salzman, G.S., Ackerman, S.D., Ding, C., Koide, A., Leon, K., Luo, R., Stoveken, H.M., Fernandez, C.G., Tall, G.G., Piao, X., et al. (2016). Structural basis for regulation of GPR56/ADGRG1 by its alternatively spliced extracellular domains. Neuron 91, 1292–1304.

Salzman, G.S., Zhang, S., Gupta, A., Koide, A., Koide, S., and Araç, D. (2017). Stachel-independent modulation of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. Proc. Natl. Acad. Sci. USA 114, 10095–10100.

Sando, R., Jiang, X., and Südhof, T.C. (2019). Latrophilin GPCRs direct synapse specificity by coincident binding of FLRTs and teneurins. Science 363, eaav7969.

Schmied, W.H., Elsässer, S.J., Uttamapinant, C., and Chin, J.W. (2014). Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. J. Am. Chem. Soc. 136, 15577–15583.

Scholz, N. (2018). Cancer cell mechanics: adhesion G protein-coupled receptors in action? Front. Oncol. 8, 59.

Scholz, N., Gehring, J., Guan, C., Ljaschenko, D., Fischer, R., Lakshmanan, V., Kittel, R.J., and Langenhan, T. (2015). The adhesion GPCR latrophilin/CIRL shapes mechanosensation. Cell Rep. 11, 866–874.

Scholz, N., Monk, K.R., Kittel, R.J., and Langenhan, T. (2016). Adhesion GPCRs as a putative class of metabotropic mechanosensors. Handb. Exp. Pharmacol. 234, 221–247.

Scholz, N., Guan, C., Nieberler, M., Grotemeyer, A., Maiellaro, I., Gao, S., Beck, S., Pawlak, M., Sauer, M., Asan, E., et al. (2017). Mechano-dependent signaling by Latrophilin/CIRL quenches cAMP in proprioceptive neurons. eLife 6, 1364.

Schreiber, G. (2002). Kinetic studies of protein-protein interactions. Curr. Opin. Struct. Biol. 12, 41–47.

Serfling, R., Lorenz, C., Etzel, M., Schicht, G., Böttke, T., Mörl, M., and Coin, I. (2018). Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells. Nucleic Acids Res. 46, 1–10.

Serfling, R., Seidel, L., Bock, A., Lohse, M.J., Annibale, P., and Coin, I. (2019). Quantitative single-residue bioorthogonal labeling of G protein-coupled receptors in live cells. ACS Chem. Biol. 14, 1141–1149.

Shapovalov, M.V., and Dunbrack, R.L., Jr. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19, 844–858.

Stoveken, H.M., Hajduczok, A.G., Xu, L., and Tall, G.G. (2015). Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. Proc. Natl. Acad. Sci. USA 112, 6194–6199.

Sun, X., Singh, S., Blumer, K.J., and Bowman, G.R. (2018). Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. eLife 7, 19.

Tiemann, J.K.S., Guixà-González, R., Hildebrand, P.W., and Rose, A.S. (2017). MDsrv: viewing and sharing molecular dynamics simulations on the web. Nat. Methods 14, 1123–1124.

van de Linde, S., Löschberger, A., Klein, T., Heidbreder, M., Wolter, S., Heilemann, M., and Sauer, M. (2011). Direct stochastic optical reconstruction microscopy with standard fluorescent probes. Nat. Protoc. 6, 991–1009.

Vizurraga, A., Adhikari, R., Yeung, J., Yu, M., and Tall, G.G. (2020). Mechanisms of adhesion G protein-coupled receptor activation. J. Biol. Chem. 295, 14065–14083.

Vu, T.K., Hung, D.T., Wheaton, V.I., and Coughlin, S.R. (1991). Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. Cell 64, 1057–1068.

Wang, T., Ward, Y., Tian, L., Lake, R., Guedez, L., Stetler-Stevenson, W.G., and Kelly, K. (2005). CD97, an adhesion receptor on inflammatory cells, stimulates angiogenesis through binding integrin counterreceptors on endothelial cells. Blood 105, 2836–2844.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46 (W1), W296–W303.

