

# Structure of an LGR dimer – an evolutionary predecessor of glycoprotein hormone receptors

Zhen Gong<sup>1</sup>, Shuobing Chen<sup>1</sup>, Ziao Fu<sup>1,†</sup>, Brian Kloss<sup>2,†</sup>, Chi Wang<sup>1</sup>, Oliver B. Clarke<sup>3,4,5</sup>, Qing R. Fan<sup>6,7</sup> and Wayne A. Hendrickson<sup>1,3,\*</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032

<sup>2</sup> New York Structural Biology Center, New York, NY 10027

<sup>3</sup> Department of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032

<sup>4</sup> Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY 10032

<sup>5</sup> Irving Institute for Clinical and Translational Research, Columbia University, New York, NY 10032

<sup>6</sup> Department of Molecular Pharmacology and Therapeutics, Columbia University, New York, NY 10032

<sup>7</sup> Department of Pathology and Cell Biology, Columbia University, New York, NY 10032

\* To whom correspondence may be addressed. E-mail: [wah2@cumc.columbia.edu](mailto:wah2@cumc.columbia.edu)

<sup>†</sup> Current addresses: Z. F., Center for the Investigation of Membrane Excitability Diseases, Washington University School of Medicine, St. Louis, MO 63110; Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110

## Summary

The glycoprotein hormones of humans, produced in the pituitary and acting through receptors in the gonads to support reproduction and in the thyroid gland for metabolism, have co-evolved from invertebrate counterparts<sup>1,2</sup>. These hormones are heterodimeric cystine-knot proteins; and their receptors bind the cognate hormone at an extracellular domain and transmit the signal of this binding through a transmembrane domain that interacts with a heterotrimeric G protein. Structures determined for the human receptors as isolated for cryogenic electron microscopy (cryo-EM) are all monomeric<sup>3-6</sup> despite compelling evidence for their functioning as dimers<sup>7-10</sup>. Here we describe the cryo-EM structure of the homologous receptor from a neuroendocrine pathway that promotes growth in a nematode<sup>11</sup>. This structure is an asymmetric dimer that can be activated by the hormone from that worm<sup>12</sup>, and it shares features especially like those of the thyroid stimulating hormone receptor (TSHR). When studied in the context of the human homologs, this dimer provides a structural explanation for the transactivation evident from functional complementation of binding-deficient and signaling-deficient receptors<sup>7</sup>, for the negative cooperativity in hormone action that is manifest in the 1:2 asymmetry of primary TSH:TSHR complexes<sup>8,9</sup>, and for switches in G-protein usage that occur as 2:2 complexes form<sup>9,10</sup>.

# Introduction

There are three glycoprotein hormone receptors (GpHRs) in humans: luteinizing hormone–choriogonadotropin receptor (LHCGR), follicle stimulating hormone receptor (FSHR) and TSHR. Together, they play important roles in reproduction and regulating metabolism<sup>1</sup>. GpHRs belong to a subfamily of class A G protein coupled receptors (GPCRs) called leucine-rich-repeat-containing GPCRs (LGRs). An LGR is characterized by a large extracellular domain (ECD) with multiple leucine-rich repeats (LRRs), a rhodopsin-like transmembrane domain (TMD) of seven transmembrane (TM) helices, and a disulfide-bridged hinge segment that connects the two. The ECD binds the hormone<sup>13,14</sup> while the TMD transmits the signal of this binding to cytoplasmic effectors, notably G proteins<sup>15</sup>.

The structures of signaling complexes for all three human GpHRs have been solved recently using cryogenic electron microscopy (cryo-EM)<sup>3-6</sup>. The three hormone-receptor-G-protein complexes are similar to one another; however, in each case the receptor is monomeric, whereas compelling evidence implicates receptor dimers (possibly other oligomers) in GpHR signaling. Förster and bioluminescence resonance energy transfer (FRET and BRET)<sup>8,16-18</sup> and super-resolution microscopy experiments<sup>19</sup> demonstrate associations; negative cooperativity in hormone binding as well as assays of binding competition for chimeric receptors imply asymmetric 1:2 hormone:receptor association<sup>8,10,20</sup>; switches in TSHR signaling from being through Gs to produce cAMP to being through Gq to produce inositol phosphate<sup>10</sup> or through Gi/o for a biphasic cAMP response, which in consort require the formation of 2:2 TSH:TSHR complexes<sup>9,21</sup>; and functional complementation of mutants defective in binding by mutants defective in signaling require transactivation from one protomer to another<sup>8,22-24</sup>. Most impressively, transgenic mice co-expressing binding-deficient and signaling-deficient LHCGRs reinstate functional hormone action in the absence of wild-type receptors<sup>7</sup>. Such complementation is intrinsic to the natural functioning of the type C heterodimeric GABA<sub>B</sub> receptors<sup>25,26</sup>.

The glycoprotein hormones and their receptors co-evolved from homologs in invertebrate metazoans<sup>27</sup>, in which the hormones are  $\alpha 2/\beta 5$  heterodimers and their type A LGR receptors are already distinguished as having 12 LRRs bounded by disulfide-bridged regions at the N-terminus and at the hinge region<sup>28</sup>. The *Caenorhabditis elegans* genome includes a single LGR gene<sup>29</sup> and genes for the  $\alpha^{30}$  and  $\beta^{31}$  chains encode the CeLGR ligand (Cea2 $\beta$ 5); gene knockouts of any one of these produce the same growth defective phenotype in the nematode<sup>11</sup>. The human homolog of the CeLGR ligand can activate human TSHR, and it is called thyrostimulin (Hsa2 $\beta$ 5)<sup>32</sup>.

We previously solved the crystal structure of Cea2 $\beta$ 5<sup>12</sup>, which has expected similarity to human glycoprotein hormones but is unglycosylated. In this study, we report a nearly full-length cryo-EM structure of the *C. elegans* LGR, which is an asymmetric dimer in an unliganded state. We also show that Cea2 $\beta$ 5 is able to activate CeLGR to produce cyclic adenosine monophosphate (cAMP) in human embryonic kidney (HEK) cells.

# Results

## Characterization of the CeLGR receptor

The full-length CeLGR contains 929 residues, including a C-terminal tail extending beyond the limits of hGpHRs and predicted by AlphaFold to be unstructured (sequence alignment in Extended

Data Fig. 1). For protein production, we truncated residues Ile713 to Ser929 of CeLGR after helix 8 and replaced its native signal peptide (residues 1 to 30) with an influenza hemagglutinin (HA) signal peptide to improve the expression of CeLGR in HEK293S cells. The purified CeLGR protein without reducing agent appears as monomer and ‘dimer’ bands on an SDS-PAGE gel. In the presence of a reducing agent, the dimer band almost disappears, residual monomer bands persist, and two new bands corresponding to lower molecular weight appear (Extended Data Fig. 2a). Complete cleavage into two subunits was observed for TSHR from human thyroid membranes<sup>33</sup>; but we observe an incomplete cleavage of hTSHR as produced from HEK293S cells, which can be completed by limited trypsin digestion (Extended Data Fig. 2b). Based on evidence that a double cleavage had excised a segment of the hinge region<sup>34</sup>, both cryo-EM structures of hTSHR employed constructs that deleted residues Ala317-Phe366<sup>4,5</sup>.

We used proteolytic mass spectrometry to delimit the cleavage site that generated the two polypeptides found in our preparations of CeLGR and of hTSHR. We analyzed each protein after reduction and subsequent digestion with three proteases, coupling the observed coverages from identified proteolytic peptides with knowledge of contiguously ordered cryo-EM structure on either side of a disordered segment, which is where the natural cleavages must have occurred. These considerations delimit the cleavage of CeLGR to within residues Arg343-Arg354 and that of hTSHR to within its residues Arg310-Ser314 (Extended Data Fig. 3). This is surprising for CeLGR since cleavage would have been expected after Arg354 at a canonical furin site (RRKR↓) and surprising for hTSHR in that it excludes the previously suggested cleavages after Asn316 and Phe366<sup>34</sup>. These severances cut at similar sites between the disulfide-bridged cysteine residues of the hinge-region, just beyond HH2 and near the C-terminal end of the hormone-binding ECD (Extended Data Fig. 1). The highly conserved P10 segment joins the ECD to the TMD.

### **Cryo-EM structure of the unliganded CeLGR dimer**

The structure of CeLGR was determined by single-particle cryo-EM to an overall resolution of 3.79 Å (Fig. 1 and Extended Data Fig. 4-6). The cryo-EM data collection, refinement and validation statistics are summarized in Extended Data Table 1. The structure of CeLGR is an asymmetric homodimer. By comparison to the known structures of hGpHRs<sup>3-6</sup>, the ECDs of both CeLGR protomers are tilted down towards the cell membrane, occupying an inactive conformation. Overall, the two protomers are similar in structure (Extended Data Fig. 7a); but protomer B embraces protomer A after an anti-clockwise rotation of 142° and a translation of 4.1 Å toward the ECD along the rotational screw axis that penetrates the dimeric interface (Fig. 1b and c). The angle between the rotation axis and the normal to cell membrane is 8.3° (Fig. 1b) based on a membrane boundary calculated using the Positioning of Proteins in Membrane (PPM 3.0) server<sup>35</sup>. The two protomers are not identical; rather, ECD and TMD are rigid bodies linked such that the ECD is rotated 7.3° further away from the TMD in protomer A than in B (Extended Data Fig. 7).

The density map is well resolved for both protomers from the N termini through TM5. Side chains of residues in these regions could be assigned with certainty. However, despite extensive attempts at classification to resolve ambiguities, the map remains poor at TM6, TM7 and helix 8 for both protomers with little side chain definition. This may be due to a dynamic character for the TM6-TM7 interface. Therefore, the model for TM6 and TM7 in both protomers was built by rigid-body fittings of the AlphaFold-predicted model into the density map.

Five potential N-linked glycosylation sites (Asn80, Asn92, Asn155, Asn372 and Asn410) in CeLGR are predicted by the NetNGlyc-1.0 server<sup>36</sup>. Density was observed to model carbohydrate residues at three potential glycosylation sites (Asn80, Asn92 and Asn155); Asn 372 is in a disordered segment, and density observed at Asn 410 in protomer A was too weak to model.

Since the overall structures of the two protomers are similar and protomer A has the better density, we use it to illustrate structural features in the protomer.

### **The N-terminal cysteine-rich and C-terminal cysteine-rich hinge regions**

As for the hGpHRs, the extracellular ligand-binding domain of CeLGR consists of 12 LRRs flanked by N-terminal and C-terminal cysteine-rich regions (NCR and CCR, respectively). Disulfide-bridged flanking regions are common features of LRR proteins, thought to serve LRR integrity by capping its ends, and those of CeLGR conform to cysteine motifs that characterize GpHRs<sup>37</sup>. LRRs belong to an archaic prokaryotic protein architecture that is widely used in protein-protein interactions. In eukaryotes, LRR domains developed into key recognition modules in innate immunity<sup>38</sup>.

As for the hGpHRs, the NCR region of CeLGR has an additional  $\beta$  strand, named  $\beta_0$ , arranged antiparallel to the LRR  $\beta$  strands that form the inner concave binding surface (Fig. 2a). The NCR region of CeLGR includes three disulfide bridges (Fig. 2a and 2c). The  $\beta$  strand in LRR1 includes three consecutive cysteine residues (Cys67, Cys68 and Cys69), which connect to  $\beta_0$  through the disulfide bonds of Cys69:Cys54 (as for hGpHRs) and Cys67:Cys56; and to the unique N-terminal helix  $\alpha_0$  through the disulfide bond of Cys68:Cys43 (Fig. 2a and 2c).

As first described by Jiang et al. for FSHR<sup>14</sup> and recently confirmed with structures of full-length hGpHRs<sup>3-6</sup>, the CCR hinge region is an integral component of LGR ECDs. Mutagenesis and chimeric receptor studies show importance of the CCR hinge region in GpHR function<sup>39</sup>, seeming to act as a fulcrum<sup>40</sup> in transmitting the signal from hormone binding to the ECD to G-protein activation by the TMD. In essence, the CCR of CeLGR arises as an elaborated return loop between the  $\beta$  strands of LRR11 and LRR12. For CeLGR, the CCR hinge region includes three hinge helices (HH1, HH2 and HH3), LRR12 and the highly conserved 10-residue fragment (P10, residues Leu420 to Tyr429) (Fig. 2b and 2d); this hinge region is disordered between Arg343 and Ile392. It is noteworthy that HH1, P10 and the six cysteine residues in the CCR hinge region are highly conserved in hGpHRs (Fig. 2b and 2d), indicating a common and important biological function of this region. The HH1 of CeLGR links with LRR12 and P10 through two conserved disulfide bridges Cys311:Cys413 and Cys312:Cys423, while HH2 and HH3 of CeLGR are linked by disulfide bridge Cys340:Cys397 (Fig. 2b and 2d). All three disulfides bridge the cleaved segments.

### **Interactions between the ECD and TMD domains of CeLGR**

The TMD of each CeLGR protomer arranges itself in the seven-transmembrane-helix conformation of canonical GPCRs, with ECL2 covering the extracellular end in this case (Fig. 3a). P10 in the CCR hinge region of CeLGR directly connects the ECD to the TMD, and this covalent linkage is buttressed by interactions with HH1, ECL1, ECL2 and TM2. The backbone carbonyl group of Leu420<sup>P10</sup> forms hydrogen bonds with the backbone NH and the side chain hydroxyl group of Ser582<sup>ECL2</sup> (Fig. 3b). The side chain of Asn421<sup>P10</sup> forms a hydrogen bond with



the side chain of Gln505<sup>ECL1</sup> (Fig. 3b). As mentioned previously, P10 interacts with HH1 through a disulfide bridge Cys423<sup>P10</sup>:Cys312<sup>HH1</sup> (Fig. 3b). The backbone carbonyl group of Glu424<sup>P10</sup> forms a hydrogen bond with the backbone NH of Asp495<sup>ECL1</sup> (Fig. 3b). In addition, the backbone NHs of Ile426<sup>P10</sup> and Val427<sup>P10</sup> form hydrogen bonds with the side chain of Asp490<sup>TM2</sup> (Fig. 3b).

Residue Gln505 in ECL1 is essential in interaction with HH1, P10 and ECL2. The side chain of Gln505<sup>ECL1</sup> forms hydrogen bonds with the side chain of His309<sup>HH1</sup> and Asn421<sup>P10</sup>, and the carbonyl backbone of Ser579<sup>ECL2</sup> (Fig. 3b). The ECL2 of CeLGR interacts with TM2 through a hydrogen bond between the carbonyl backbone of V583<sup>ECL2</sup> and side chain of Y482<sup>TM2</sup>, and TM3 through a disulfide bridge Cys584<sup>ECL2</sup>:Cys510<sup>TM3</sup> (Fig. 3b and 3d).

The interactions mentioned above are summarized in Fig. 3c. Most residues involved are conserved between CeLGR and all three hGpHRs (Fig. 2d and 3d). The ones which are not conserved (Leu420<sup>P10</sup>, Ser579<sup>ECL2</sup>, Val427<sup>P10</sup> and V583<sup>ECL2</sup>) interact through backbone atoms; thus, variations at these residues will not affect their interactions. His309<sup>HH1</sup> in CeLGR has evolved to serine in all three hGpHRs (Fig. 2d). A systematic study of substitutions of Ser277 in hLHCGR, corresponding to His309<sup>HH1</sup> in CeLGR, indicated that the activity of S277H LHCGR did not change significantly from the wild-type receptor<sup>41</sup>.

