



OPEN

# Coevolution underlies GPCR-G protein selectivity and functionality

Min Jae Seo<sup>1</sup>, Joongyu Heo<sup>2</sup>, Kyunghui Kim<sup>2</sup>, Ka Young Chung<sup>4</sup> & Wookyung Yu<sup>1,3</sup>✉

G protein-coupled receptors (GPCRs) regulate diverse physiological events, which makes them as the major targets for many approved drugs. G proteins are downstream molecules that receive signals from GPCRs and trigger cell responses. The GPCR-G protein selectivity mechanism on how they properly and timely interact is still unclear. Here, we analyzed model GPCRs (i.e. HTR, DAR) and G $\alpha$  proteins with a coevolutionary tool, statistical coupling analysis. The results suggested that 5-hydroxytryptamine receptors and dopamine receptors have common conserved and coevolved residues. The G $\alpha$  protein also have conserved and coevolved residues. These coevolved residues were implicated in the molecular functions of the analyzed proteins. We also found specific coevolving pairs related to the selectivity between GPCR and G protein were identified. We propose that these results would contribute to better understandings of not only the functional residues of GPCRs and G $\alpha$  proteins but also GPCR-G protein selectivity mechanisms.

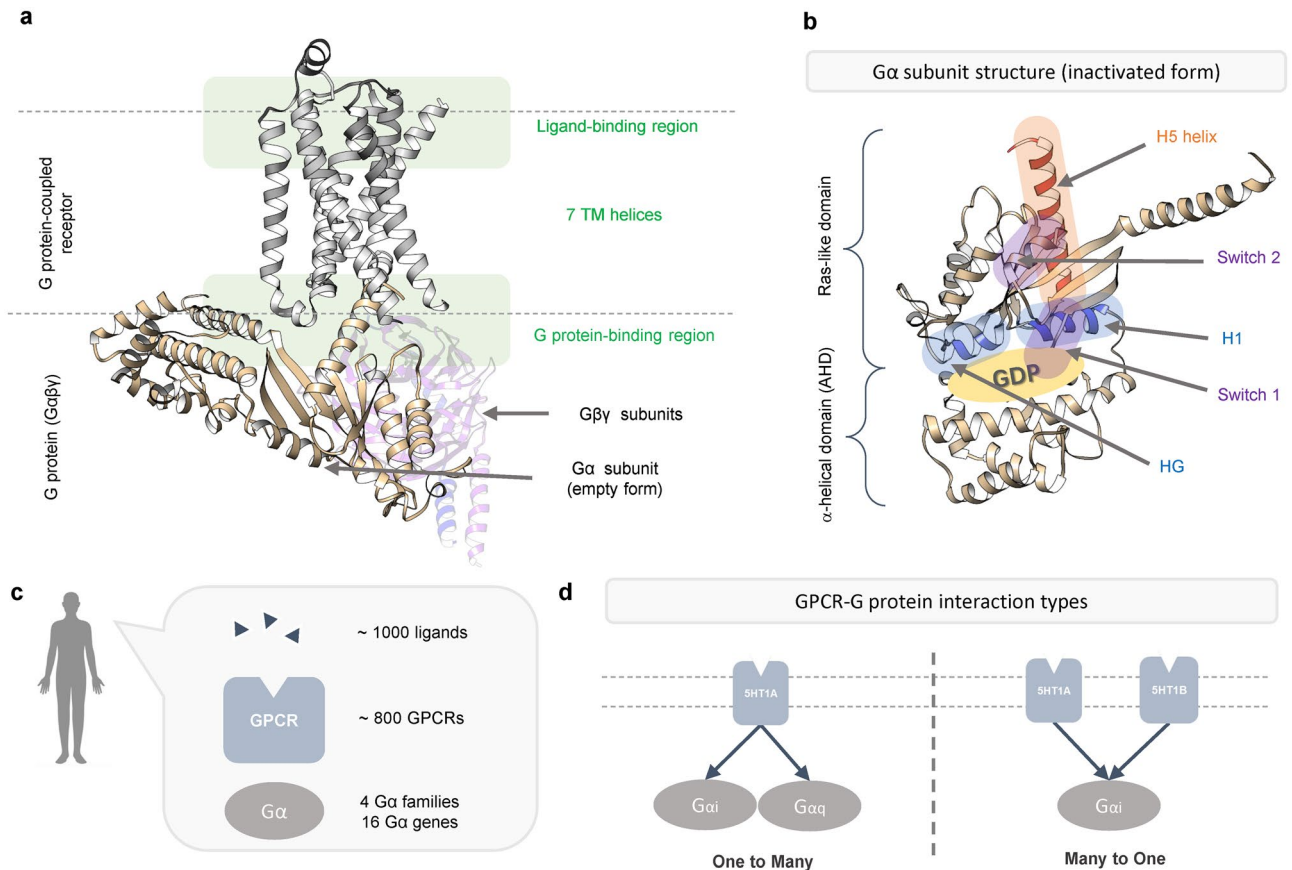
G protein-coupled receptors (GPCRs) are one of the most important signal transduction systems to transmit extracellular signals (e.g., light, odorants, and hormones) into cells. GPCRs regulate critical physiological functions including sense recognition, neural transmission, and hormonal responses<sup>1</sup>. Since they play important roles in human physiology and pathology<sup>2</sup>, GPCRs have been extensively studied over the past decades. However, a number of questions regarding GPCR systems still remain to be addressed, and moreover new phenomena have been identified day by day in recent years. Current active research into GPCRs would help to understand the fundamentals of cell signaling and would be useful for GPCR-targeted drug developments<sup>2-4</sup>.