White, J.P., Wrann, C.D., Rao, R.R., Nair, S.K., Jedrychowski, M.P., You, J.-S., Martínez-Redondo, V., Gygi, S.P., Ruas, J.L., Hornberger, T.A., et al. (2014). G protein-coupled receptor 56 regulates mechanical overload-induced muscle hypertrophy. Proc. Natl. Acad. Sci. USA *111*, 15756–15761.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).

Wilde, C., Fischer, L., Lede, V., Kirchberger, J., Rothemund, S., Schöneberg, T., and Liebscher, I. (2016). The constitutive activity of the adhesion GPCR GPR114/ADGRG5 is mediated by its tethered agonist. FASEB J. *30*, 666–673.

Wolter, S., Löschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M.-C., van de Linde, S., and Sauer, M. (2012). rapidSTORM: accurate, fast open-source software for localization microscopy. Nat. Methods *9*, 1040–1041.

Yin, Y., Xu, X., Tang, J., Zhang, W., Zhangyuan, G., Ji, J., Deng, L., Lu, S., Zhuo, H., and Sun, B. (2018). CD97 Promotes tumor aggressiveness through the traditional G protein-coupled receptor-mediated signaling in hepatocellular carcinoma. Hepatology *68*, 1865–1878.

Zhang, L., and Hermans, J. (1996). Hydrophilicity of cavities in proteins. Proteins *24*, 433–438.

Zhang, C., Srinivasan, Y., Arlow, D.H., Fung, J.J., Palmer, D., Zheng, Y., Green, H.F., Pandey, A., Dror, R.O., Shaw, D.E., et al. (2012). High-resolution crystal structure of human protease-activated receptor 1. Nature *492*, 387–392.

Zhu, B., Luo, R., Jin, P., Li, T., Oak, H.C., Giera, S., Monk, K.R., Lak, P., Shoichet, B.K., and Piao, X. (2019). GAIN domain-mediated cleavage is required for activation of G protein-coupled receptor 56 (GPR56) by its natural ligands and a small-molecule agonist. J. Biol. Chem. *294*, 19246–19254.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
α-CD97/1 (α-E5^{EGF1})	Eichler et al., 1994	N/A
α-CD97/3 (α-E5^{GAIN})	Kwakkenbos et al., 2000	N/A
Mouse monoclonal anti-HA Tag	ThermoFischer Scientific	Cat# 26183-A488; RRID:AB_2610624
Mouse monoclonal tubulin beta antibody	DSHB	Cat# E7;, RRID: AB_528499
Rabbit polyclonal anti-ADGRE5	Sigma-Aldrich	Cat# HPA013707; RRID:AB_1846345
Goat anti-Rabbit IgG	ThermoFischer Scientific	Cat# A-11008; RRID:AB_143165
Goat anti-Mouse IgG (WB)	Li-Cor	Cat# 926-32210
Goat anti-Rabbit IgG (WB)	Li-Cor	Cat# 926-68071
Bacterial and virus strains		
E. coli XL1-Blue	New England Biolabs	https://international.neb.com/
Chemicals, peptides, and recombinant proteins		
jetPRIME transfection reagent	Polyplus	Cat# 114-01
Lipofectamine 2000 transfection reagent	Invitrogen	Cat# 11668019
Click Amino Acid / trans-Cyclooct-2-en – L - Lysine (axial isomer)	SiChem	Cat# SC-8008
Protease Inhibitor Cocktail	Sigma-Aldrich	Cat# P8340
α-Bungarotoxin, Alexa Fluor 488 conjugate	ThermoFischer Scientific	Cat# B13422
α-Bungarotoxin, Alexa Fluor 594 conjugate	ThermoFischer Scientific	Cat# B13423
Deposited data		
Static models and dynamics structures of MD data	This paper	https://zenodo.org/record/4114651
Imaging and WB datasets	This paper	https://dx.doi.org/10.17632/5fpk9g52rw.1
Experimental models: cell lines		
Human: HEK293T	German Collection of Microorganisms and Cell Culture (Braunschweig, Germany)	#ACC635
Monkey: COS-7	Cell Lines Service GmbH (Eppelheim, Germany)	#605470
Oligonucleotides		
Primer for cloning and sequence verification, see Table S2	This paper	N/A
Recombinant DNA		
Genetic constructs and Plasmids, see Table S1	This paper	N/A
Software and algorithms		
rapidSTORM	N/A	Wolter et al., 2012
LAS X (Leica)	Leica Microsystems, Germany	RRID:SCR_013673
MarvinSketch v18.16	ChemAxon, Hungary	https://www.chemaxon.com
DOWSER	N/A	Zhang and Hermans, 1996
OriginPro 2016G	OriginLab, Northampton, MA)	https://www.originlab.com
ZENblack	Zeiss, Germany	N/A
Prism 7	GraphPad	RRID:SCR_002798

RESOURCE AVAILIBILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Tobias Langenhan (tobias.langenhan@gmail.com).