In summary, the ECD and TMD of CeLGR interact with each other through the CCR hinge region and extracellular loops in a delicate arrangement, which keep the pore to the TMD cavity 'blocked'. The ECD of CeLGR could be viewed as a tethered antagonist from this point.

### Dimer interface between CeLGR protomers A and B

The CeLGR protomers A and B form a compact homodimer through an extensive dimer interface that primarily involves their TMDs. In particular, TMs1, 6 & 7 from protomer A contact TMs1, 5, 6 & 7 from protomer B, with a buried surface area of 2075.2 Å<sup>2</sup> from the TMDs (Fig. 4a to d). As mentioned earlier, the density map of TM6 and TM7 in both protomers is poor with no density observed for many side chains. The electron density for half of TM6 in protomer B could not be resolved at all, and it is indicated in grey in Figs. 4a to c. In addition, the central layers of direct homodimer contacts are flanked by an assigned cholesterol molecule, which facilitates the interaction between TM6 of protomer A and TM1 of protomer B (Fig. 4a to d).

Direct contacts are also observed between the CCR hinge region of protomer A and the ECD of protomer B (Fig. 4e) with a buried surface area of 185.3 Å<sup>2</sup>. The backbone carbonyl group of Ile392 in protomer A forms a hydrogen bond with the side chain amino group of Gln230 in protomer B (Fig. 4e). The side chains of Glu395 in protomer A and Lys233 in protomer B make an ionic interaction (Fig. 4e). In addition, favored van der Waals contacts are observed between Leu232 in protomer B and Gly393, Ala394, Glu395 in protomer A, Gln230 in protomer B and Ile392 in protomer A, as well as Arg209 in protomer B and Pro396 in protomer A (Fig. 4e).

Urizar et al. have reported that hTSHR and hLHCGR form homo- and heterodimers via interactions involving primarily their TMDs and their ECDs are dispensable for dimerization. However, the ECDs must play a role in the dimeric interaction because differences were observed in the FRET signals obtained from dimers made of intact or truncated hTSHR devoid of ECD<sup>8</sup>. Our CeLGR structure supports their observation that dimeric interactions involve primarily the TMDs. It is unlikely that the spare interactions between ECDs of the protomers (Fig. 4e) will affect

receptor dimerization very much since the major dimer interface resides in the TMD; however, hormone binding may generate greater interaction.

### Conformational dynamics

Various features of the analysis indicate that CeLGR in these preparations is highly dynamic. Distinct sub-populations could not be detected. Thus, we undertook a 3D variability analysis (3DVA) of the particles by *cryoSPARC*<sup>42</sup>, which corroborates and quantifies this impression (Supplementary Videos 1-2). The CeLGR protomers appear to be in complex oscillatory motions about the interprotomer screw axis. Rotational orientations of the protomers vary smoothly through the trajectory, most obviously for the ECDs; each protomer contracts and extends along the screw axis; and the TMD micelle appears to expand and contract in diameter.

To further analyze these dynamics, we built models at the motional extrema by rigid body fittings of the ECD (from N terminus to LRR12) and TMD (TM1 to TM7) from each protomer into the initial and final frame of the computed 3DVA progression. These two models (conformers 1 and 2, respectively) were then compared with the model in the 'rest' state as built with the consensus frame (same with Fig. 1b). Protomers A and B fluctuate in opposite directions asymmetrically about the interprotomer screw axis (Fig. 5 a and b). As a result, relative to the rest state, the angle between the ECDs decreases by 6.9° for conformer 1 and increases by 11.1° for conformer 2. Similarly, the interprotomer angle for the TMDs decreases by 5.3° for conformer 1 and increases by 10.8° for conformer 2.

Concomitant with the rotations, the ECD of each protomer also moves relative to the deduced membrane surface (Fig. 5c). To measure this angle, we defined a pseudo axis from C<sub>α</sub> 55 in β0 to C<sub>α</sub> 413 in LRR12 and determined the angle between this axis and the plane normal to the screw axis. Taken again relative to the rest state, this angle decreases by 2.2° and 3.6° in protomers A and B, respectively, on going to conformer 1 and it increases by 5.0° and 6.1° in A and B, respectively on going to conformer 2. TMD positions relative to the membrane plane are relatively constant through the oscillations (Fig. 5c); however, screw dispositions of the TMDs do shift and they do so asymmetrically [(Δχ, Δt<sub>χ</sub>) = (-3.2°, - 0.5Å) and (+10.1°, - 1.6Å) respectively for conformers 1 and 2 relative to the rest screw (χ=140.0°, t<sub>χ</sub>=5.1Å) for TMD<sub>B</sub> relative to TMD<sub>A</sub>]. At the same time, the TMD separations also vary (Fig. 5b); and, to quantify these changes, we define an outer surface width by the distance between C<sub>α</sub> 573 in TM4 of protomer A and C<sub>α</sub> 502 in TM3' of protomer B. On going from conformer 1 to conformer 2, this distance decreases from 50.5 Å to 48.7 Å through the rest distance of 49.3 Å.

The extents of rotational displacements and ECD elevations are appreciably greater in protomer B than in protomer A. This is consistent with the observation that the density map quality is worse in protomer B than in protomer A. The extensive TMD interface is both intimate and highly dynamic, seeming to involve correlated movements between the two protomers.

### Activation of CeLGR by Cea2β5 *in vitro*

The *C. elegans* genome contains only one LGR sequence<sup>29</sup>, the type A LGR that is the subject of this study and only one LGR ligand, Cea2β5<sup>43</sup>. It has recently been reported that the Cea2β5:CeLGR (called GPA2/GPB5:FSHR-1 in Ref. 10) pathway controls growth in *C. elegans*<sup>11</sup>. Cea2β5 signaling is required for normal body size and intestinal lumens, acting through activation

of CeLGR for cAMP production<sup>11</sup>. In addition, CeLGR (FSHR-1) has been implicated in responses to infection, oxidative and freezing-thawing stress in *C. elegans*<sup>11</sup>, which suggest that the thyrostimulin (alias for  $\alpha 2\beta 5$  hormone) signaling pathway may also regulate protective immune and stress responses in *C. elegans*.

To determine whether Ce $\alpha 2\beta 5$  can activate CeLGR, we applied a cAMP-based competitive assay in HEK293S cells. HEK293S cells infected with baculovirus carrying the CeLGR gene were incubated with serially diluted Ce $\alpha 2\beta 5$  hormone. The cAMP production was quantitatively determined to calculate the half maximal effective concentration (EC<sub>50</sub>) of Ce $\alpha 2\beta 5$ , which is 159.1 nM (Fig. 6a). The EC<sub>50</sub> of recombinantly expressed and purified hTSH to wild type hTSHR was also measured as a control. The measured EC<sub>50</sub> of recombinantly purified hTSH is 7.5 nM (Extended Data Fig. 8), which is comparable to that of the wild type hTSH (3 nM)<sup>44</sup>. Of interest, the G<sub>s</sub> is highly conserved between human and nematode with sequence identity percentage of 66%<sup>45</sup>, consistent with the observed coupling between CeLGR and G<sub>s</sub> protein found in HEK293S cells. The potency of Ce $\alpha 2\beta 5$  is lower than that of hTSH according to our measured EC<sub>50</sub>. The imperfect 'match' between CeLGR and mammalian G protein, as well as the tight truncation of CeLGR right after helix 8 may account for the lower potency of Ce $\alpha 2\beta 5$ .

In addition, the dissociation constant (K<sub>d</sub>) of the Ce $\alpha 2\beta 5$ :CeLGR complex was determined as 153.3 nM using microscale thermophoresis (MST) (Fig. 6b), which is in a comparable range to the EC<sub>50</sub> of Ce $\alpha 2\beta 5$ .

## Discussion

This study reveals several structural features of the LGR from *C. elegans* that are of interest in relation to the functioning of this particular receptor and, by analogy, to the activity of related receptors. Most prominently, this nearly full-length LGR is dimeric, both in solution after detergent extraction and as imaged on cryo-EM grids. These dimers show pronounced asymmetry, and they are highly dynamic in the apo state, without the cognate  $\alpha 2/\beta 5$  hormone and heterotrimeric G protein being present. Moreover, the receptor chains have been cleaved in the CCR hinge region at a site between disulfide-bridged cysteine residues.

### CeLGR in relation to mammalian glycoprotein hormone receptors (GpHRs)

CeLGR and the mammalian GpHRs are similar in sequence and in structure, and the relationship is closest to TSHR. The co-evolution of type A LGRs and their activating ligands is thought to have proceeded in early vertebrates in two rounds of whole genome duplication from invertebrate counterparts, the first producing progenitors of TSH and TSHR and the second adding FSH and LH and their receptors; thus, by evolution, CeLGR is expected to be most closely related to TSHR. Indeed, CeLGR is like hTSHR in having a sizable insertion, relative to hFSHR and hLHCGR, within its CCR hinge region and, also, in being cleaved just beyond HH2 between the disulfide-bridged cysteine residues of the CCR (Extended Data Fig. 1). As for CeLGR, helices HH1, HH2 and HH3 are all ordered in structures of TSHR<sup>4,5</sup>, whereas only HH1 is ordered in the structures of LHCGR<sup>3</sup> and FSHR<sup>6</sup>. Additionally, just as CeLGR binds Ce $\alpha 2\beta 5$ , human TSHR uniquely retained high affinity for Hs $\alpha 2\beta 5$  (thyrostimulin) while it acquired affinity for hTSH (thyrotropin) as well. Finally, the asymmetric CeLGR dimer structure is consistent with abundant evidence for functional GpHR dimerization<sup>7,8,10,16-24</sup>, which is particularly compelling for hTSHR<sup>8,10,16,20,21</sup>.

The CeLGR dimer differs from other GpHR dimers in the strength of its self-association. Whereas we found no monomers in our 2D classifications of cryo-EM particles, and only a small fraction (1.5%) of trimers (Extended Data Fig. 4), only monomer classes were reported for the cryo-EM analyses of human GpHRs<sup>3-6</sup>. Monomers also predominate (70%) in imaging analyses of FSHR as expressed in HEK293 cells; however, a substantial fraction of dimers (15.5%) and other oligomers are also present in basal conditions, with some variation in the presence of hormone variations<sup>46</sup>. Similar oligomerization levels (41.4%) were found for LHR expressed in HEK293 cells<sup>19</sup>. Of course, dimerization is concentration dependent; and, in the case of TSHR, the expression level in thyrocyte cultures needed to be at the thyroid level *in situ* to exhibit physiological biphasic responses reflective of dimerization<sup>9</sup>.

The puzzle as to why detergent-extracted CeLGR is dimeric whereas similarly purified human GpHRs are monomeric may be explained by properties of the dimer interface of CeLGR. Nearly all of the surface area buried into that interface comes from the transmembrane helices (91.8%), for which the sequences are highly conserved among the three human GpHRs (68-72% identity overall) and especially so for residues at the TMD interior (84-88% identity). The respective identity levels between the CeLGR and human GpHR sequences, while appreciably lower, are also quite high (35-37% overall and 58-63% for interior residues). The dimer-interface residues of CeLGR are more conserved than outward-facing residues generally (29-37% identity with human GpHR homologs as compared to 25-28% overall). Similarly, among human GpHRs, the homologs of dimer-interface residues are more conserved than are all outward-facing residues (63-74% compared to 61-67% identity overall). The CeLGR interface includes seven residues that are distinguished from counterparts that are identical in human GpHRs and in *Drosophila* LGR1 as well (A442L, T637K, V644F, L647F, V675L, F685C and V692A). As for the human GpHRs, detergent-extracted *Dm*LGR1 purifies as a monomer (data not shown). We postulate that the CeLGR interface distinctions stabilize these dimers, whereas the similar but distinctive interfaces of the homologous receptors are more labile and do not survive detergent solubilization from the lipid bilayer.

### **The CeLGR dimer as related to dimerization of other GPCRs**

Whether the CeLGR structure truly reflects the nature of functional dimers of TSHR and other GpHRs remains to be seen. Meanwhile, since the evidence for physiological relevance of GpHR dimers is compelling, the CeLGR dimer provides us with hypotheses on the functioning of GpHRs. The growing body of evidence on the obligate dimers of class C and class D GPCRs<sup>47</sup> supplements this structural foundation. In particular, functional complementation such as that employed in GpHR experiments<sup>7,8,21-24</sup> is intrinsic to GABA<sub>B</sub> receptors, which are heterodimeric class C GPCRs wherein the B1 subunit binds the ligand but does not contact Gα and the B2 subunit does not interact with GABA but does bind Gα<sup>25,26</sup>. The homodimeric class C GPCRs, metabotropic glutamate receptor<sup>48</sup> and calcium-sensing receptor (CaSR)<sup>49</sup>, also have only one G protein coupled to the receptor dimer.

The TMD interfaces in the known class C structures are more symmetric than that in the CeLGR structure; but they nevertheless involve similar TMD surfaces. For the active GABA<sub>B</sub> and CaSR receptors, these are centered at TM6-TM6' interactions with only peripheral involvement of the adjacent TM5 and TM7 helices; for CeLGR, being off center, TM6 and TM7 of protomer A

contact the TM5-TM7 surface. Also, whereas direct contacts are appreciable in the active states for class C receptors, and they vary but are typically minimal in inactive states<sup>25,48,50,51</sup>, the CeLGR dimer interface in the hormone-free (presumably inactive) state here is intimate albeit mobile. Such TM6-centered dimers may not be universal, however; cross-linking evidence for 5HT<sub>2c</sub> and some other class A receptors contacts involving TM4 and TM5<sup>52</sup>.

### Implications for receptor activation

The structure of the CeLGR dimer sheds light on the activation mechanism of GpHRs. Its asymmetric assembly suggests a structural basis for the negative cooperativity observed in hormone binding<sup>8,20</sup>, for its corollary in switching from G<sub>s</sub> signaling from 1:2 complexes to G<sub>q</sub> or G<sub>i/o</sub> signaling in 2:2 complexes<sup>9,10,21</sup>, and for the functional complementation provided by the co-expression of binding-deficient and signaling-deficient receptors<sup>7,8,22-24</sup>. The current activation model based on structures of monomeric GpHRs<sup>3-6</sup> could not explain these effects, whereas the compactly associated CeLGR dimer makes allosteric modulation and rescue of signaling possible. Atomic details of the conformational changes found in the full-length GpHR monomers upon binding of their hormones coupled with constraints from the dimer structure allow us to contemplate how this might happen.

As found for the hTSHR, hFSHR and hLHCGR monomers in their inactive hormone-free states, both CeLGR protomers have their ECDs in a 'down' orientation. Upon activation for hormone binding, each GpHR then adopts an 'up' conformation with its ECD rotated upward by 45-55° relative to the ligand-free state<sup>3-6</sup>. We presume that CeLGR similarly adopts an ECD-up conformation when hormone associated and that the CeLGR asymmetry is retained as it forms a 1:2 complex. Whether such dimers have both protomers ECD-Up or only the one that is hormone-bound is uncertain; however, we successfully modeled both alternatives. We built Cea2β5-activated CeLGR dimers by first superimposing the TMD of hTSHR in a hTSH:hTSHR:G<sub>s</sub> complex (PDBid 7UTZ) onto each CeLGR protomer, then moving one or both of the CeLGR ECDs into superposition with the ECDs of this TSHR 'dimer', and finally replacing hTSH on protomer A with the Cea2β5 ligand<sup>27</sup> (Extended Data Fig. 9a).