Much of the recent progress in GPCRs has been achieved after the crystal structure of  $\beta_2$ -adrenergic receptor ( $\beta_2$ AR) has been revealed<sup>5,6</sup>. GPCRs share common structural features with seven-transmembrane (TM) helices (Fig. 1a). Upon ligand-binding, the seven TM helices undergo conformational changes, which promote the interaction of the receptor with downstream signaling molecules in cells<sup>7</sup>. The canonical downstream signaling molecules for GPCRs are heterotrimeric G proteins. G proteins consist of three subunits: G $\alpha$ , G $\beta$ , and G $\gamma$  (Fig. 1a). The G $\alpha$  subunit consists of a Ras-like domain and  $\alpha$ -helical domain, and the nucleotide-binding pocket is located between these two domains (Fig. 1b). The Ras-like domain directly interacts with the receptor (Fig. 1b) and transduces signals to other effectors<sup>8</sup>. In the basal state, the G $\alpha$  subunit is occupied by GDP and forms a heterotrimer with the G $\beta\gamma$  subunits. Active GPCRs act as guanine nucleotide exchange factors (GEFs) by releasing GDP from G proteins. Under physiological conditions, GTP is quickly inserted into the G $\alpha$  subunit, which then dissociates from the receptor and G $\beta\gamma$  subunit.

G proteins are sub-categorized depending on the G $\alpha$  subunit. In human, there are 16 G $\alpha$  subunit genes, which are classified into four subtype families (G<sub>s</sub>, G<sub>i/o</sub>, G<sub>q/11</sub>, and G<sub>12/13</sub>) (Fig. 1c). Each subtype family of G proteins interacts with different downstream effector molecules to induce distinct cell responses<sup>9</sup>. In human, approximately 1000 ligands and 800 GPCRs have been identified (Fig. 1c). Therefore, one G protein subtype interacts with more than one receptor. Moreover, many receptors can couple to not only one G protein subtype but also more than one G protein subtypes (Fig. 1d).

In recent years there has been an increased research interest in explaining how GPCRs and G proteins recognize each other in a timely manner<sup>10,11</sup>. However, complicated interaction patterns exacerbate to solve the selectivity problem (Fig. 1d). To determine the GPCR-G protein coupling selectivity, Flock et al. recently used an evolutionary approach<sup>10</sup>. They used the concepts of paralogue and orthologue conservation. If some positions in G $\alpha$  protein have pronounced conservation of orthologue and paralogue, they are defined as the conserved group concerning their major function. Positions with a highly conserved orthologue and low paralogue conservation are defined as 'selectivity barcode'. The authors identified the selectivity barcode positions critical for

<sup>1</sup>Department of Brain and Cognitive Sciences, DGIST, 333 Techno jungang-daero, Daegu 42988, Republic of Korea. <sup>2</sup>Department of Undergraduate Studies, DGIST, 333 Techno jungang-daero, Daegu 42988, Republic of Korea. <sup>3</sup>Core Protein Resources Center, DGIST, 333 Techno jungang-daero, Daegu 42988, Republic of Korea. <sup>4</sup>School of Pharmacy, Sungkyunkwan University, 2066 Seoburo, Jangan-gu, Suwon 16419, Republic of Korea. ✉email: wkyu@dgist.ac.kr