Materials availability

All unique reagents generated in this study are available from the lead contact without restriction.

Data and code availibility

Links to sessions of representative molecular simulations discussed in this manuscript are available at Zenodo: https://zenodo.org/record/4114651. Confocal imaging, *d*STORM imaging, and western blot raw datasets are deposited at Mendeley Data: https://dx.doi.org/10.17632/5fpk9g52rw.1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

HEK293T (RRID:CVCL_0063) cells were maintained in T25-culture flasks (Thermo Scientific, #156340) or Petri dishes (Greiner Bio-One, #664160) with cell growth medium consisting of DMEM (Sigma-Aldrich, #D5796 or GIBCO, #11995065) containing 10% (v/v) FCS or FBS (Sigma-Aldrich, #F7524 or GIBCO, #10500) and 1% (v/v) Penicillin-Streptomycin (Sigma-Aldrich, #P4333 or GIBCO, #15140122) in a 5% CO_2 atmosphere at $37^\circ C$. COS7 cells (RRID:CVCL_0224) were maintained in DMEM (Sigma-Aldrich, #D8062) with 10% (v/v) FCS and 1% (v/v) Penicillin-Streptomycin in a 5% CO_2 atmosphere at $37^\circ C$. Prior to seeding the dishes were coated with 0.01% (w/v) poly-D-Lysine (Sigma-Aldrich, #P6407) or 0.01% (w/v) poly-L-Lysine (Sigma-Aldrich, #P9404) for 30-60 min at room temperature (RT). For imaging experiments cells were seeded at least 16 h before transfection on 4-well chambered Lab-Tek II chambered cover glass (Thermo Scientific, #155409) or μ-slide 8 well (ibidi, #80826). For surface ELISA HEK293T cells were seeded on 96-well plate (Greiner Bio-One, #655090 or #655098). For western blot analysis cells were seeded on 6-well plates (Greiner Bio-One, #657160).

METHOD DETAILS

Molecular biology

The constructs and cloning procedures of all plasmids used in this study (B3, E2, E5 and L1) are shown in Table S1, primers used for their construction and confirmation are listed in Table S2. For expression of the PylRS/tRNAPyl-pair in mammalian cells we used the pcDNA3.1-tRNAPyl/NESPylRSAF plasmid (for HEK293T cells) (Nikić et al., 2016) or pcDNA3.1-*Mb*PylRSF/tRNAM15 (for HEK293T and COS7 cells) (Serfling et al., 2018), which were gifts from Edward Lemke and Irene Coin, respectively. Plasmid amplification was performed via transformation in *E. coli* (XL1-Blue) and DNA isolation via MIDI-prep (NucleoBond® Xtra Midi, Macherey & Nagel, #740410), CompactPrep Plasmid Midi Kit (QIAGEN, #12843) or QIAprep Spin Miniprep Kit (QIAGEN, #27104).

Transfection

Cells were transfected at 60 – 80% confluency with the transfection reagent JetPrime (Polyplus, #114-01) or Lipofectamine 2000 (Invitrogen, #11668019) with the suitable 1:1 plasmid/reagent mixture (according to manufacturer's recommendations) and incubated for at least 24 h in a 5% CO_2 atmosphere at $37^\circ C$. The plasmids encoding for constructs with an amber stop codon were mixed in a 1:1 ratio with pcDNA3.1-tRNAPyl/NESPylRSAF plasmid or pcDNA3.1-*Mb*PylRSF/tRNAM15 plasmid. In addition, the UAA TCO*A (SiChem, #SC-8008) was supplemented in the medium at a final concentration of 250 μM, diluted 1:4 with 1 M HEPES. After 4-8 h the medium was exchanged to fresh cell growth medium. The cells were incubated ∼12-36 h before labeling and fixation.