Cea2β5 could only be fitted to protomer A because the binding site in protomer B, whether ECD-up or ECD-down, is blocked by the CCR hinge region of protomer A as it is for both ECDs down (Fig. 4e). As noted for hGpHRs<sup>3-6</sup>, Cea2β5 would be expected to collide with the cell membrane if bound to CeLGR in the down conformation (Extended Data Fig. 9b). Moreover, there is no room for a second Cea2β5 unless the subunits are rotated, with both ECDs up, from asymmetry (142° apart) to symmetry (180°). In accord with the observed negative cooperativity, this symmetrization entails an energetic penalty for the second-site binding event, which is lessened when the receptor is pre-disposed by activating mutation<sup>20</sup> to having both protomer ECD-up for the 2:2 state.

Another question is how many heterotrimeric G proteins couple to a CeLGR dimer? There is sufficient room for only a single heterotrimeric G protein at a time to bind to the 1:2 CeLGR dimer; however, G-protein coupling with either protomer A (cis activation) or B (trans activation) seems possible (Extended Data Fig. 9a). In light of functional complementation experiments for hGpHRs and structural results from class C GPCRs, trans-activation would be expected. The



actual activation mechanism for dimeric CeLGR is complicated and remains to be determined experimentally.

### Implications for physiology

Little is known about the molecular physiology of the LGR- $\alpha 2\beta 5$  system in nematodes, but much is known about the GpHR-hormone systems of humans. What role the dimer may play in the observed impact of the LGR system in worm physiology, and whether the tight association in the CeLGR dimer is important for its functioning is unknown. As discussed above, there is considerable evidence for the role of GpHR dimers in human physiology; however, these dimers are relatively loosely self-associated since they do not survive detergent-extraction from cell membranes. Moreover, TSHR dimerization in cells requires expression at the level of thyrocytes *in vivo*<sup>9</sup>. How labile GpHR dimerization impacts biological function is unclear; but the resulting effects of negative cooperativity are evident: whereas the 1:1 hormone:dimer complex promotes Gs signaling and cAMP production, the 2:1 complex from higher TSH levels promotes Gi/Go signaling to elicit IP1 production and a biphasic cAMP response<sup>9,10,21</sup>.

Negative cooperativity, resulting from allosteric communication between protomers of the dimer, is essential for the function of GpHRs. For instance, while the hCG level is very low (0.04 to 5.5 ng/ml) for unpregnant women, it rises rapidly during pregnancy and can reach 1,952 to 19,958 ng/ml at a peak around 8-10 weeks<sup>53</sup>. Negative cooperativity provides exquisite sensitivity at low hormone concentrations while buffering against acutely elevated hormone levels<sup>54</sup>.

## Materials and methods

### Homolog screen and construct design

Homologs of GpHR from the following species were cloned into a BacMam vector with oxidizing environment-optimized GFP (oxGFP)<sup>55</sup> fused at the C-terminus for fluorescence-detection size exclusion chromatography (FSEC) screen<sup>56</sup>: *Caenorhabditis elegans* (UniProt code: G5EG04), *Drosophila melanogaster* (Fruit fly; UniProt code: Q9VEG4), *Petromyzon marinus* (Sea lamprey, UniProt code: Q0P6K8), *Danio rerio* (Zebrafish; UniProt code: F1R4X9), *Xenopus laevis* (African clawed frog; UniProt code: G9M6I0), *Bothrops jararaca* (Jararaca; UniProt code: Q6K0L4), *Gallus gallus* (Chicken; UniProt code: P79763), *Rattus norvegicus* (Rat; UniProt code: P20395), *Mus musculus* (Mouse; UniProt code: P35378), *Bos taurus* (Bovine; UniProt code: P35376) and *Homo sapiens* (Human; UniProt code for hFSHR: P23945, hLHCGR: P22888 and hTSHR: P16473). The codons from each species other than human have been optimized for expression in human cells. An influenza hemagglutinin (HA) signal peptide was fused at the N terminus of each gene in place of its original signal peptide to enhance protein insertion into cellular membranes. CeLGR (UniProt accession code: G5EG04) was identified as the best candidate for structural studies because of its large expression yield and good solubility in detergent.

CeLGR contains a large intracellular domain (Ile713 to Ser929) at its C-terminus compared with hGpHRs (Extended Data Fig. 1), which is predicted by AlphaFold to be unstructured. In order to reduce structural flexibility potentially introduced by the unstructured C-terminus, CeLGR was truncated after helix 8. Therefore, the final construct contains an HA signal peptide replacing the natural signal peptide (CeLGR residues 1 -31) at the N-terminus, followed with a Flag tag,

residues Gln31 to Arg712 of CeLGR, a triple alanine linker and a Flag tag at the C-terminus for affinity purification. The oxGFP was removed by inserting a stop codon after the C-terminal Flag tag. The presence of both N-terminal and C-terminal Flag tag is because a proteolytic posttranslational cleavage was observed during the maturation event of CeLGR. The original plan was to purify the N-terminal Flag-tagged ECD with anti-Flag M1 antibody affinity resin and the C-terminal Flag-tagged TMD with anti-Flag M2 antibody affinity resin. Actually, the cleaved ECD and TMD of CeLGR are linked by three disulfide bridges in the CCR hinge region. Therefore, either one of the two Flag tags should be sufficient to purify the CeLGR.

### **Expression and purification of CeLGR**

The gene of the expression construct was cloned into a BacMam expression vector containing a human cytomegalovirus (CMV) promoter. Baculovirus was made in Sf9 cells. For large-scale expression, suspension-adapted HEK293S, which lack N-acetyl-glucosaminyltransferase I (GnTI), were grown in Freestyle 293 expression medium (Life Technology) supplemented with 2% Fetal Bovine Serum (FBS) at 37 °C in the presence of 8% CO<sub>2</sub>. The culture was transduced with P3 baculovirus once cell density reached 2.0 x 10<sup>6</sup> to 3.5 x 10<sup>6</sup> cells per milliliter. After incubation for 8-16 h at 37 °C, 10 mM sodium butyrate was added to enhance the expression level, and the temperature was changed to 30 °C for protein expression. Cells were harvested 72 h after adding sodium butyrate and resuspended in a buffer containing 20 mM HEPES, pH 7.5, 300 mM NaCl and 10 mM MgCl<sub>2</sub>, supplemented with EDTA-free protease inhibitor cocktail tablet (Sigma) and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell pellets were stored at -80 °C until use or lysed immediately for protein purification.

Cells were lysed using an EmulsiFlex-C3 high pressure homogenizer (Avestin) with three passes at 10,000-15,000 psi. Cell debris was removed by centrifugation at 2850 x g for 15 min. Cell membranes were collected by ultracentrifugation for 1 h in a Beckman 45 Ti rotor at 40,000 r.p.m. The membranes were homogenized and subsequently solubilized for 2 h in a buffer composed of 20 mM HEPES, pH 7.5, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 1% lauryl maltose neopentyl glycol (LMNG), and 0.2% cholesteryl hemisuccinate tris (CHS), supplemented with protease inhibitor cocktail tablet (Sigma) and 1 mM (PMSF). After insoluble material was removed by ultracentrifugation for 1 h at 32,000 r.p.m. in a Beckman 50.2 Ti rotor, the supernatant was incubated with anti-FLAG M2 resin (Sigma) by rotating overnight. The resin was washed with 30 column volumes of wash buffer containing 50 mM Tris HCl, pH 7.4, 150 mM NaCl, 0.01% LMNG and 0.002% CHS and the protein was eluted using wash buffer supplemented with 0.2 mg/ml Flag peptide (Sigma). The eluted protein was concentrated using a 100-kDa cutoff concentrator and injected into a Superose 6 Increase size exclusion column (Cytiva) equilibrated in a buffer composed of 20 mM HEPES, pH 7.5, 150 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.002% LMNG, and 0.0004% CHS. The peak fraction was collected for imaging. All purification steps were conducted on ice or at 4 °C.

### **Construct design, expression and purification of hTSHR**

The wild type hTSHR (UniProt code: P16473, residues 21-764) was cloned into a BacMam expression vector containing a CMV promoter with an HA signal peptide at its N-terminus in place of its native signal peptide and a Flag tag at its C-terminus. The expression and purification procedures for hTSHR were the same as those used for CeLGR.

## Characterization of cellular proteolysis of expressed CeLGR and hTSHR

The purified CeLGR and hTSHR receptors were analyzed by SDS polyacrylamide gel electrophoresis (SDS PAGE) with or without the presence of a reducing agent DTT to break the disulfide bridges. To test whether residual monomer bands reflected intact receptors or other effects, the purified proteins also went through limited trypsin digestion by adding trypsin to a final concentration of 0.005% to 0.01% followed with incubation at room temperature for 15 min. A 2-fold excess of trypsin inhibitor (Sigma) was then added, and incubation continued for 10 min before adding DTT.

The site of natural protease cleavage was delimited by mass spectrometric (MS) analysis of proteolytic peptides after digestion of the denatured receptors. Because natural cleavage was incomplete for the receptors as expressed in HEK293S cells, we used proteolytic MS analysis of only the post-cleavage bands from SDS PAGE (highlighted in red boxes in Extended Data Fig. 2). The gel band was excised, and *in-gel* digestion was performed as previously described<sup>57</sup> with minor modifications. Briefly, the proteins were digested using trypsin, chymotrypsin or GluC. The resulting peptides were dissolved in a solution of 3% acetonitrile and 0.1% formic acid. Peptides were separated using a reversed-phase C18 column (25 cm × 75 μm, 1.6 μm, IonOpticks). Peptide MS/MS analysis was performed using a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer.

Raw mass spectrometric data were analyzed using the MaxQuant environment v.2.6.1.0, utilizing Andromeda for database searching with default settings with a few modifications. The default tolerances were used for the first search (20 ppm) and the main search (6 ppm). MaxQuant was configured to search against the reference human/*C. elegans* proteome database, which was downloaded from UniProt. The search parameters included trypsin/ chymotrypsin /GluC digestion with up to 2 missed cleavages. The false discovery rates (FDRs) for peptides, sites, and proteins were all set to 1%. The output combined folder was uploaded into Scaffold 5.0 (Proteome Software). Spectral counting was employed for comparative sample analysis.

## Expression and purification of Cea2β5 and hTSH

Cea2β5 was expressed and purified as previously described<sup>12</sup> with a brief summary here. The genes coding for Cea2 (UniProt code: A0T3A2, residues 29-120) and Ceβ5 (UniProt code: A7DT38, residues 20-125) were cloned into two BacMam expression vectors. Both Cea2 and Ceβ5 have an N-terminal HA signal peptide in place of the original signal peptide. A Flag affinity tag was engineered between the HA signal peptide and Cea2, while an HA affinity tag was engineered between the HA signal peptide and Ceβ5 to facilitate affinity purification. Suspension-adapted HEK293S (GnTI<sup>-</sup>) cells were used for co-expression of Cea2β5 using baculovirus produced by Sf9 cells. The Cea2β5 protein was purified using anti-Flag M2 antibody affinity column and size exclusion chromatography.

The hTSH construct comprises the native signal peptide from hTSHβ (UniProt code: P01222, residues 1-20), followed by an HA affinity tag, the mature sequence of hTSHβ (UniProt code: P01222, residues 21-138), a 15-residue linker (GGGSGGGSGGGSGGG), the mature sequence of hα (UniProt code: P01215, residues 25-116), a triple alanine linker and a Flag affinity tag at C-terminus, which is analogous to the construct used by Fan and Hendrickson for the structural determination of hFSH bound to the extracellular hormone-binding domain of its receptor using

X-ray crystallography<sup>13</sup>. Suspension-adapted HEK293S (GnT1<sup>-</sup>) cells infected with baculovirus secreted the hTSH into the cell medium. Both anti-HA (Sigma) and anti-Flag M2 antibody affinity resins were tested to purify the hTSH. No protein could be purified using the anti-HA affinity resin. It is very likely that the anti-HA resin could not bind to HA-tagged proteins well (not limited to hTSH) according to our experience. Finally, the hTSH was purified using an anti-Flag M2 antibody affinity column and Superdex 75 Increase size exclusion chromatography.

### **Cryo-EM sample preparation and data collection**

EM grids (Quantifoil R0.6/1, 400 mesh, holey gold) were glow discharged for 25 s (6.4 sccm H<sub>2</sub>, 27.5 sccm O<sub>2</sub>, 10 watts) using a Gatan Solarus model 950 Advanced Plasma Cleaning System prior to sample application. 3  $\mu$ l of protein solution at 0.3-0.4 mg/ml was applied onto a grid and then plunge-frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (FEI). The Vitrobot chamber was set to 100% humidity at 4 °C. The sample was blotted for 8 s with a wait time of 30 s and a blot force of 3.

Cryo-EM imaging was performed on a Titan Krios electron microscope (FEI) operating at 300 kV, equipped with a post-column GIF Quantum energy filter with a slit width of 20eV, and a K2 summit direct electron camera (Gatan) in counting mode. Micrographs were collected with 70  $\mu$ m C2 aperture at a calibrated magnification of 130,000 corresponding to a magnified pixel size of 1.06 Å. Each micrograph comprises 50 sub-frames with a total accumulated dose of 71.01 e<sup>-</sup>/Å<sup>2</sup>, recorded for a total exposure time of 10 sec at a dose rate of 8 e<sup>-</sup>/pixel/sec. Data acquisition was done using Leginon across a nominal defocus range of -1.2  $\mu$ m to -2.0  $\mu$ m. A total of 8,028 micrographs were acquired as does-fractionated image stacks.

### **Cryo-EM image processing**

The cryo-EM data processing workflow is summarized in Extended Data Fig. 4. Maps and raw videos have been deposited at EMDB (Extended data Table 1) and EMPIAR (EMPIAR-11917). Subsequent steps were performed in cryoSPARC (v4.5.0) unless otherwise indicated<sup>58</sup>.

Patch-based motion correction and dose weighting of the 8,028 cryo-EM movies were executed in cryoSPARC using the Patch Motion job type. Patch-based CTF estimation was conducted on the aligned averages utilizing Patch CTF. Out of these, 7,031 micrographs were selected based on a CTF estimate resolution of < 4 Å, relative ice thickness between 0.9-1.1, total motion distance of < 50 pixels, and a defocus range of < 1200.

The initial round of particle picking employed Blob Picker, extracting particles in a 256-pixel box. Six typical classes emerged post 2D classification. The secondary round of particle picking used template picking with the six identified classes, resulting in 5,634,393 particles being picked. Subsequent multiple rounds of 2D classification yielded 997,598 selected particles. One well-defined class, consisting of 185,961 particles, was identified and selected after 3D classification. Nonuniform refinement of this class achieved a resolution of 3.87 Å. For the third-round particle picking, 75 exemplary micrographs, containing 70-90 good particles each (totaling 5,538 particles), were selected as the training dataset.

A tertiary round of particle picking utilized Topaz<sup>59</sup>, initially training the Topaz neural network with the 75 micrographs in the training dataset. In total, 965,789 particles were selected using the Topaz training model. Post 2D classification, 940,614 particles were chosen, with one well-

defined class being selected after 3D classification. Subsequently, 286,046 particles corresponding to CeLGR were re-extracted with re-centering in a 256-pixel box.

An initial refinement of all 286,046 CeLGR particles, using nonuniform refinement, yielded a consensus refinement with a resolution of 3.82 Å. The application of dynamic masking and real-space windowing of particle images during refinement was not enabled. Further 3D classification without alignment resulted in the best class with complete ECD and TMD, consisting of 72,220 particles, achieving a nonuniform refinement resolution of 3.79 Å.