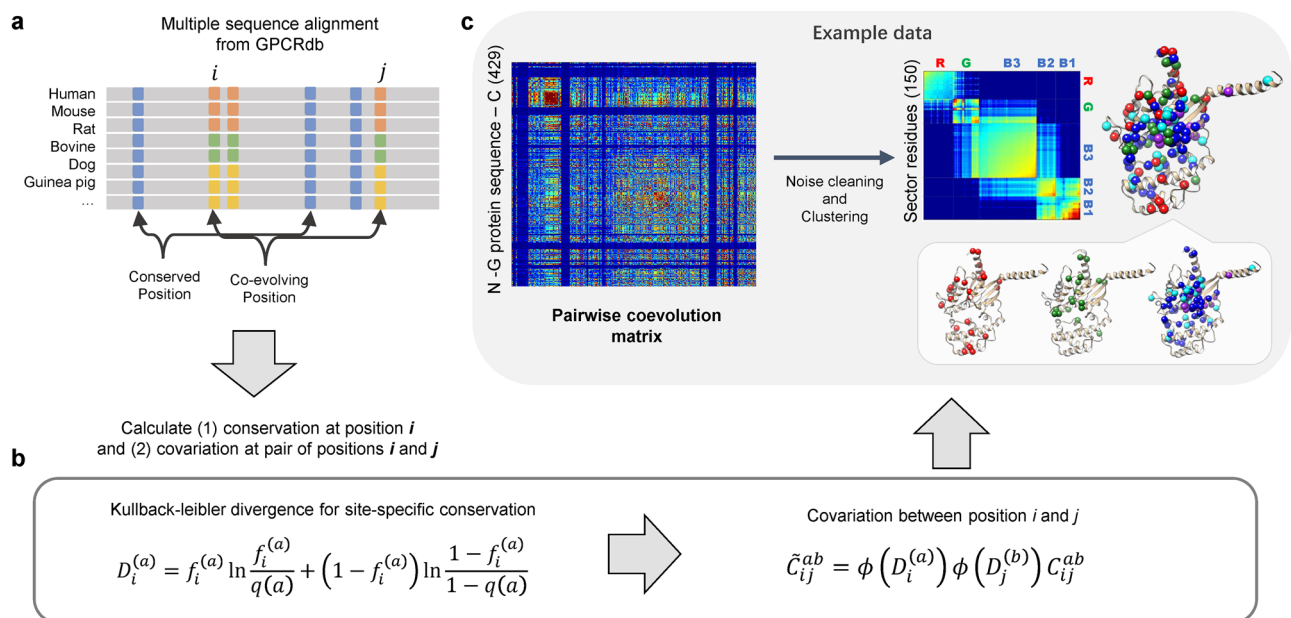
**Figure 1.** Structure and selectivity of GPCR and G protein. (a) Structure of GPCR-G protein complex. This structure is obtained from GPCRdb (modeling structure of 5HT2B-Gq complex). (b) Structure of G protein alpha subunit. It is consisted of Ras-like domain and  $\alpha$ -helical domain (AHD). Please note that AHD is displaced in the GPCR-G protein complex structure in (a). (c) Approximate number of currently identified ligands, GPCRs and G $\alpha$  families. (d) Diverse mechanism of GPCR-G protein activation. One GPCR can activate different G proteins, and one G protein can be activated by diverse GPCRs.

GPCR-G protein selectivity, which are available at GPCRdb (<https://GPCRdb.org>). This concept explaining selectivity determinants is innovative and persuasive. However, it is difficult to define the key components of selectivity determinants because (as discussed later in the paper) approximately 30–40% of amino acids are defined as selectivity barcodes (Supplementary Fig. S5a). Flock et al.<sup>10</sup> mainly focused on GPCR-G protein interface regions, hence they could not fully explain the effects on selective binding of other residues outside the GPCR-G protein interface.

Here, we adopted another approach, sequence coevolution, to understand the GPCR-G protein coupling system. Sequence coevolution is used to predict protein contacts, folding network, allosteric network, functional sites, and other features<sup>12</sup>. It is similar to the evolutionary approach but differs in that it can coincidentally observe pairs of amino acids. Among the various coevolutionary approaches, statistical coupling analysis (SCA) defines functional clusters, also termed sectors. Sectors consist of highly coevolved residues<sup>13</sup>. We applied SCA to explain the selectivity and functionality of the GPCR-G protein coupling system.

## Results

To analyze the GPCR-G protein coupling system from a coevolutionary perspective, we applied the SCA which is one of the biophysical and bioinformatic tools. Originally, SCA was developed to measure the energetic coupling between residues in a protein<sup>14</sup>. It can now define functional coevolutionary units<sup>13,15</sup>. This method measures conservation at position  $i$  and covariance between pairs of positions  $i$  and  $j$  in a protein multiple sequence alignment (Fig. 2a,b). After cleaning noise and clustering highly coevolving residues together from a pairwise coevolution matrix, a sector matrix is generated (Fig. 2c). Generally, a protein family with high sequence identity is a good target for SCA analysis. Therefore, G $\alpha$  proteins are good candidates for SCA analysis. However, SCA has practical problems for analyzing GPCR families. GPCRs have numerous subfamilies, but they share very low sequence identity, even in the same subfamilies. The extremely low sequence identities make it difficult to interpret SCA results. We seek to choose GPCR subfamilies with relatively well-defined coupling G proteins and selected two subfamilies of Class A GPCRs as model receptors: 5-hydroxytryptamine (5-HT) receptors (HTRs) and dopamine receptors (DARs) (Supplementary Fig. S1).