Surface ELISA

Transfected HEK293T cells were fixed with 60 μL of 4% (w/v) paraformaldehyde (PFA) at RT for 10 min. Next, cells were blocked with 1x PBS containing 5% (v/v) goat serum (Sigma-Aldrich, #G6767) at RT for 30 min. Then, cells were incubated for 1 h at RT with 1:1000 dilution of α-HA-Peroxidase (RRID:AB_390917) in 1x PBS containing 5% (v/v) goat serum. After at least three washing steps using 1x PBS, the cells were incubated with a substrate solution (1 mg/mL o-phenylenediamine and 1 μL/mL hydrogen peroxide solved in a 0.05 M citric acid and 0.05 M disodium phosphate solution [pH 5]) at RT. The reaction was stopped within 10 min using 2.5 M sulfuric acid. The absorbance of the supernatants was measured at 490 nm with a multimode-reader (Promega, GloMax-Multi+ or Molecular Devices, SpectraMax M).

Western blot

HEK293T cells were chemically lysed with 200 μL M-PER buffer (Thermo Scientific, #78503) containing 1:100 protease inhibitor (Sigma-Aldrich, #P8340) on an orbital shaker for 5 min at $4^\circ C$. The lysate was collected and centrifuged at 15.200 xg for 5 min at

4°C. Supernatants were diluted 4:1 with 4x protein loading buffer (Li-Cor, #928-40004) containing 10% (v/v) β-mercaptoethanol. Samples were loaded on a Novex 4%–12% Tris-Glycine Mini Gel (Invitrogen) or a Novex 8%–16% Tris-Glycine Mini Gel (Invitrogen), which were applied in a Mini Gel Tank (Invitrogen). Electrophoresis was performed in running buffer (25 mM Tris, 190 mM glycine, 0.1% (w/v) SDS) at RT. Protein bands were transferred from the gel on a nitrocellulose membrane with 0.2 μm pore size (Invitrogen). Transfer was performed in an iBlot 2 Dry Blotting System (Invitrogen). After the transfer, the membrane was incubated with a blocking solution on a shaker for 1 h at RT. Blocking solution consisted of a 1:2 dilution of blocking buffer (Li-Cor, #927-40000) and 1x PBS. Next, the membrane was treated with 1:1000 rabbit-α-HA (RRID:AB_1549585), 1:1000 mouse-α-V5 (RRID:AB_2792973) and 1:5000 mouse-α-E7 (RRID:AB_2315513) diluted in blocking solution containing 0.1% (v/v) Tween-20 on a shaker overnight at 4°C. Following at least three washing steps with 0.1% PBS-T, the membrane was incubated with 1:15000 IRDye 680RD Goat anti-Rabbit IgG (Li-Cor, #926-68071) and 1:15000 IRDye 800CW Goat anti-Mouse IgG (Li-Cor, #926-32210) diluted in blocking solution containing 0.1% (v/v) Tween-20 on a shaker for 1 h at RT. After the membrane was washed several times with 0.1% PBST, the protein bands were detected by the excitation at 685 nm and 785 nm using a solid-state diode laser of the Odyssey Fc Imaging System (Li-Cor).