### 3D variability analysis

After 3D classification, 3D Variability Analysis<sup>42</sup> (3DVA) was carried out with 286,046 selected particles in cryoSPARC, solving in three modes. Reconstructions were calculated along each mode. A mode corresponding to rotation of each protomer, in combination of contraction and extension of the CeLGR dimer was identified. This mode was then used (using the 3D Variability Display job type) to split the particles into 20 clusters, and a reconstruction calculated for each to display the 3D variability.

### Atomic model building and refinement

An initial model for the CeLGR homodimer, which includes residues from the N-terminus to TM5 in each protomer, was generated using phenix.predict\_and\_build<sup>60</sup>. This model was then extended manually and completed in Coot<sup>61</sup>. The cholesterol molecule and glycans were added where justified by the density map and chemical environment. The model was finalized by rebuilding in ISOLDE<sup>62</sup> and refinement with phenix.real\_space\_refine<sup>63</sup>. Figures were prepared using ChimeraX<sup>64</sup>. Refinement and validation statistics are provided in Extended Data Table 1.

### cAMP assay

The same construct for expression of CeLGR and hTSHR was cloned into a pcDNA3.1 vector for the cAMP assay. HEK293S (GnTI<sup>-</sup>) cells grown in Freestyle 293 expression medium supplemented with 2% FBS were seeded onto a 6-well plate at  $1.0 \times 10^6$  cells per milliliter. Two hours later, cells were transiently transfected with hTSHR plasmids using Lipofectamine 3000 (Invitrogen) or infected with P2 virus of CeLGR, the same construct used for protein preparation for structural analysis. Transiently transfected (hTSHR) or virus-infected (CeLGR) cells were incubated at 37 °C in the presence of 8% CO<sub>2</sub> for two days. Next, cells were harvested with centrifugation at 200 xg for 5 min, washed with phosphate-buffered saline (PBS), resuspended in stimulating buffer (Cisbio), seeded onto a 384-well plate and incubated with serially diluted ligand for 2 h at 37 °C in the presence of 8% CO<sub>2</sub>. The cAMP accumulation was measured using the cAMP G<sub>s</sub> dynamic kit (Cisbio) following manufacturer's instructions. In this competitive assay, cAMP is covalently labeled with a red-emissive acceptor dye d2, and an anti-cAMP antibody is labeled with a cryptate donor. A Förster Resonance Energy Transfer (FRET) signal is detected when the d2-labeled cAMP (acceptor) binds to an cryptate-labeled anti-cAMP antibody (donor)<sup>65</sup>. Upon stimulation of CeLGR in HEK293S cells with addition of Cex2β5, native cAMP is generated. After cell lysis and the addition of two labeled compounds, unlabeled native cAMP will compete with the d2-labeled cAMP, and thus a decrease of FRET is observed. The specific signal is inversely proportional to the concentration of native cAMP produced in the sample. The fluorescence emissions at two different wavelengths (665 nm and 620 nm) were read on the



EnVision 2104 Multilabel Microplate Reader (Perkin Elmer). Data were analyzed using GraphPad Prism 7.0 and were presented as mean  $\pm$  standard deviation (s.d.) of measurements from an experiment of n=3 biological replicates.

### **MST measurements**

The dissociation constant ( $K_d$ ) of Cea2 $\beta$ 5 to CeLGR was measured using Microscale Thermophoresis (MST). Cea2 $\beta$ 5 and CeLGR were purified separately as described above. Purified CeLGR was labeled with the RED fluorescent dye NT-647-MALEIMIDE using the Monolith NT Protein Labeling Kit RED - MALEIMIDE (NanoTemper Technologies) following manufacturer's instructions. 50 nM of labelled CeLGR dimer was incubated with serially diluted Cea2 $\beta$ 5 with a 1:1 (v:v) ratio. The normalized fluorescence ( $F_{norm}$ ) of the mixture was read on the Monolith NT.115 instrument (NanoTemper Technologies) selecting LED color in red. Data were analyzed using GraphPad Prism 7.0 and were presented as mean  $\pm$  s.d. of measurements from an experiment of n=4 biological replicates.

### **Data availability**

The cryo-EM density map for CeLGR in the apo state have been deposited in the Electron Microscopy Data Bank under accession code EMD-43735. Atomic coordinates for the CeLGR structure in apo state have been deposited in the RCSB Protein Data Bank under accession code 8W1Z. Raw cryo-EM images have been deposited in the Electron Microscopy Public Image Archive (<https://www.ebi.ac.uk/pdbe/emdb/empiar/>) under accession code EMPIAR-11917.

### **Acknowledgements**

Cryo-EM data were collected at Columbia Cryo-EM facility and the Simons Electron Microscopy Center (SEMC), with the assistance of staff from both SEMC and the Columbia University Cryo-Electron Microscopy Center. R. Grassucci and Z. Zhang from the Columbia Cryo-EM Center assisted with data collection. Some of the work was performed using equipment from the Center for Membrane Protein Production and Analysis (COMPPA; grant no. NIH P41 GM116799 to W.A. Hendrickson). We thank Dr. R. Bruni for the gift of BacMam-oxGFP vector. Z. Gong is partially supported by Endocrinology Training grant from NIH (grant number 5 T32 DK 7328-40).

### **Author contributions**

Z.G., Q.R.F. and W.A.H. designed the expression constructs. Z.G. performed gene cloning, protein purification, cAMP assay and MST measurement. B.K. and Z.G. carried out FSEC screens. S.C., Z.F. and Z.G. screened and optimized sample vitrification, and generated cryo-EM data. Single-particle cryo-EM data analysis was conducted by S.C., O.B.C., C.W., and Z.G. 3D Variability Analysis was carried out by Z. G. Model building was conducted by Z.G. The project was supervised by Q.R.F. and W.A.H. The manuscript was written by Z.G. and W.A.H with input from S.C., O.B.C., and Q.R.F.

**Extended Data Table 1. Cryo-EM data collection, refinement, and validation statistics**

<b>Composite/full map PDB</b>	<b>apo CeLGR (EMD-43735) (PDB 8W1Z)</b>
<b>Data collection and processing</b>	
<b>Magnification</b>	130,000
<b>Voltage (kV)</b>	300
<b>Electron exposure (e-/Å<sup>2</sup>)</b>	71
<b>Defocus range (µm)</b>	-1.2 to -2.0
<b>Pixel size (Å)</b>	1.06
<b>Symmetry imposed</b>	C1
<b>Initial particle images (no.)</b>	965,798
<b>Final particle images (no.)</b>	72,220
<b>Map resolution (Å)</b>	3.79
<b>FSC threshold</b>	0.143
<b>Refinement</b>	
<b>Initial model used</b>	AlphaFold
<b>Model resolution (Å)</b>	3.79
<b>Map sharpening B factor (Å<sup>2</sup>)</b>	86.2
<b>Model composition</b>	
<b>Protein residues</b>	1192
<b>Ligands</b>	NAG: 6, CLR: 1
<b>RMSD</b>	
<b>Bond length (Å)</b>	0.002
<b>Bond angles (°)</b>	0.570
<b>Validation</b>	
<b>MolProbity score</b>	2.48
<b>Clashscore</b>	9.26
<b>Poor rotamers (%)</b>	6.10
<b>CaBLAM outliers (%)</b>	2.39
<b>Ramachandran plot</b>	
<b>Favored (%)</b>	94.16
<b>Allowed (%)</b>	5.67
<b>Disallowed (%)</b>	0.17

# References

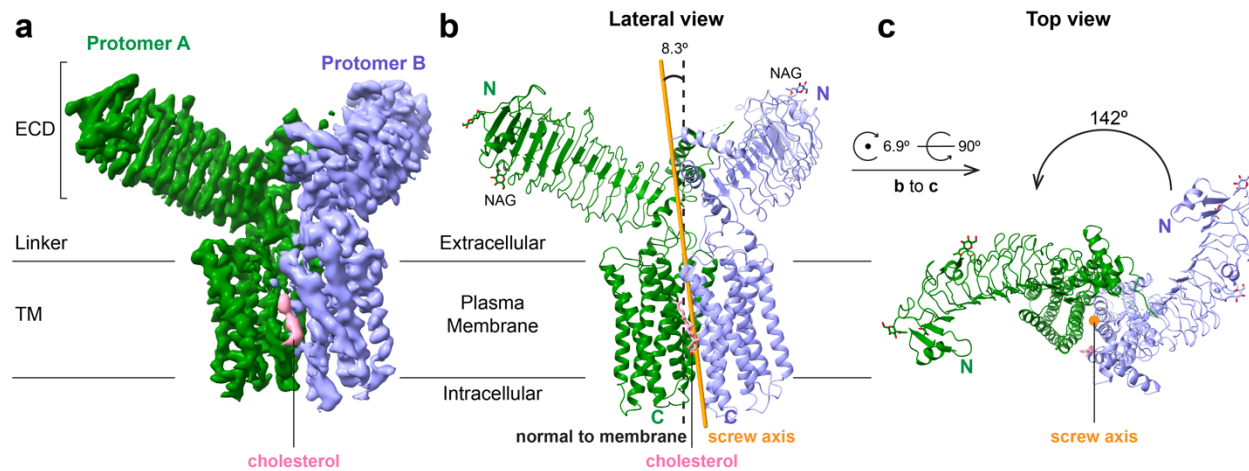
- 1 Pierce, J. G. & Parsons, T. F. Glycoprotein hormones: structure and function. *Annu Rev Biochem* **50**, 465-495, doi:10.1146/annurev.bi.50.070181.002341 (1981).
- 2 Querat, B. Unconventional Actions of Glycoprotein Hormone Subunits: A Comprehensive Review. *Front Endocrinol (Lausanne)* **12**, 731966, doi:10.3389/fendo.2021.731966 (2021).
- 3 Duan, J., Xu, P., Cheng, X., Mao, C., Croll, T., He, X., Shi, J., Luan, X., Yin, W., You, E., Liu, Q., Zhang, S., Jiang, H., Zhang, Y., Jiang, Y. & Xu, H. E. Structures of full-length glycoprotein hormone receptor signalling complexes. *Nature* **598**, 688-692, doi:10.1038/s41586-021-03924-2 (2021).
- 4 Duan, J., Xu, P., Luan, X., Ji, Y., He, X., Song, N., Yuan, Q., Jin, Y., Cheng, X., Jiang, H., Zheng, J., Zhang, S., Jiang, Y. & Xu, H. E. Hormone- and antibody-mediated activation of the thyrotropin receptor. *Nature*, doi:10.1038/s41586-022-05173-3 (2022).
- 5 Faust, B., Billesbolle, C. B., Suomivuori, C. M., Singh, I., Zhang, K., Hoppe, N., Pinto, A. F. M., Diedrich, J. K., Muftuoglu, Y., Szkudlinski, M. W., Saghatelian, A., Dror, R. O., Cheng, Y. & Manglik, A. Autoantibody mimicry of hormone action at the thyrotropin receptor. *Nature*, doi:10.1038/s41586-022-05159-1 (2022).
- 6 Duan, J., Xu, P., Zhang, H., Luan, X., Yang, J., He, X., Mao, C., Shen, D. D., Ji, Y., Cheng, X., Jiang, H., Jiang, Y., Zhang, S., Zhang, Y. & Xu, H. E. Mechanism of hormone and allosteric agonist mediated activation of follicle stimulating hormone receptor. *Nat Commun* **14**, 519, doi:10.1038/s41467-023-36170-3 (2023).
- 7 Rivero-Muller, A., Chou, Y. Y., Ji, I., Lajic, S., Hanyaloglu, A. C., Jonas, K., Rahman, N., Ji, T. H. & Huhtaniemi, I. Rescue of defective G protein-coupled receptor function in vivo by intermolecular cooperation. *Proc Natl Acad Sci U S A* **107**, 2319-2324, doi:10.1073/pnas.0906695106 (2010).
- 8 Urizar, E., Montanelli, L., Loy, T., Bonomi, M., Swillens, S., Gales, C., Bouvier, M., Smits, G., Vassart, G. & Costagliola, S. Glycoprotein hormone receptors: link between receptor homodimerization and negative cooperativity. *EMBO J* **24**, 1954-1964, doi:10.1038/sj.emboj.7600686 (2005).
- 9 Boutin, A., Krieger, C. C., Marcus-Samuels, B., Klubo-Gwiedzinska, J., Neumann, S. & Gershengorn, M. C. TSH Receptor Homodimerization in Regulation of cAMP Production in Human Thyrocytes in vitro. *Front Endocrinol (Lausanne)* **11**, 276, doi:10.3389/fendo.2020.00276 (2020).
- 10 Allen, M. D., Neumann, S. & Gershengorn, M. C. Occupancy of both sites on the thyrotropin (TSH) receptor dimer is necessary for phosphoinositide signaling. *Faseb J* **25**, 3687-3694, doi:10.1096/fj.11-188961 (2011).
- 11 Kenis, S., Istiban, M. N., Van Damme, S., Vandeweyer, E., Watteyne, J., Schoofs, L. & Beets, I. Ancestral glycoprotein hormone-receptor pathway controls growth in *C. elegans*. *Front Endocrinol (Lausanne)* **14**, 1200407, doi:10.3389/fendo.2023.1200407 (2023).
- 12 Gong, Z., Wang, W., El Omari, K., Lebedev, A. A., Clarke, O. B. & Hendrickson, W. A. Crystal structure of LGR ligand alpha2/beta5 from *Caenorhabditis elegans* with implications for the evolution of glycoprotein hormones. *Proc Natl Acad Sci U S A* **120**, e2218630120, doi:10.1073/pnas.2218630120 (2023).
- 13 Fan, Q. R. & Hendrickson, W. A. Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* **433**, 269-277, doi:10.1038/nature03206 (2005).
- 14 Jiang, X., Liu, H., Chen, X., Chen, P. H., Fischer, D., Sriraman, V., Yu, H. N., Arkinstall, S. & He, X. Structure of follicle-stimulating hormone in complex with the entire ectodomain of its receptor. *Proc Natl Acad Sci U S A* **109**, 12491-12496, doi:10.1073/pnas.1206643109 (2012).
- 15 Vassart, G., Pardo, L. & Costagliola, S. A molecular dissection of the glycoprotein hormone receptors. *Trends Biochem Sci* **29**, 119-126, doi:10.1016/j.tibs.2004.01.006 (2004).
- 16 Latif, R., Graves, P. & Davies, T. F. Oligomerization of the human thyrotropin receptor: fluorescent protein-tagged hTSHR reveals post-translational complexes. *J Biol Chem* **276**, 45217-45224, doi:10.1074/jbc.M103727200 (2001).
- 17 Guan, R. B., Feng, X. Y., Wu, X. Q., Zhang, M. L., Zhang, X. S., Hébert, T. E. & Segaloff, D. L. Bioluminescence Resonance Energy Transfer Studies Reveal Constitutive Dimerization of the Human Lutropin Receptor and a Lack of Correlation between Receptor Activation and the