**Figure 2.** Overview of statistical coupling analysis (SCA). **(a)** Multiple sequence alignment is provided by GPCRdb (GPCRdb.org). **(b)** Calculation of Kullback–Leibler divergence for site-specific conservation and covariation between pairwise position. **(c)** Pairwise coevolution matrix is generated from **(b)**. Then, through the noise cleaning and clustering process, sectors are defined.

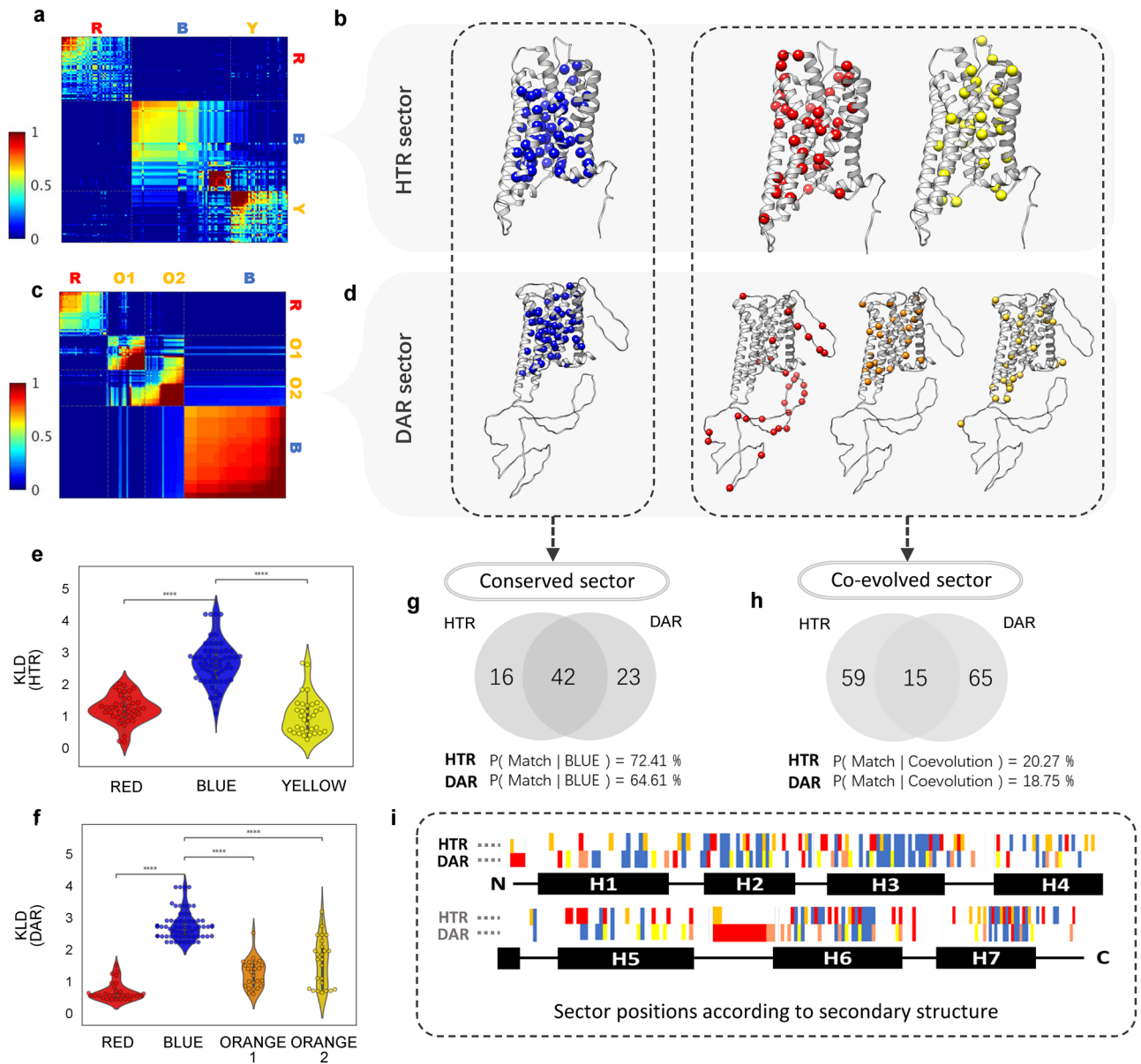
**GPCRs have coevolved sectors and a common conserved sector.** We used multiple sequence alignments from GPCRdb because they provide well-labeled alignments using the Generic-GPCRdb numbering system by each receptor<sup>16</sup>. This labeling system makes comparisons much easier among the results of each receptor. We selected HTRs and DARs as model systems, as discussed above. Applying SCA and sector refinement algorithms to HTRs, we defined three sectors that were designated HTR-Red, HTR-Blue, and HTR-Yellow (Fig. 3a,b, Supplementary Table S1). Four DAR sectors were defined: DAR-Red, DAR-Blue, DAR-Orange1, and DAR-Orange2 (Fig. 3c,d, Supplementary Table S2). The total number of residues in the HTR and DAR sectors were 132 and 145, respectively.