Bioorthogonal labeling and immunolabeling
For single bioorthogonal labeling of UAA, the transfected HEK293T or COS7 cells were labeled with 1.5 μM tetrazine dye derivative H-Tet-Cy5 (Jena Bioscience, #CLK-015-05) in cell growth medium for 10-30 min at 37°C and 5% CO_2. After labeling, the cells were rinsed once with cell growth medium and fixed at RT with 4% (v/v) formaldehyde (containing 0.25% (v/v) glutaraldehyde) or 4% (w/v) PFA for 10-15 min. The fixed cells were washed three times with 1x PBS before imaging. α-CD97/1 (α-E5^{EGF1}) (Eichler et al., 1994) and α-CD97/3 (α-E5^{GAIN}) (Kwakkenbos et al., 2000) were fluorescently labeled following the supplier's protocol with Alexa Fluor 532 (Thermo Scientific, #A20101MP) or Alexa Fluor 647 (Thermo Scientific, #A37566), respectively. Briefly, 50 μg of the purified antibodies were incubated in a 5 M excess with the desired NHS-dye in 100 mM $NaHCO_3$ at RT for 3 h in the dark. Antibody conjugates were purified on gel filtration columns (GE Healthcare, NAP-5, Sephadex G25 DNA Grade). Finally, the degree of labeling of the purified antibody was determined by a UV/VIS spectrophotometer (Jasco, V-650) to DOL = ∼1 (for α-CD97/1- Alexa Fluor 532) and DOL = ∼5 (for α-CD97/3- Alexa Fluor 647; antibody working concentration: 10 μg/ml). The conjugated antibodies were stored at 4°C. Dual-labeling of UAA and tag was achieved through a co-incubation of H-Tet-Cy5 or Me-Tet-ATTO 488 (generous gift from ATTO-TEC) and antibody (α-HA- Alexa Fluor 488 1:500 (RRID:AB_2610624), α-CD97/1- Alexa Fluor 532 or α-CD97/3-Alexa647) for 30 min at 4°C. After labeling, the cells were rinsed once with cell growth medium and fixed at RT with 4% (v/v) formaldehyde (containing 0.25% (v/v) glutaraldehyde) for 10-15 min. The fixed cells were washed three times with 1x PBS before imaging. The GAIN domain of E5 was labeled using 1:200 α-E5-GAIN (RRID:AB_1846345) for 30 min at 37°C. After labeling, the cells were rinsed once with cell growth medium and fixed at RT with 4% (w/v) PFA for 10 min. Fixed cells were blocked with 1x DPBS (GIBCO, #14190144) containing 2% (w/v) BSA (Sigma-Aldrich, #126593) at RT for 1 h and labeled with 1:500 Alexa Fluor 488 conjugated goat α-rabbit IgG (RRID:AB_143165) over night at 4°C. Cells were washed three times with 1x DPBS before imaging.

Labeling of α-bungarotoxin binding site
HEK293T cells were incubated with 1:500 α-BuTX- Alexa Fluor 488 (Invitrogen, #B13422) or α-BuTX- Alexa Fluor 594 (Invitrogen, #B13423) in cell growth medium for 30 min at 37°C and 5% CO_2. After one washing step using cell growth medium, the cells were fixed at RT with 4% (w/v) PFA for 10 min before imaging.

Cell vibration
HEK293T cells were seeded on a μ-slide 8 well, which was coated with 0.01% (w/v) chondroitin sulfate B (Sigma-Aldrich, #C3788) for 60 min at RT. Transfected HEK293T cells were labeled with 1.5 μM H-Tet-Cy5 in cell growth medium for 20 min at 37°C on a orbital shaker at 750 rpm. Control cells were labeled without shaking. After labeling, the cells were washed once with cell growth medium and fixed at RT with 4% (w/v) PFA for 10 min before imaging.

Confocal microscopy
Confocal images were obtained with a LSM700 setup (63x/1.4 oil objective, Zeiss) or SP8 setup (63x/1.3 glycerol objective, Leica) setup. Single plane images on the LSM700 were acquired with suitable settings for the respective dye (AlexaFluor 488, ATTO 488, Alexa Fluor 532: excitation laser with 488 nm; Alexa Fluor 647, Cy5: excitation laser with 640 nm). Same imaging settings were chosen when different constructs were compared. Images were adjusted for brightness and contrast. Single plane images on the SP8 were acquired with suitable settings for the respective dye (Alexa Fluor 488: excitation laser with 488 nm; Cy5: excitation laser with 633 nm). Same imaging settings were chosen when receptor variants were compared.

dSTORM imaging
dSTORM images were acquired using an inverted wide-field fluorescence microscope (IX-71, Olympus). For excitation of Cy5, a 641-nm diode laser (Cube 640-100C, Coherent) which was spectrally cleaned by a clean-up filter (Laser Clean-up filter 640/10, Chroma) was used. The laser beam was focused onto the back focal plane of the oil-immersion objective (60x/1.45 oil objective, Olympus). Emission light was separated from the illumination light using a dichroic mirror (635rpc, Chroma) spectrally filtered by a bandpass filter (Em01-R442/514/647-25, Semrock) and projected onto an electron multiplying CCD camera chip (Ixon DU 897,

512 × 512 pixel a 16 μm, Andor). Placing additional lenses in the detection path, a final pixel size of 128 nm was generated. For each dSTORM measurement, at least 15.000 images with an exposure time of 20 ms and irradiation intensities ∼5 kW/cm² were recorded. Basal membranes were imaged by internal reflection fluorescence (TIRF) illumination. Experiments were performed in PBS-based photoswitching buffer containing 100 mM β-mercaptoethylamine (Sigma-Aldrich, #641022) and an oxygen scavenger system (2% (w/v) glucose, 4 U/ml glucose oxidase and 80 U/ml catalase) adjusted to pH 7.4 (van de Linde et al., 2011). dSTORM image reconstruction: Open source software for single-molecule localizations and super-resolution image reconstruction, rapidSTORM3.3 (Wolter et al., 2012).