- Propensity for Dimerization. *Journal of Biological Chemistry* **284**, 7483-7494, doi:10.1074/jbc.M809150200 (2009).
- 18 Zhang, M., Feng, X., Guan, R., Hebert, T. E. & Segaloff, D. L. A cell surface inactive mutant of the human lutropin receptor (hLHR) attenuates signaling of wild-type or constitutively active receptors via heterodimerization. *Cell Signal* **21**, 1663-1671, doi:10.1016/j.cellsig.2009.07.003 (2009).
- 19 Jonas, K. C., Fanelli, F., Huhtaniemi, I. T. & Hanyaloglu, A. C. Single molecule analysis of functionally asymmetric G protein-coupled receptor (GPCR) oligomers reveals diverse spatial and structural assemblies. *J Biol Chem* **290**, 3875-3892, doi:10.1074/jbc.M114.622498 (2015).
- 20 Zoenen, M., Urizar, E., Swillens, S., Vassart, G. & Costagliola, S. Evidence for activity-regulated hormone-binding cooperativity across glycoprotein hormone receptor homomers. *Nat Commun* **3**, 1007, doi:10.1038/ncomms1991 (2012).
- 21 Neumann, S., Malik, S. S., Marcus-Samuels, B., Eliseeva, E., Jang, D., Klubo-Gwiedzinska, J., Krieger, C. C. & Gershengorn, M. C. Thyrotropin Causes Dose-dependent Biphasic Regulation of cAMP Production Mediated by G(s) and G(i/o) Proteins. *Mol Pharmacol* **97**, 2-8, doi:10.1124/mol.119.117382 (2020).
- 22 Osuga, Y., Hayashi, M., Kudo, M., Conti, M., Kobilka, B. & Hsueh, A. J. Co-expression of defective luteinizing hormone receptor fragments partially reconstitutes ligand-induced signal generation. *J Biol Chem* **272**, 25006-25012, doi:10.1074/jbc.272.40.25006 (1997).
- 23 Ji, I., Lee, C., Song, Y., Conn, P. M. & Ji, T. H. Cis- and trans-activation of hormone receptors: the LH receptor. *Mol Endocrinol* **16**, 1299-1308, doi:10.1210/mend.16.6.0852 (2002).
- 24 Ji, I., Lee, C., Jeoung, M., Koo, Y., Sievert, G. A. & Ji, T. H. Trans-activation of mutant follicle-stimulating hormone receptors selectively generates only one of two hormone signals. *Mol Endocrinol* **18**, 968-978, doi:10.1210/me.2003-0443 (2004).
- 25 Park, J., Fu, Z., Frangaj, A., Liu, J., Mosyak, L., Shen, T., Slavkovich, V. N., Ray, K. M., Taura, J., Cao, B., Geng, Y., Zuo, H., Kou, Y., Grassucci, R., Chen, S., Liu, Z., Lin, X., Williams, J. P., Rice, W. J., Eng, E. T., Huang, R. K., Soni, R. K., Kloss, B., Yu, Z., Javitch, J. A., Hendrickson, W. A., Slesinger, P. A., Quick, M., Graziano, J., Yu, H., Fiehn, O., Clarke, O. B., Frank, J. & Fan, Q. R. Structure of human GABAB receptor in an inactive state. *Nature* **584**, 304-309, doi:10.1038/s41586-020-2452-0 (2020).
- 26 Shen, C., Mao, C., Xu, C., Jin, N., Zhang, H., Shen, D. D., Shen, Q., Wang, X., Hou, T., Chen, Z., Rondard, P., Pin, J. P., Zhang, Y. & Liu, J. Structural basis of GABA(B) receptor-G(i) protein coupling. *Nature* **594**, 594-598, doi:10.1038/s41586-021-03507-1 (2021).
- 27 Dos Santos, S., Mazan, S., Venkatesh, B., Cohen-Tannoudji, J. & Quérat, B. Emergence and evolution of the glycoprotein hormone and neurotrophin gene families in vertebrates. *BMC Evol Biol* **11**, 332, doi:10.1186/1471-2148-11-332 (2011).
- 28 Van Hiel, M. B., Vandersmissen, H. P., Van Loy, T. & Vanden Broeck, J. An evolutionary comparison of leucine-rich repeat containing G protein-coupled receptors reveals a novel LGR subtype. *Peptides* **34**, 193-200, doi:10.1016/j.peptides.2011.11.004 (2012).
- 29 Kudo, M., Chen, T., Nakabayashi, K., Hsu, S. Y. & Hsueh, A. J. The nematode leucine-rich repeat-containing, G protein-coupled receptor (LGR) protein homologous to vertebrate gonadotropin and thyrotropin receptors is constitutively active in mammalian cells. *Mol Endocrinol* **14**, 272-284, doi:10.1210/mend.14.2.0422 (2000).
- 30 Park, J. I., Semyonov, J., Chang, C. L. & Hsu, S. Y. Conservation of the heterodimeric glycoprotein hormone subunit family proteins and the LGR signaling system from nematodes to humans. *Endocrine* **26**, 267-276, doi:10.1385/ENDO:26:3:267 (2005).
- 31 Hsu, S. Y., Nakabayashi, K. & Bhalla, A. Evolution of glycoprotein hormone subunit genes in bilateral metazoa: identification of two novel human glycoprotein hormone subunit family genes, GPA2 and GPB5. *Mol Endocrinol* **16**, 1538-1551, doi:10.1210/mend.16.7.0871 (2002).
- 32 Nakabayashi, K., Matsumi, H., Bhalla, A., Bae, J., Mosselman, S., Hsu, S. Y. & Hsueh, A. J. Thyrostimulin, a heterodimer of two new human glycoprotein hormone subunits, activates the thyroid-stimulating hormone receptor. *J Clin Invest* **109**, 1445-1452, doi:10.1172/JCI14340 (2002).
- 33 Loosfelt, H., Pichon, C., Jolivet, A., Misrahi, M., Caillou, B., Jamous, M., Vannier, B. & Milgrom, E. Two-subunit structure of the human thyrotropin receptor. *Proc Natl Acad Sci U S A* **89**, 3765-3769 (1992).

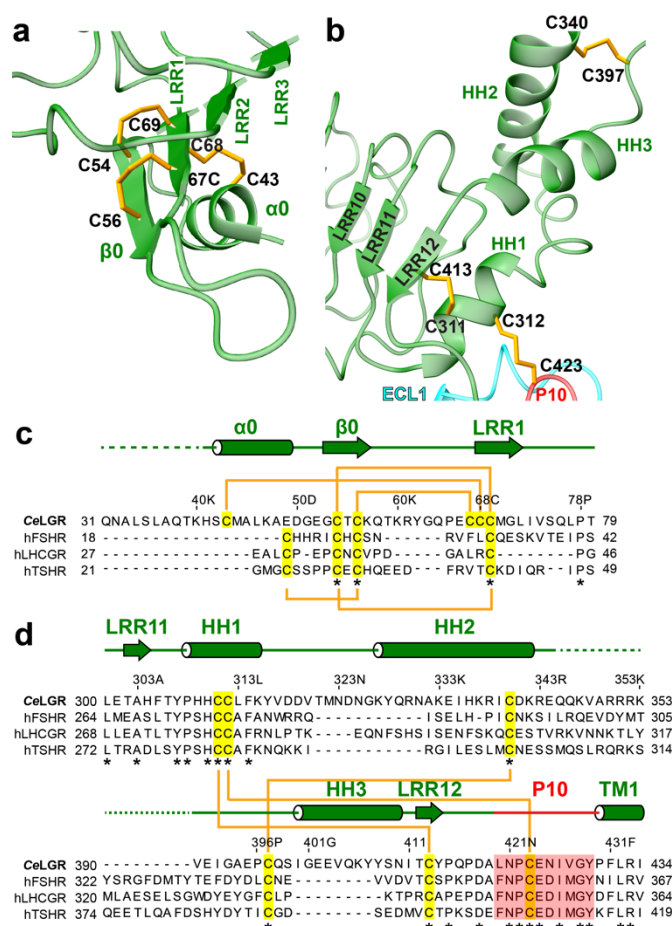
- 34 Chazenbalk, G. D., Tanaka, K., Nagayama, Y., Kakinuma, A., Jaume, J. C., McLachlan, S. M. & Rapoport, B. Evidence that the thyrotropin receptor ectodomain contains not one, but two, cleavage sites. *Endocrinology* **138**, 2893-2899, doi:10.1210/endo.138.7.5259 (1997).
- 35 Lomize, A. L., Todd, S. C. & Pogozheva, I. D. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Sci* **31**, 209-220, doi:10.1002/pro.4219 (2022).
- 36 Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*, 310-322 (2002).
- 37 Kobe, B. & Kajava, A. V. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struc Biol* **11**, 725-732, doi:10.1016/S0959-440x(01)00266-4 (2001).
- 38 Ng, A. C., Eisenberg, J. M., Heath, R. J., Huett, A., Robinson, C. M., Nau, G. J. & Xavier, R. J. Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4631-4638, doi:10.1073/pnas.1000093107 (2011).
- 39 Mueller, S., Jaeschke, H., Gunther, R. & Paschke, R. The hinge region: an important receptor component for GPCR function. *Trends Endocrinol Metab* **21**, 111-122, doi:10.1016/j.tem.2009.09.001 (2010).
- 40 Mueller, S., Kleinau, G., Jaeschke, H., Neumann, S., Krause, G. & Paschke, R. Significance of ectodomain cysteine boxes 2 and 3 for the activation mechanism of the thyroid-stimulating hormone receptor. *J Biol Chem* **281**, 31638-31646, doi:10.1074/jbc.M604770200 (2006).
- 41 Nakabayashi, K., Kudo, M., Kobilka, B. & Hsueh, A. J. Activation of the luteinizing hormone receptor following substitution of Ser-277 with selective hydrophobic residues in the ectodomain hinge region. *J Biol Chem* **275**, 30264-30271, doi:10.1074/jbc.M005568200 (2000).
- 42 Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702, doi:10.1016/j.jsb.2021.107702 (2021).
- 43 Dos Santos, S., Bardet, C., Bertrand, S., Escriva, H., Habert, D. & Querat, B. Distinct expression patterns of glycoprotein hormone- $\alpha$ 2 and - $\beta$ 5 in a basal chordate suggest independent developmental functions. *Endocrinology* **150**, 3815-3822, doi:10.1210/en.2008-1743 (2009).
- 44 Nagayama, Y., Takeshita, A., Luo, W., Ashizawa, K., Yokoyama, N. & Nagataki, S. High affinity binding of thyrotropin (TSH) and thyroid-stimulating autoantibody for the TSH receptor extracellular domain. *Thyroid* **4**, 155-159, doi:10.1089/thy.1994.4.155 (1994).
- 45 Park, J. H., Ohshima, S., Tani, T. & Ohshima, Y. Structure and expression of the gsa-1 gene encoding a G protein  $\alpha$ (s) subunit in *C. elegans*. *Gene* **194**, 183-190, doi:10.1016/s0378-1119(97)00122-4 (1997).
- 46 Agwuegbo, U. T., Colley, E., Albert, A. P., Butnev, V. Y., Bousfield, G. R. & Jonas, K. C. Differential FSH Glycosylation Modulates FSHR Oligomerization and Subsequent cAMP Signaling. *Front Endocrinol (Lausanne)* **12**, 765727, doi:10.3389/fendo.2021.765727 (2021).
- 47 Gusach, A., García-Nafria, J. & Tate, C. G. New insights into GPCR coupling and dimerisation from cryo-EM structures. *Curr Opin Struc Biol* **80**, doi:10.1016/j.sbi.2023.102574 (2023).
- 48 Seven, A. B., Barros-Alvarez, X., de Lapeyriere, M., Papasergi-Scott, M. M., Robertson, M. J., Zhang, C., Nwokonko, R. M., Gao, Y., Meyerowitz, J. G., Rocher, J. P., Schelshorn, D., Kobilka, B. K., Mathiesen, J. M. & Skiniotis, G. G-protein activation by a metabotropic glutamate receptor. *Nature* **595**, 450-454, doi:10.1038/s41586-021-03680-3 (2021).
- 49 Zuo, H., Park, J., Frangaj, A., Ye, J., Lu, G., Manning, J. J., Asher, W. B., Lu, Z., Hu, G. B., Wang, L., Mendez, J., Eng, E., Zhang, Z., Lin, X., Grassucci, R., Hendrickson, W. A., Clarke, O. B., Javitch, J. A., Conigrave, A. D. & Fan, Q. R. Promiscuous G-protein activation by the calcium-sensing receptor. *Nature* **629**, 481-488, doi:10.1038/s41586-024-07331-1 (2024).
- 50 Du, J., Wang, D. J., Fan, H. C., Xu, C. J., Tai, L. H., Lin, S. L., Han, S., Tan, Q. X., Wang, X. W., Xu, T., Zhang, H., Chu, X. J., Yi, C. Y., Liu, P., Wang, X. M., Zhou, Y., Pin, J. P., Rondard, P., Liu, H., Liu, J. F., Sun, F., Wu, B. L. & Zhao, Q. Structures of human mGlu2 and mGlu7 homo- and heterodimers. *Nature* **594**, 589+, doi:10.1038/s41586-021-03641-w (2021).
- 51 Park, J., Zuo, H., Frangaj, A., Fu, Z., Yen, L. Y., Zhang, Z., Mosyak, L., Slavkovich, V. N., Liu, J., Ray, K. M., Cao, B., Vallese, F., Geng, Y., Chen, S., Grassucci, R., Dandey, V. P., Tan, Y. Z., Eng, E., Lee, Y., Kloss, B., Liu, Z., Hendrickson, W. A., Potter, C. S., Carragher, B., Graziano, J., Conigrave, A. D., Frank, J., Clarke, O. B. & Fan, Q. R. Symmetric activation and modulation of



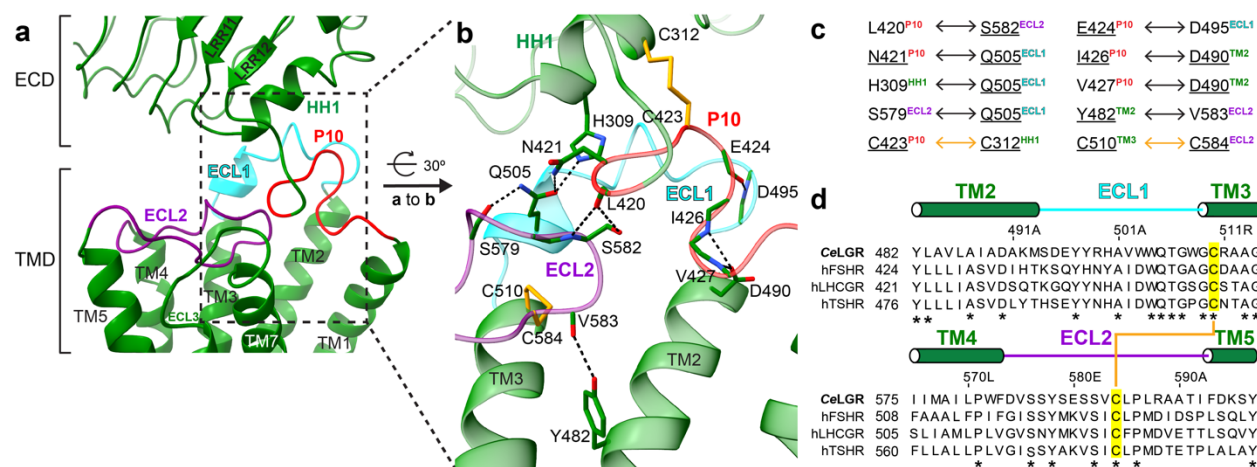
- the human calcium-sensing receptor. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2115849118 (2021).
- 52 Mancia, F., Assur, Z., Herman, A. G., Siegel, R. & Hendrickson, W. A. Ligand sensitivity in dimeric associations of the serotonin 5HT<sub>2c</sub> receptor. *EMBO Rep* **9**, 363-369, doi:10.1038/embor.2008.27 (2008).
- 53 Cole, L. A. hCG, the wonder of today's science. *Reprod Biol Endocrinol* **10**, 24, doi:10.1186/1477-7827-10-24 (2012).
- 54 Koshland, D. E., Jr. The structural basis of negative cooperativity: receptors and enzymes. *Curr Opin Struct Biol* **6**, 757-761, doi:10.1016/s0959-440x(96)80004-2 (1996).
- 55 Costantini, L. M., Baloban, M., Markwardt, M. L., Rizzo, M. A., Guo, F., Verkhusha, V. V. & Snapp, E. L. A palette of fluorescent proteins optimized for diverse cellular environments. *Nat Commun* **6**, 7670, doi:10.1038/ncomms8670 (2015).
- 56 Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precystallization screening of integral membrane proteins. *Structure* **14**, 673-681, doi:10.1016/j.str.2006.01.013 (2006).
- 57 Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856-2860, doi:10.1038/nprot.2006.468 (2006).
- 58 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
- 59 Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).
- 60 Terwilliger, T. C., Poon, B. K., Afonine, P. V., Schlicksup, C. J., Croll, T. I., Millan, C., Richardson, J. S., Read, R. J. & Adams, P. D. Improved AlphaFold modeling with implicit experimental information. *Nat Methods* **19**, 1376-1382, doi:10.1038/s41592-022-01645-6 (2022).
- 61 Casanal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci* **29**, 1069-1078, doi:10.1002/pro.3791 (2020).
- 62 Croll, T. I.: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D* **74**, 519-530, doi:10.1107/S2059798318002425 (2018).
- 63 Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A. & Adams, P. D. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).
- 64 Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).
- 65 Zwier, J. M., Bazin, H., Lamarque, L. & Mathis, G. Luminescent lanthanide cryptates: from the bench to the bedside. *Inorg Chem* **53**, 1854-1866, doi:10.1021/ic402234k (2014).