To characterize the conservation of each sector, we examined the Kullback–Leibler divergence (KLD) of each sector (Fig. 3e,f) using the Mann–Whitney test. Higher KLD values meant more conservation. The HTR-Blue and DAR-Blue sectors consisted of more conserved residues than the other sectors. Generally, SCA sectors are coevolutionary groups<sup>13</sup>, but the HTR-Blue and DAR-Blue sectors are grouped by conservation. Therefore, we defined HTR-Blue and DAR-Blue sectors as conserved sectors. The others were defined as coevolutionary sectors.

HTR-Blue and DAR-Blue sectors were formed by highly conserved positions (Fig. 3e,f) and were found mainly in TM regions facing the inside of the receptors (Fig. 3b,d, Supplementary Fig. S2). These findings suggested that the blue sectors may be important to stabilize the TM structures. In addition, functionally important conserved motifs, such as DRY<sup>17–23</sup>, NPxxY(N,P,Y)<sup>19–25</sup>, CWxP(W,P)<sup>19,21,23–25</sup>, and PIF(P,I)<sup>22,23,26</sup> were identified in the blue sectors (Supplementary Fig. S2). Moreover, conserved fingerprint residue<sup>21</sup>, disulfide bonds, and Na<sup>+</sup> pockets were involved<sup>21,23,27,28</sup>. Some ligand-binding pockets<sup>23,28</sup> and G protein coupling regions were also in the blue sectors<sup>21</sup>. These analyses indicated the involvement of the blue sectors in the conserved functions of GPCRs<sup>17–20</sup>.