Molecular modeling

E5 GAIN domain homology modeling was performed using the SWISS-MODEL server (Waterhouse et al., 2018). The search for homolog protein templates yielded three top-ranked crystal structures, namely Latrophilin-1 (PDB:4DLQ), brain-specific angiogenesis inhibitor 3 (PDB:4DLO), and GPR56/ADGRG1 (PDB:5KVM). Homology models using each one of the templates were built and evaluated using different geometrical properties of the model (Benkert et al., 2011). Based on this evaluation, the brain-specific angiogenesis inhibitor 3 was picked as a template to generate the final model. Chemical structures of TCO and H-Tet-Cy5 were drawn and characterize using MarvinSketch v18.16, 2018, ChemAxon (https://www.chemaxon.com). UCSF Chimera (Pettersen et al., 2004) and VMD (Humphrey et al., 1996) software were used to manually dock both structures to the homology model of the E5 GAIN domain. To this end, UCSF Chimera was used to mutate positions +3 and +6 into a lysine and then generate all potential rotamers of this residue. Prediction of lysine rotamers was performed using the Shapovalov and Dunbrack (2011) rotamer library implementation of UCSF Chimera. The preferred rotamer (i.e., pointing toward the solvent accessible opening) was chosen and completed to a TCO residue. Finally, H-Tet-Cy5 was manually linked to TCO using UCSF Chimera and VMD. All images were rendered using the built-in ray tracing engine tachyon implemented in VMD.

Molecular dynamics simulations

A list of simulations performed in this work is given in Table S3. Protein preparation, modeling, and general set-up of initial structures was performed using VMD1.9 (Humphrey et al., 1996). Crystal structures of the rat Latrophilin-1 GAIN/HormR (PDB:4DLQ) and mouse GPR56 GAIN/PLL (PDB:5KVM) domains were used to build the L1 and G1 systems, respectively. All co-crystallization atoms except for crystallization water molecules closer than 5 Å to the protein were removed from the crystal structure including the FN3 monobody present in the G1 structure. Histidines 851 and 852 (part of the polyhistidine tag used in protein purification) were removed from the L1 structure. E5 and E2 homology models were used to build the E5 and systems. We then modeled unresolved water molecules using the DOWSER software (Zhang and Hermans, 1996).

The CHARMM-GUI builder (Jo et al., 2008) was used to model, solvate, and build all systems. The native protein sequence was preserved (i.e., no mutation were introduced) in all systems. Only in the case of the E2^C492Y system, a mutation (Cys492Tyr) was introduced. L1 residues Glu774, Glu808, His836, and His846; G1 residues His47 and His170; and E5 residue His484 were protonated. All other titratable residues of the protein were left in the dominant protonation state at pH 7.0. Disulfide bridges were inserted between: Cys480-Cys515, Cys503-Cys532, Cys801-Cys832, and Cys820-Cys834 in L1; Cys35-Cys 91, Cys 121-Cys 177, Cys 346-Cys 377, and Cys 366-Cys 379 in G1; Cys495-Cys525 and Cys513-Cys527 in E5; and Cys 483-Cys512, and Cys 500-Cys 514 in E2. An extra disulfide bridge between Cys328-Cys492 was inserted to study the effect of this bridge in the stability of the protein (i.e., system E2^Cys-Cys). No sugars were added to the system, except for the control glycosylated L1 system where six N-Acetylglucosamine units were linked via β1→N-glycosidic bonds to Asn531, Asn640, Asn741, Asn800 (two N-Acetylglucosamine units linked via β1→4 glycosidic bond), Asn805, and Asn826, based on PDB:4DLQ and UniprotKB:O88917.