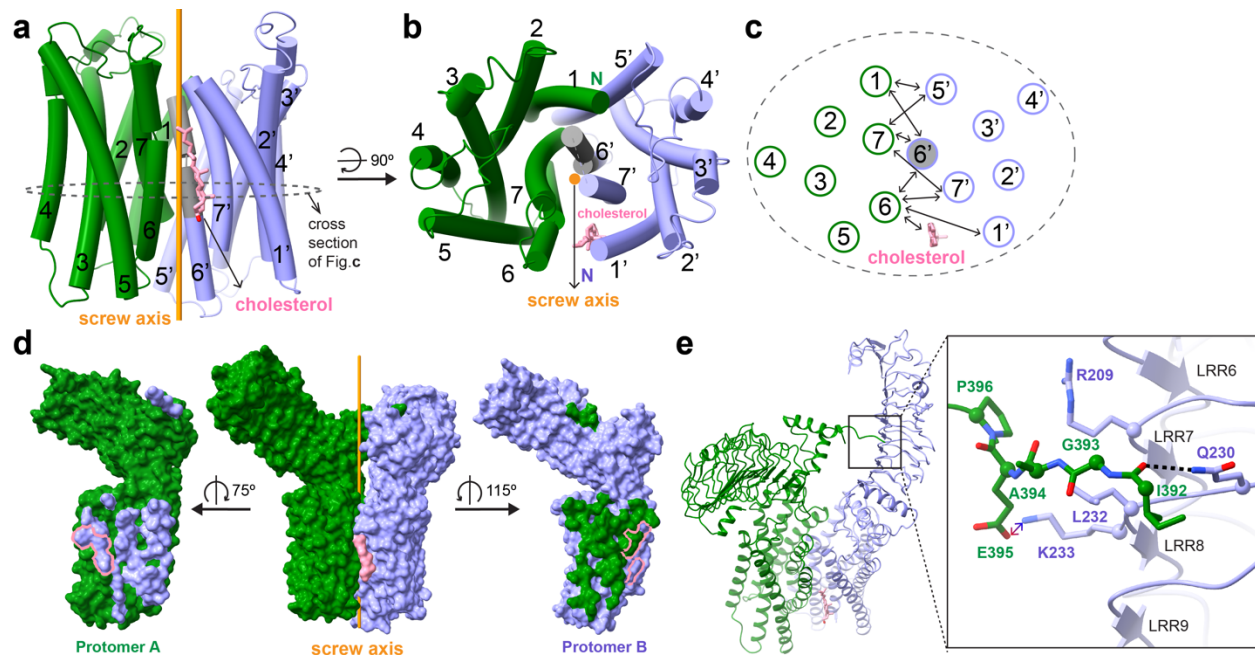
**Fig. 1. Cryo-EM structure of the CeLGR homodimer in an unliganded apo state.** (a) Cryo-EM density map of the CeLGR homodimer at an overall resolution of 3.79 Å. (b) Lateral view of a ribbon diagram of the atomic model fitted to the cryo-EM density for the CeLGR structure (same view as in a). The orange stick is the screw axis relating protomer B to protomer A. N-acetylglucosamines (NAGs) and cholesterol are shown as sticks. (c) Top view of the CeLGR structure in ribbon diagram, looking directly down along the screw axis (orange stick). Coloring has protomer A (green), protomer B (light purple) and assigned cholesterol (pink).



**Fig. 2. Structure and sequence alignment of the N-terminal cysteine-rich region (a, c) and the C-terminal cysteine-rich hinge region (b, d) of CeLGR protomer A.** For the ribbon diagram in **a** and **b**, *N*-acetylglucosamines (NAGs) and CeLGR protomer B are not displayed for clarity. Structural elements and cysteine residues are labeled. The P10 region is colored in red. For the sequence alignment in **c** and **d**, the secondary structural elements are indicated by cylinders for  $\alpha$  helices and arrows for  $\beta$  strands. Cysteine residues are highlighted with yellow background, and disulfide bridges are indicated by orange connections. The conserved residues are underscored by asterisks (\*). Regions that are unstructured in the model are indicated with green dash lines.

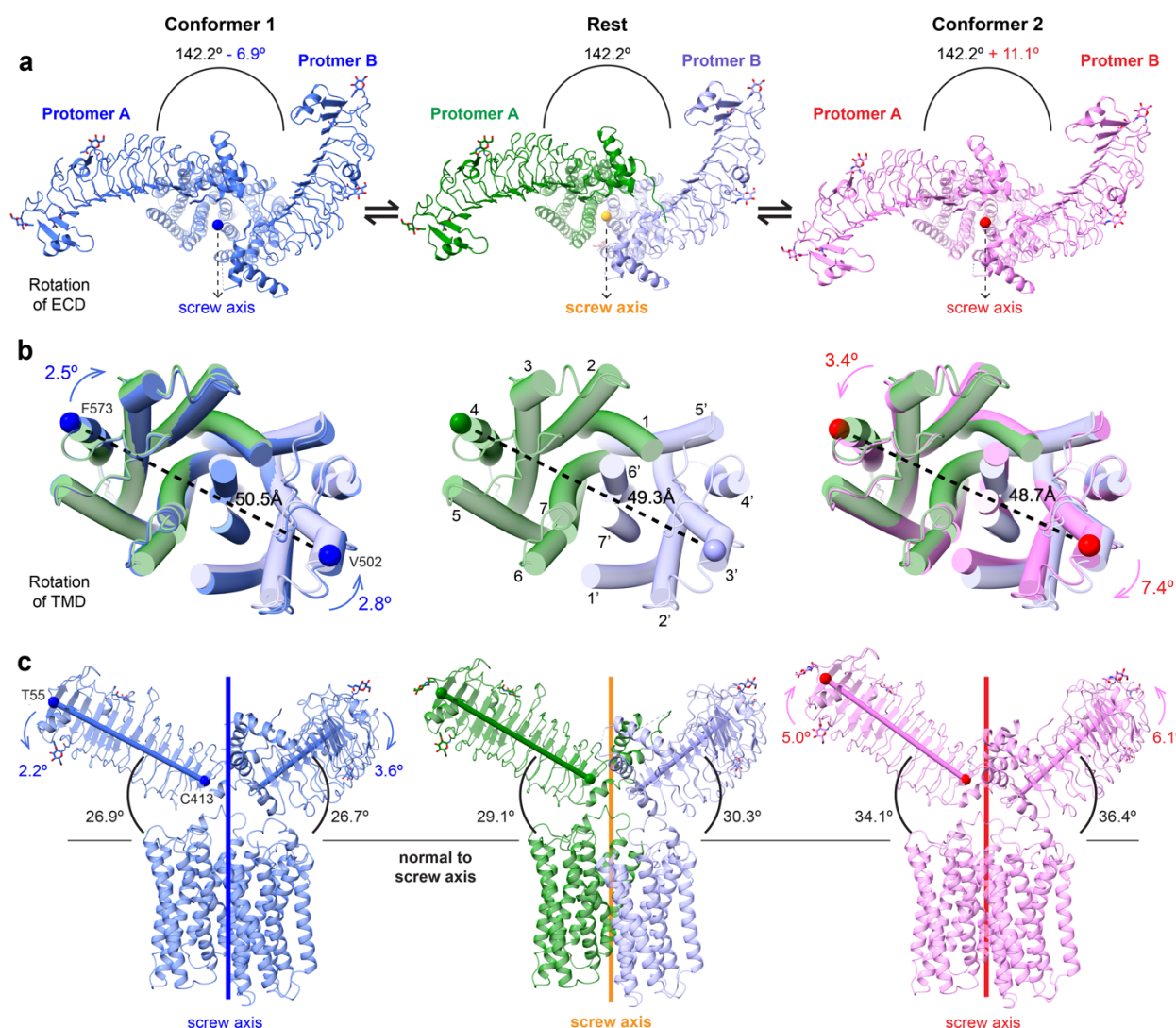


**Fig. 3. Interactions between the ECD and TMD of CeLGR protomer A.** (a) P10 in the CCR hinge region of CeLGR plays a major role in linking the ECD with the TMD through interactions with HH1, ECL1, ECL2 and TM2. P10, ECL1 and ECL2 are colored in red, cyan and purple, respectively. (b) A magnified illustration of the region boxed in **a** to show atomic interactions between HH1, P10, TM2, ECL1, TM3 and ECL2. The ECL3 and TM7 are not shown for clarity. Hydrogen bonds are shown in black dashed lines. Disulfide bridges are represented as orange sticks. The carbon, oxygen and nitrogen atoms are shown in green, red and blue, respectively. (c) A summary of interactions shown in **b**. Conserved residues among CeLGR and hGpHRs are underscored. Hydrogen bonds and disulfide bridges are indicated by black and orange arrows, respectively. (d) Sequence alignments in the vicinity of ECL1 and ECL2 comparing CeLGR with the hGpHRs. Cysteine residues are highlighted with yellow background, and disulfide bridges are indicated by orange connections. The conserved residues are underscored by asterisks (\*).

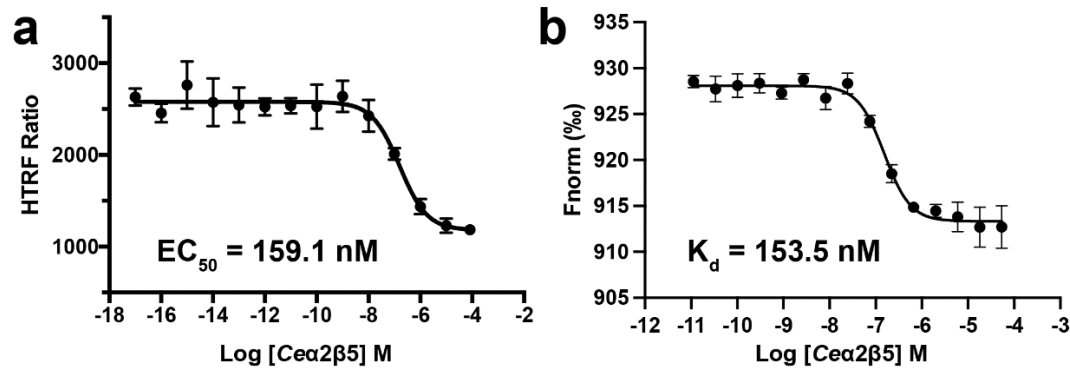


**Fig. 4. Dimer interface between CeLGR protomers A and B.** (a) Dimer interface between the TMDs of CeLGR protomer A (green) and B (light purple). The transmembrane helices (TMs) are illustrated as cylinders and numbered. Half of TM6 in protomer B is shown in grey because its density could not be resolved for modeling. The orange stick is the screw axis from protomer B to protomer A. The cholesterol is shown as sticks (pink). (b) View of the TMD dimer interface from outside the cell. (c) View of a cross section of the homodimer as circled in a, but in the same view orientation as b. (d) Molecular surface of the homodimer (center) and then of protomer A (left) and protomer B (right) after opening by rotations as indicated and left with the imprint of the surface buried into the interface as colored by the partner protomers or by the assigned cholesterol outlined in pink. (e) Direct interaction between the CCR hinge region of protomer A and the ECD of protomer B. A hydrogen bond is shown as a black dashed line, and an ionic interaction is indicated by a double-headed arrow colored as for the contacting charged atoms.

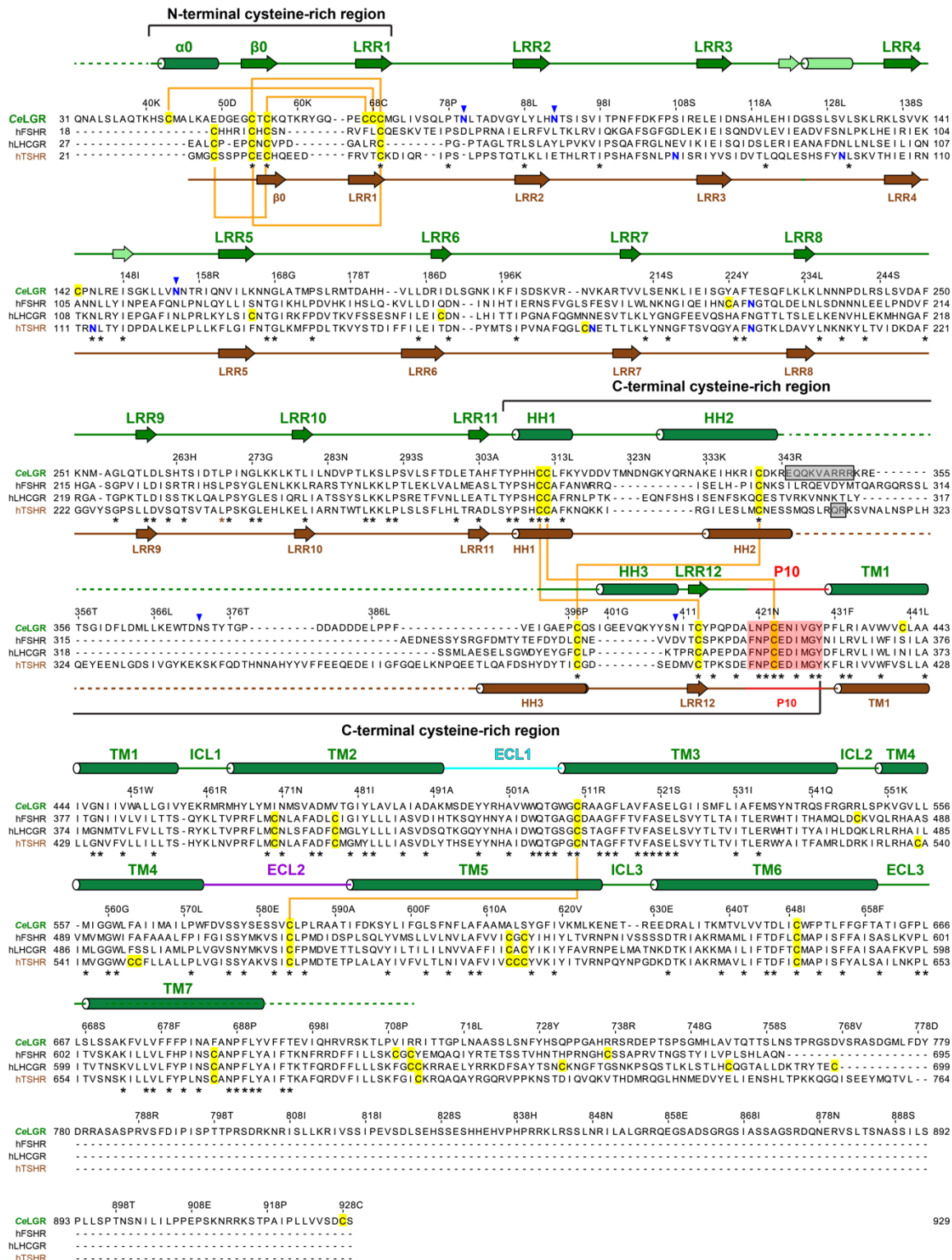




**Fig. 5. Fluctuational dynamics of the apo-state CeLGR homodimer.** The rest state model is in the center column (colored as in Fig. 1) and extreme conformers 1 (blue) and 2 (pink) are in the columns at left and right, respectively. (a) ECDs as viewed down the indicated screw axes with interprotomer rotations indicated. (b) TMDs as viewed into the membrane from extracellular side with TMs drawn as cylinders. The 'rest' state model (center) is superimposed on the extreme conformers (1, left; 2, right) with excursion rotations indicated by directed arrows. Measures of expansion and contraction are given as distances between markers Cα 573 in protomer A and Cα 502 in protomer B. (c) Side views of the CeLGR homodimer as in Fig. 1b. The angles between the ECD axis indicator (Cα 55 and Cα 413) and the plane normal to the screw axis are given for each chain, and excursion rotations are indicated by directed arrows.