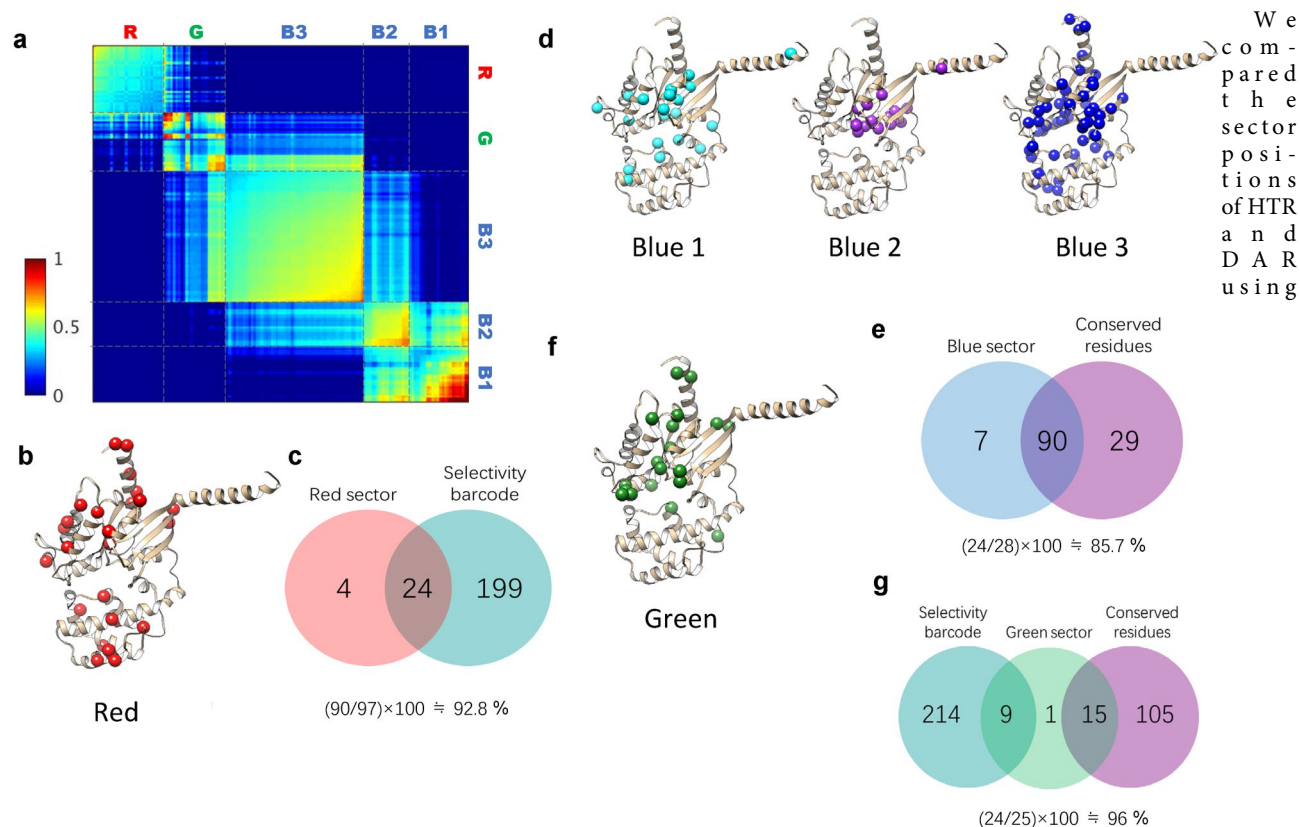
The red sectors in HTR and DAR showed interesting correlations with the GPCR-G protein coupling patterns (Supplementary Fig. S1). The HTR-Red positions could be roughly divided into Gi-coupled receptors and Gq coupled receptors (Supplementary Fig. S1a and Supplementary Dataset S1). Similarly, residues in DAR-Red correlated with G protein selectivity which distinguish D1-like and D2-like subtypes (Supplementary Fig. S1b). The positions of HTR-Red were determined to be located in the intracellular regions of TM 3–7, intracellular loop 2 (ICL2), and helix 8 (Fig. 3b and Supplementary Fig. S2a). These regions have been suggested to be the major contact sites for G proteins<sup>10</sup>. The positions of the DAR-Red sector were mostly located in ICL3 (Fig. 3d and Supplementary Fig. S2b), which may selectively interact with downstream signaling proteins. Additionally, HTR-Red contained residues from the ligand-binding pocket and conserved motifs for GPCR activation (Fig. 3b and Supplementary Fig. S2a)<sup>21</sup>. These results support the suggestion that the red sectors are involved in G protein selectivity, which is discussed further below.

There were other sectors, such as HTR-Yellow, DAR-Orange1, and DAR-Orange2 sectors. HTR-Yellow belonged to the coevolutionary groups and it has lower KLD values than the blue sectors. (Fig. 3e). Interestingly, HTR-Yellow contained several ligand-binding residues (Fig. 3b, Supplementary Table S4, Supplementary Fig. S2a). The pattern of amino acid sequences in HTR-Yellow allowed further subdivision into HTR subtypes (Supplementary Fig. S1a, Supplementary Dataset S1). We analyzed mutation effects of HTR-yellow with mutant browser at GPCRdb (Supplementary Fig. S3). The results showed the residues of HTR-Yellow could make the changes to ligand binding potency and it was different for each position and subtypes. Therefore, we presumed



**Figure 3.** GPCRs have coevolved sectors and a common conserved sector. **(a)** Sector matrix of 5-HT receptor. **(b)** Structure representations of sectors of 5-HT receptor. **(c)** Sector matrix of DA receptor. **(d)** Structure representations of sectors of DA receptor. DAR-Orange1 is colored in orange, and DAR-Orange2 is colored in yellow. **(e)** Kullback–Leibler divergence of sectors of 5-HT receptor. Mann–Whitney test is performed between sectors. (\*\*\*\* $p \leq 1.00e-04$ ) **(f)** Kullback–Leibler divergence of sectors of DA receptor. Mann–Whitney test is performed between sectors. (\*\*\*\* $p \leq 1.00e-04$ ) **(g)** Venn-diagram of matched residue between HTR-Blue sector and DAR-Blue sector. **(h)** Venn-diagram of matched residue between coevolved sectors of HTR and DAR. **(i)** Sector positions of HTR and DAR according to secondary structure. At DAR row of **(i)**, Orange color represents DAR-Orange1 and yellow represents DAR-Orange2.