Systems were then solvated, neutralized, and the ionic strength adjusted. Prior to production runs, all systems were geometry-optimized, and equilibrated with harmonic positional restraints applied to all C_α atoms of the protein. These restraints were sequentially released in a series of equilibration steps. Production simulations were run in the NPT ensemble at 1 bar and 310 K using Gromacs v5 (Abraham et al., 2015) in combination with the CHARMM36m force field (Huang and MacKerell, 2013). VMD1.9 and MDsrv (Tiemann et al., 2017) were used to visualize all MD simulations. The Gromacs v5 suite, the mdtraj v1.9.3 python module (McGibbon et al., 2015) and the R software were used to post-process and analyze all MD trajectories. Figures were rendered using VMD1.9, and both the python Matplotlib (Hunter, 2007) and R ggplot2 (Wickham, 2016) libraries. Online sessions to static models and dynamics structures were generated using MDsrv (Tiemann et al., 2017) and can also be accessed here: https://zenodo.org/record/4114651.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of imaging data

Quantifications were based on single plane images with an optimized resolution (calculated by LAS X software, Leica) at a zoom level of 0.75. The intensity of H-Tet-Cy5 signal was measured without accumulation or averaging using a hybrid photodetector of an SP8 microscope (63x/1.3 glycerol objective, Leica), which returns linear measurements of single photon counts. Single photon counts of regions of interests (ROIs) were calculated with LAS X software (Leica) based on the raw 8-bit images. ROIs were defined as HEK293T cell patches expressing labeled receptors within the focus level and with intact membrane. ROIs were manually outlined. The sum of

all photons above the background noise were calculated and divided through the number of labeled cells within the ROI. Cells transfected with empty vector (EV) served as a negative control to calculate the background noise. The number of labeled cells were manually counted. Quantification of bioorthogonal H-Tet-Cy5 labels was calculated based on indicated number of images or cells per experiment (n). The experiment was independently repeated by indicated number of sets (N). Individual data points were normalized to the EV and the indicated control construct. Ratios of TA label versus NTF quantities (TA/NTF) were calculated based on the normalized mean value of the quantified bioorthogonal H-Tet-Cy5 labels (TA) and the normalized mean value of the ELISA experiments (NTF).

FRET experiments

FRET experiments were performed on a LSM700 (Zeiss) and were acquired with suitable settings for the different dyes (α-Bungarotoxin-Alexa Fluor 594 [#B13423, Thermo Fisher] excitation: 555 nm and H-Tet-Cy5 [# CLK-015-05, Jena Bioscience] excitation: 639 nm laser) and appropriate detection filters. Same imaging settings were chosen when different constructs were compared. Laser intensity was adjusted for minimal photobleaching and crosstalk, the pixel dwell time was set to 1.2 μsec and imaging was perfomed at room temperature (23°C). Cells were labeled with 1.5 μM H-Tet-Cy5 and 1 μg/ml α-Bungarotoxin-Alexa Fluor 594 simultaneously in CGM at 37°C for 30 min and washed with fresh CGM prior to fixation (with 4% (v/v) formaldehyde for 10 min in HBSS). Before imaging cells were washed 3 times with HBSS. The fluorescence signal was recorded at the basal membrane of individual cells.

Prior to acceptor photobleaching five images of both individual channels were acquired to determine the fluorescence intensities. A selected region of interest (ROI) on the basal membrane was irradiated with the 639 nm laser intenisty (100% intensity) for ~30 s (100 iterations) to photobleach the acceptor Cy5. Postbleach images of both channels were recorded immediately following acceptor photobleaching. FRET was assessed as an increase in donor fluorescence intensity upon acceptor photobleaching.The FRET efficiency was calculated as (donor postbleach - donor prebleach)/(donor postbleach) for all individual recordings.

Statistical methods

ELISA experiments were performed with the indicated number of multiplicates (n) and each experiment was independently repeated by indicated number (N). Individual measurements were normalized to the empty vector (EV) and the indicated control construct. Normalized datasets were tested for their Gaussian distribution; depending on the result datasets from individual receptor variants were compared via an unpaired t test or Mann-Whitney U test.

Normalized datasets of bioorthogonal H-Tet-Cy5 labels were tested for normal distribution and, depending on the result, compared via an unpaired t test or Mann-Whitney U test.

If not indicated otherwise, all graphs show the mean and standard error of the mean (SEM) of all normalized datasets.