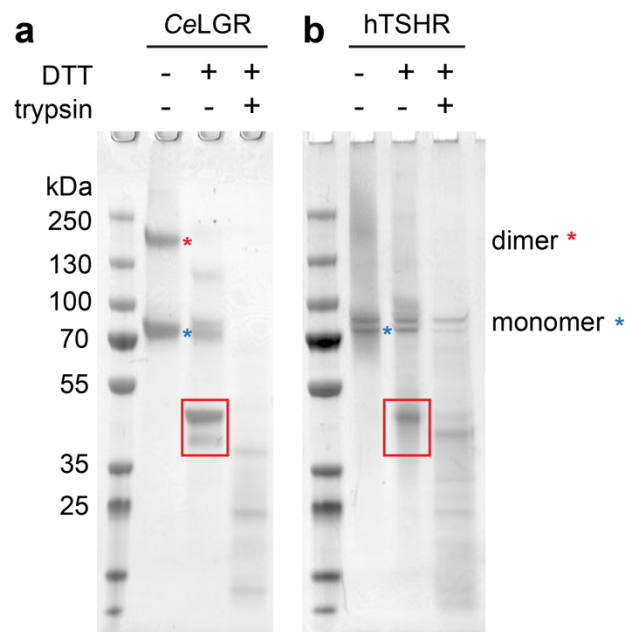
**Fig. 6. Activating interactions of Cea2β5 with CeLGR in vitro.** (a) The EC<sub>50</sub> for Cea2β5 activation of CeLGR was measured using a cAMP-based competitive assay in HEK 293S cells. The specific signal, homogenous time resolved fluorescence (HTRF) ratio, is inversely proportional to the concentration of native cAMP produced in the sample upon addition of Cea2β5 ligand. Data points (mean ± s.d) are plotted for measurements from an experiment of n=3 biological replicates. (b) The equilibrium dissociation constant (K<sub>d</sub>) of the Cea2β5:CeLGR complex was determined using MST. Data points (mean ± s.d.) are plotted for measurements from an experiment of n=4 biological replicates (see Methods for details).



## Structure-based Percent Identity Matrix

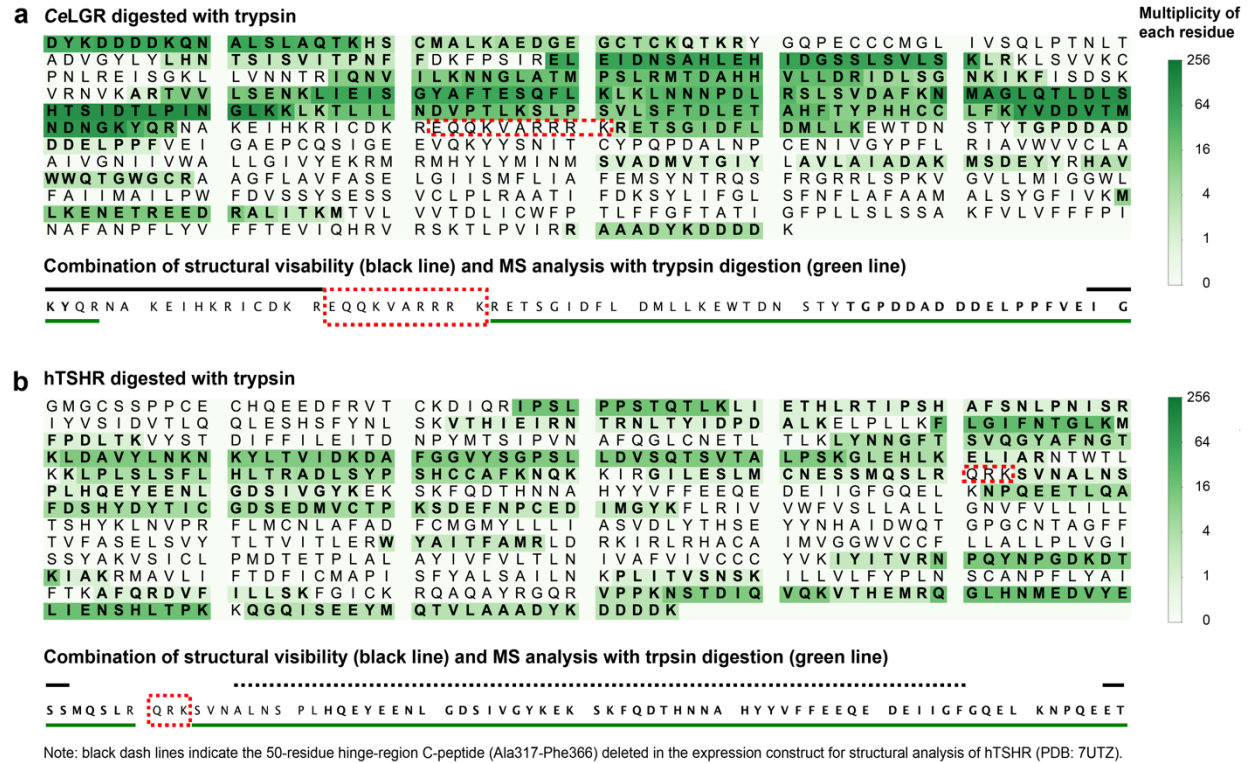
	ECD + TMD + Helix8					ECD					TMD + Helix8				
<b>CeLGR</b>	100.0					100.0					100.0				
<i>Dm</i> LGR-1	30.2	100.0				24.5	100.0				34.5	100.0			
hFSHR	30.5	36.7	100.0			30.3	24.9	100.0			33.3	49.5	100.0		
hLHCGR	30.0	40.0	56.9	100.0		26.7	27.7	45.7	100.0		34.1	52.3	69.6	100.0	
hTSHR	30.5	37.3	52.3	55.5	100.0	26.2	25.5	39.2	42.6	100.0	35.5	52.7	68.2	71.0	100.0

**Extended Data Fig. 1.** Structure-based sequence alignment of LGRs. The secondary structural elements are indicated by cylinders for  $\alpha$  helices and arrows for  $\beta$  strands, green for CeLGR and brown for hTSHR. Strands at the ECD repeats are in dark green and elements at the convex face are in light green. Potential N-linked glycosylation sites of CeLGR as predicted by the NetNGlyc-1.0 server are indicated with blue inverse triangles. N-linked glycosylation sites confirmed in solved structures are highlighted in blue letters: FSHR (PDBid 1XWD) and TSHR (PDBid 3G04). No glycosylation site was reported in the structure of LHCGR (PDBid 7FIH). Cysteine residues are highlighted with yellow background, and disulfide pairs are indicated by orange connections. Residues in P10 are highlighted with red background. Residues subject to proteolytic removal are indicated with grey background. Conserved residues are underscored by asterisks (\*).

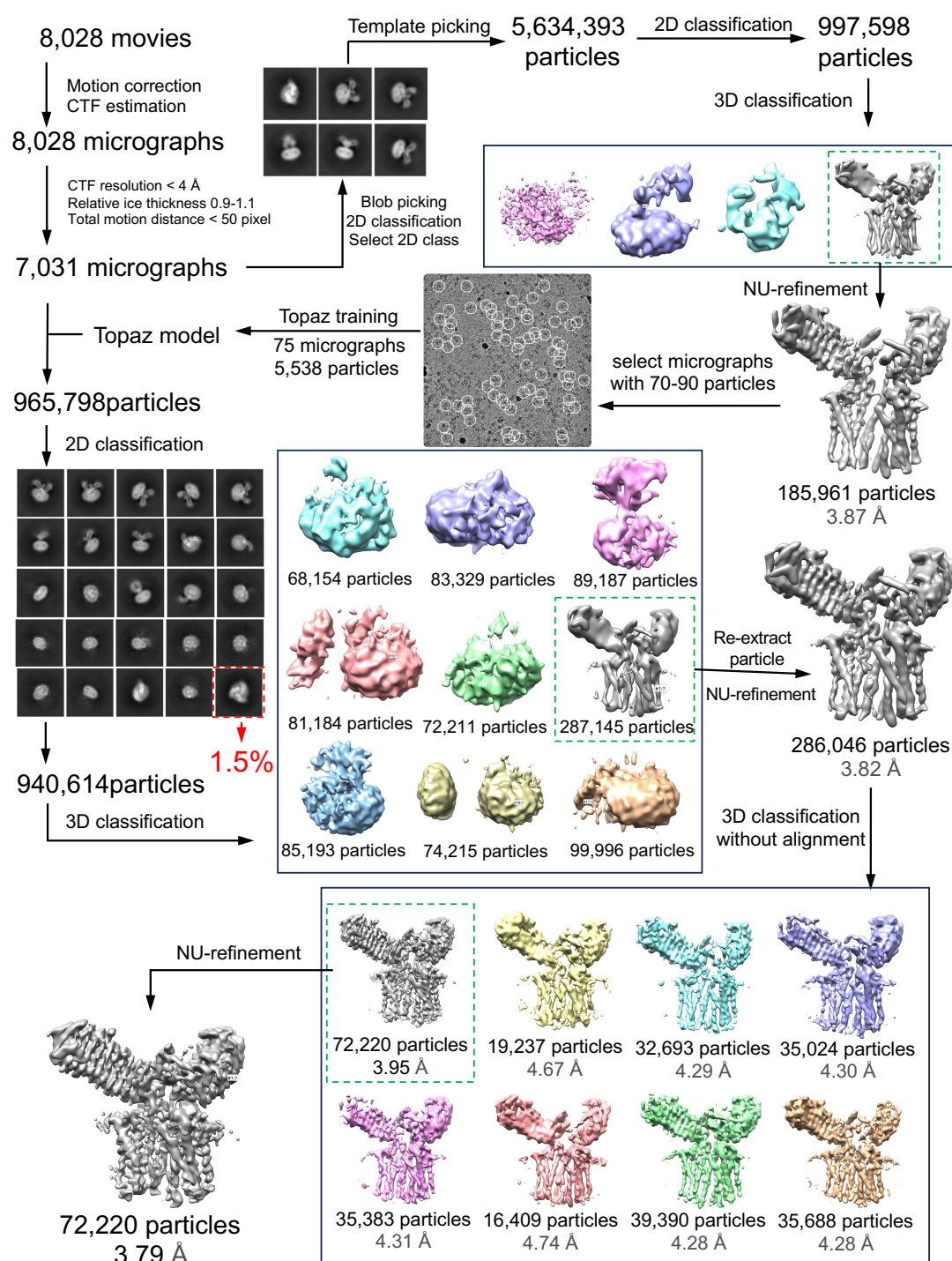


**Extended Data Fig. 2. SDS-PAGE analysis of purified receptor proteins. (a) CeLGR.** CeLGR appears as a monomer band and a dimer band in the absence of reducing agent dithiothreitol (DTT). Two new bands corresponding to lower molecular weight appear for CeLGR in the presence of DTT. After limited trypsin digestion, the residual monomer band is almost entirely displaced to smaller fragment bands. **(b) hTSHR.** The putative 'dimer' portion of the gel for hTSHR in non-reducing conditions is diffuse. A new band, similar to the one for CeLGR, plus additional diffuse densities appear for hTSHR in DTT. After limited trypsin digestion under the chosen conditions, residual monomer density is much reduced and new densities appear for fragments. The bands excised from the gel for analysis by proteolytic mass spectrometry are boxed in red.