that this sector may be related to the fine-tuning of ligand-binding responses and may regulate subtype specific reactions. The DAR-Orange1 and DAR-Orange2 sectors were also coevolved sectors (Fig. 3d), as they were less conserved than the blue sector. (Fig. 3f) The DAR-Orange1 and DAR-Orange2 positions clearly showed the subtype specific sequences (Supplementary Fig. S1b). Positions of DAR-Orange1 had D3-D4 specific amino acid pattern and positions of DAR-Orange2 had a D4-specific amino acid pattern (Supplementary Fig. S1b, Supplementary Dataset S1). Several positions from DAR-Orange1 and DAR-Orange2 were distributed at the ligand-binding residues in the intracellular regions of TM5 and TM6, ICL2, and conserved motifs that include the DRY, CWxP, and NPxxY motifs (Fig. 3d and Supplementary Table S4, Supplementary Fig. S2b). These analyses suggested that DAR-Orange1 and DAR-Orange2 might also be involved in GPCR-G protein coupling and subtype specific characteristics, although certain functions of DAR-Orange1 and DAR-Orange2 were ambiguous. These results affirmed that specific subtype characteristics could be checked using coevolutionary analysis.

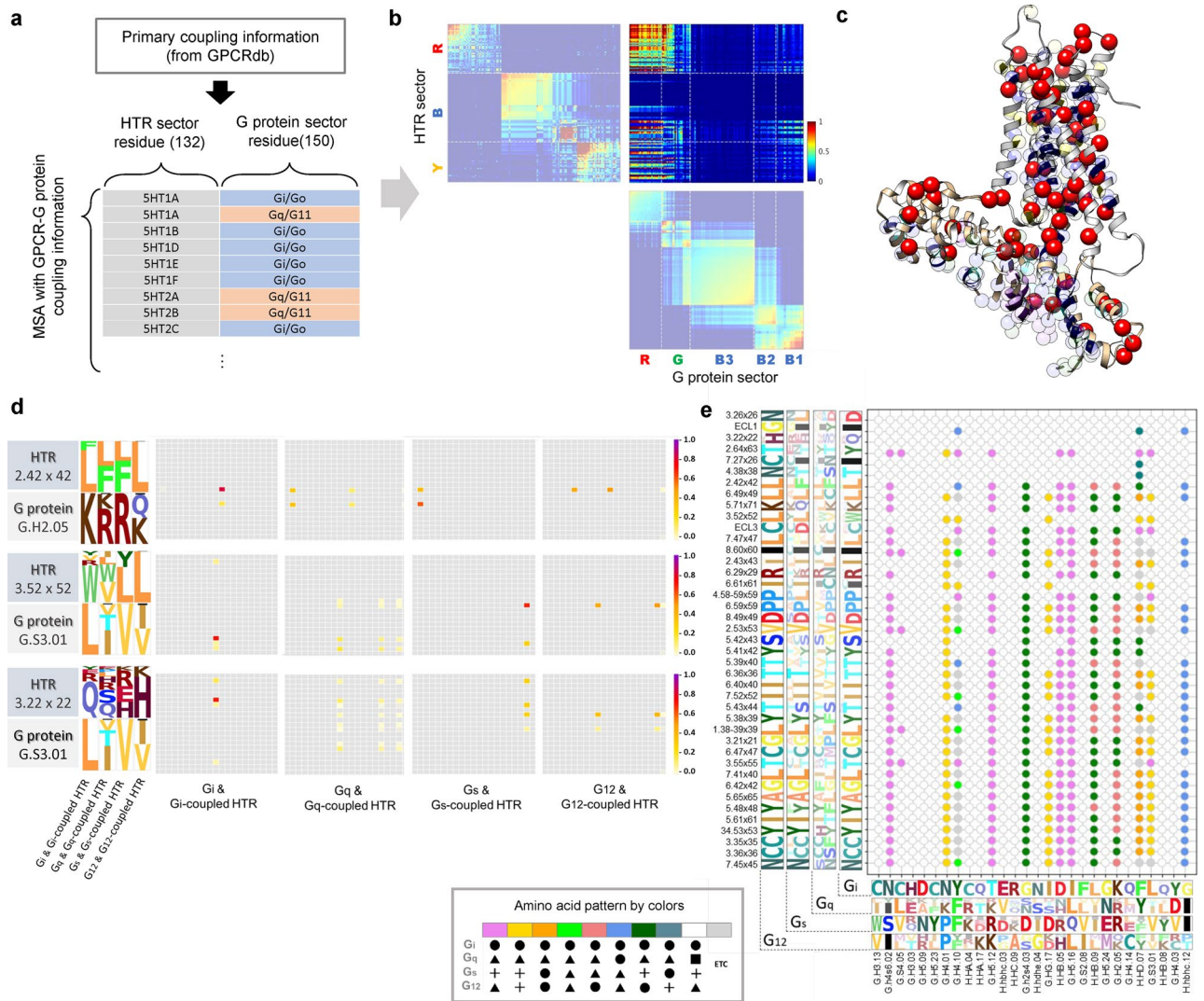


**Figure 4.** Functional networks in G protein from recognition to signal transduction. **(a)** Sector matrix of G protein. Three sectors (G-Red, G-Green, G-Blue) are existed and G-Blue sectors are subdivided into three subsectors (G-Blue1, G-Blue2, G-Blue3). **(b)** Structure representation of red sector. **(c)** Venn-diagram of matched residue between G-Red and selectivity barcode defined by previous work<sup>10</sup>. **(d)** Structure representation of blue sectors. **(e)** Venn-diagram of matched residue between G-Blue and conserved residues defined by previous work. **(f)** Structure representation of green sector. **(g)** Venn-diagram of matched residues among G-Green, Selectivity barcode and conserved residue. Close form of G protein structure is the modeling structure by Alhadeff et al.<sup>34</sup>.

GPCRdb numbering systems. Comparing HTR-Blue to DAR-Blue, 70% of the positions matched (Fig. 3g,i). Almost of matched positions tended to be conserved in aminergic receptor family while some positions showed relatively low conservation at class A level (Supplementary Fig. S4). This results may be caused by the differences between aminergic-wide and class A-wide conservation. This indicated that major functions have evolved through a group of residues that are conserved in each family of other GPCRs. However, coevolved sectors (red, yellow, oranges) had their own characteristics, and they shared only 20% of residues (Fig. 3h,i). The coevolved sectors showed interesting features at ligand–receptor interactions. We analyzed available crystal structures and their ligand interactions, then we found the specific receptor–ligand interactions (Supplementary Table S4, Supplementary Dataset 2). The common ligand-interactive residues were mostly blue sectors and ligand-specific interactive residues were mostly the coevolved sectors. This result may indicate the coevolved residues have modulated interactions to various ligands.