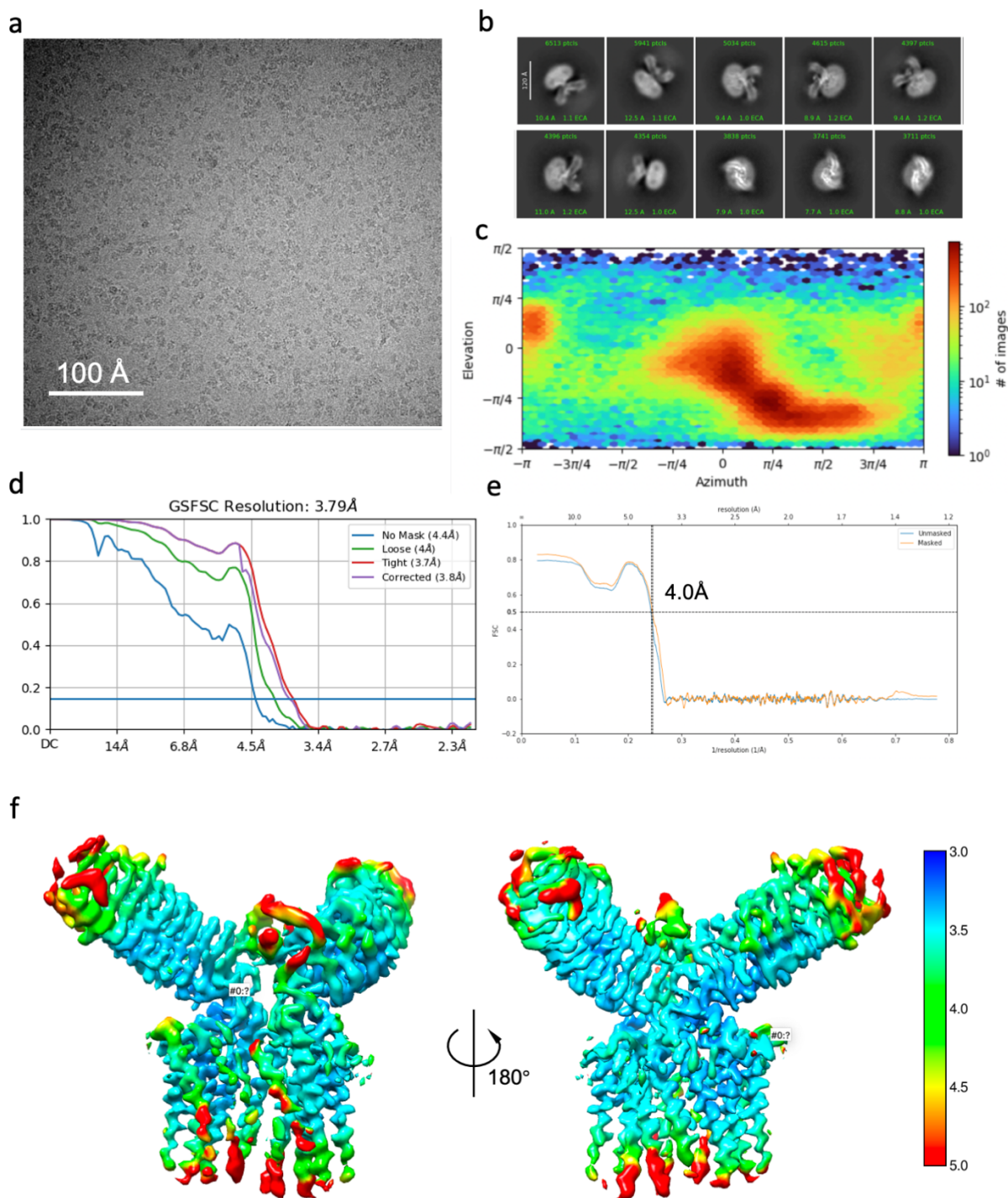




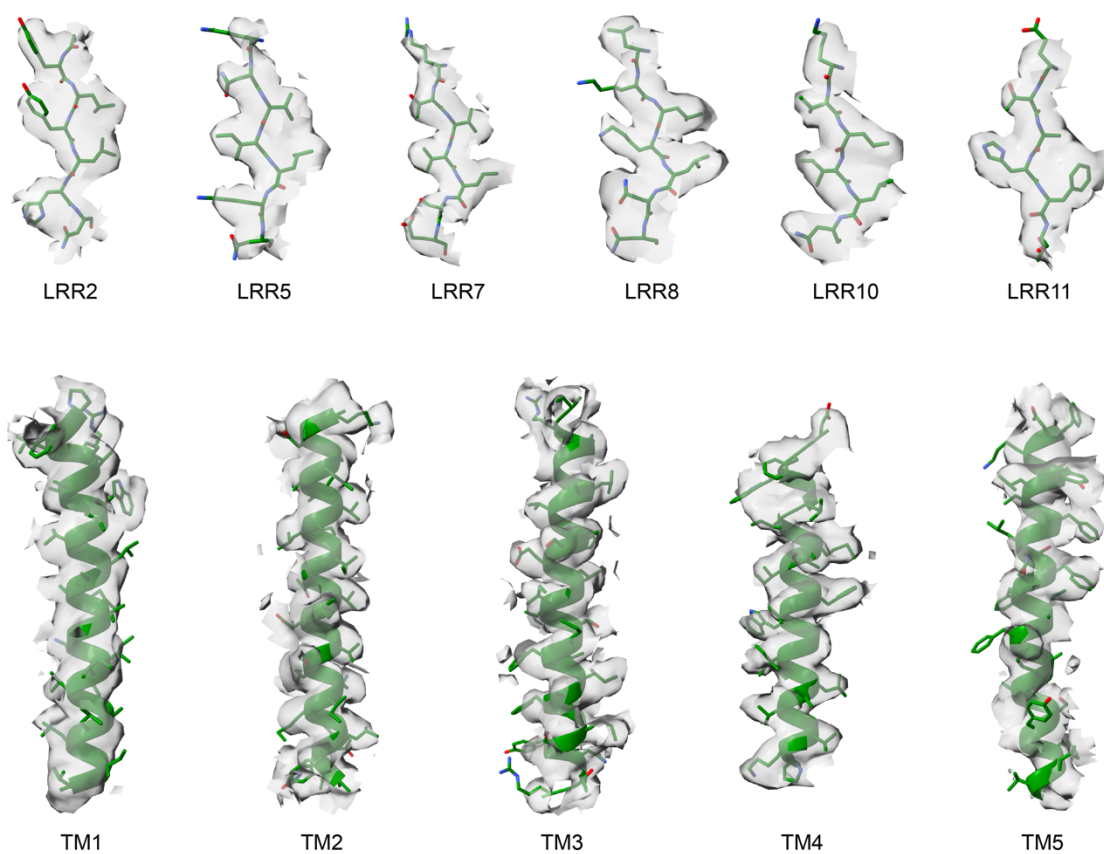
**Extended Data Fig. 3. Delimitation of the cleavage site(s) with a combination of proteolytic mass spectrometry and structural coverage. (a) CeLGR. (b) hTSHR.** For each receptor and each protease, the multiplicity of tryptic peptides is represented with gradient opacity of color as identified in the color bar at the right. The cleavage efficiencies for chymotrypsin and GluC were lower, giving less coverage from identified peptides; thus, these enzymes did not further delimit the site. The combined results for each receptor are summarized below the peptide maps as colored lines below the delimiting sequence; the span of ordered structure in the respective cryo-EM maps are shown in black above each sequence; and the integrated delimitation of possible cleavage is highlighted in a dashed red box covering the sequence.



**Extended Data Fig. 4. Flowchart outlining cryo-EM image acquisition and processing for the structure of CeLGR homodimer in apo state.** The green dashed box displays the chosen high-quality 3D class, while the red dashed box represents a top view of a putative CeLGR trimer. All processing was performed using cryoSPARC (see Methods for details).

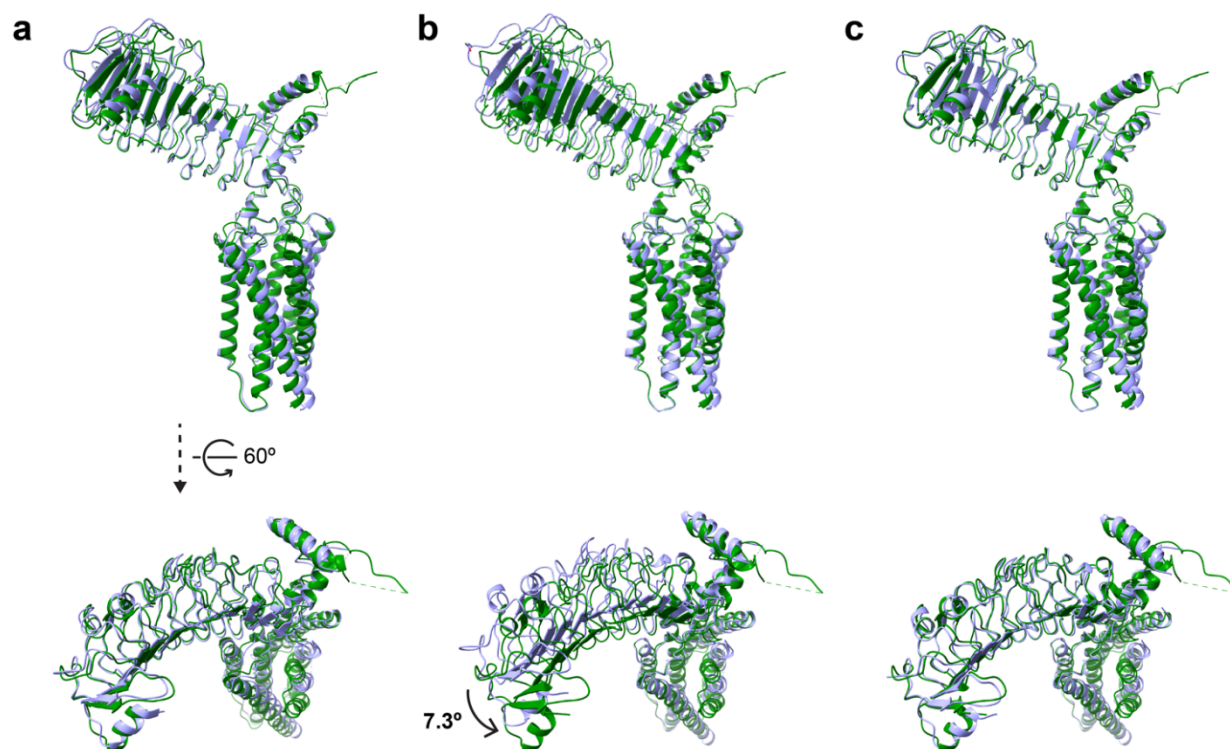


**Extended Data Fig 5. Cryo-EM data analysis for the CeLGR homodimer in apo state.** Representative micrograph (a) and 2D classes (b) for the CeLGR homodimer. (c) Heat map of the Euler angle distribution of particle orientations from the final 3D reconstruction in CryoSPARC v4.5.0. (d) Fourier shell correlation (FSC) curves between the two independently refined half-maps of final cryo-EM 3D reconstruction. (e) Map to model FSC curves. (f) Sharpened map of CeLGR homodimer colored by local resolution.



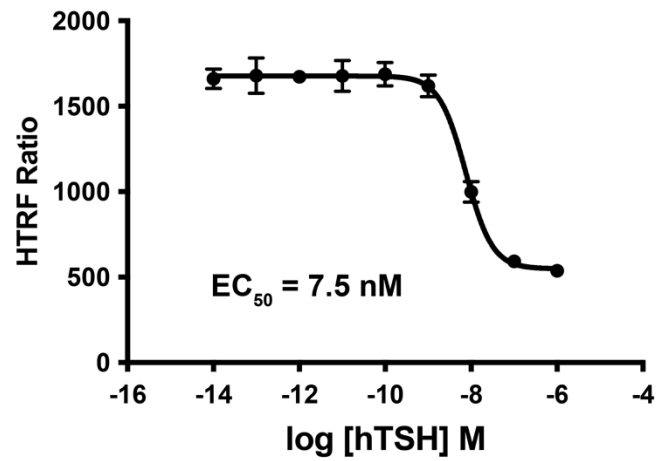
**Extended Data Fig 6. Model-to-map fittings for selected features.** Cryo-EM densities (transparent grey surface) are shown with corresponding segments of the atomic model for CeLGR protomer A.



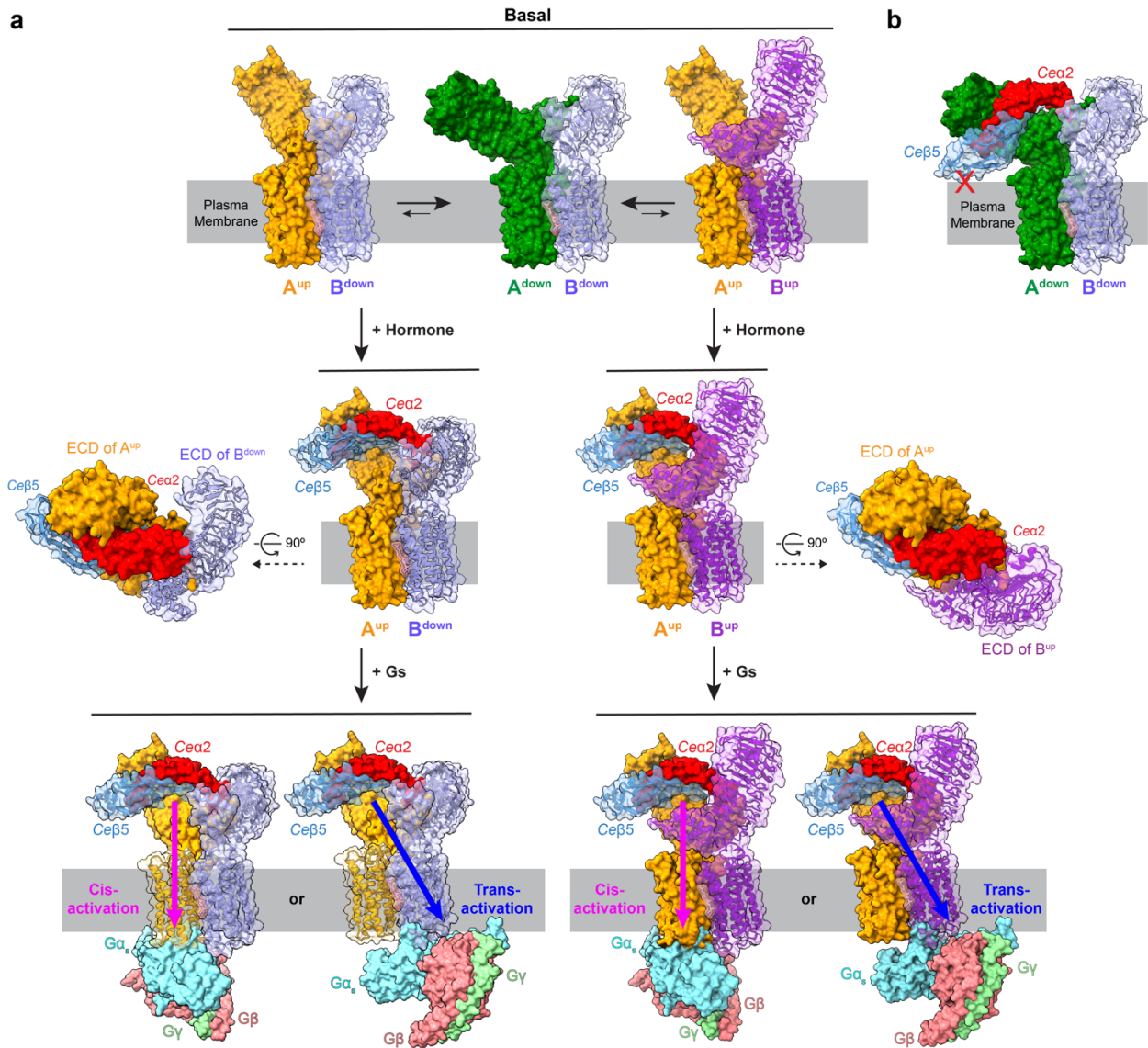


**Extended Data Fig. 7. Comparison between protomers A (green) and B (light purple) of the apo-state CeLGR homodimer. (a)** Overall superimposition of protomer B onto A. **(b)** Superimposition of the TMD of protomer B onto the TMD of protomer A. **(c)** Superimposition of the ECD of protomer B onto the ECD of protomer A after first superimposing the TMDs as in (b).





**Extended Data Fig. 8. Activation of hTSHR by hTSH.** The EC<sub>50</sub> of hTSHR activation by hTSH was measured using a cAMP-based competitive assay in HEK293S cells as a control. Data points (mean  $\pm$  s.d) are plotted for measurements from an experiment of n=3 biological replicates.



**Extended Data Fig. 9. Implications for receptor activation. (a)** In the basal state, the ECD of unliganded CeLGR can spontaneously transit to the up conformation: either one ECD-up (left) or two ECDs-up (right) seems possible. Binding of Cea2β5 ligand will stabilize the ligand-bound ECD in the 'up' conformation. G protein coupling to either protomer A (cis-activation) or protomer B (trans-activation) seems feasible. In each image, protomer A and Cea2 are drawn as opaque surfaces while protomer B and Ceβ5 are drawn as transparent surfaces exposing ribbon diagrams of the polypeptide backbone. **(b)** Binding of Cea2β5 onto a CeLGR dimer in the 'down' conformation might cause collision between the Ceβ5 subunit and cell membrane.

**Supplementary Video 1. Top view of the global conformational variability analysis of CeLGR homodimer in apo state.** The angle between the ECDs of protomer A (left) and protomer B (right) changes continuously. Simultaneously, the TMDs contract and extend like a 'heart'.

**Supplementary Video 2. Side view of the global conformational variability analysis of CeLGR homodimer in apo state.** Protomer A is on the left and protomer B is on the right.