**Functional networks in G protein from recognition to signal transduction.** The interaction between GPCRs and G proteins mostly occurs at the  $G\alpha$  subunit.  $G\alpha$  subunits, but not the  $G\beta\gamma$  subunits, have been suggested to be the key protein for GPCR-G protein selectivity<sup>29–31</sup>. Therefore, we performed SCA with 16  $G\alpha$  proteins. The  $G\alpha$  protein residues comprised three sectors (G-Red, G-Green, and G-Blue) (Fig. 4b and Supplementary Table S3). Total number of sector residues is 150. Most of these sectors were identified in the core of the Ras-like domain, with some in the  $\alpha$ -helical domain and  $\alpha N$  (Fig. 4 and Supplementary Fig. S5). The Ras-like domain has been suggested to have critical functions of  $G\alpha$  proteins, such as receptor binding, GTPase activity, and effector binding<sup>29, 30, 32</sup>. Thus, the identified sectors might be involved in G protein functions.

The majority of G-Red sector residues were positioned at the interface with GPCRs (e.g.  $\beta 1$ ,  $\beta 2$  loop, C-terminal part of  $\alpha 3$  and  $\alpha 4$ , and  $\alpha 5$ ), although some were positioned in the  $\alpha$ -helical domain (Fig. 4b and Supplementary Fig. S5b). Flock et al. defined the ‘selectivity barcode’ that determines GPCR-G protein selectivity using evolutionary concepts (Supplementary Fig. S5a), and identified 25 CGN positions in these barcodes that contacted the receptor<sup>10</sup>. Interestingly, most of the G-Red sector positions (24 of 28 positions) overlapped with the selectivity barcode suggested by Flock et al. (Fig. 4c and Supplementary Fig. S5)<sup>10</sup>. This result suggested that



**Figure 5.** Red sectors are mediator between GPCR and G protein signaling. **(a)** Construction of new multiple sequence alignments for integrated analysis of GPCR-G protein. **(b)** Results of integrated SCA. Blurred matrices are original sector matrix of HTR and G protein. The cross section of each matrix is covariation between HTR sectors and G protein sectors. **(c)** Structure representation of HTR-Red and G-Red interaction which have high coevolution signals. **(d)** Sequence logo and amino acids patterns of HTR-Red and G-Red interaction. In amino acid pattern figure, x, y axis represent the amino acid order of three letter code alphabetically from top to bottom, left to right. (one letter code order: ARNDQCQEGHILKMFPTWYV, - represents gap). **(e)** Amino acid pattern analysis between HTR-Red and G-Red residues. Patterns of each color are explained in legend. Same shape in each column represents similar distribution of amino acids' pair.

G-Red might serve as a determinant of GPCR-G protein selectivity. This suggestion is supported by the analysis described in Fig. 5. Flock et al.<sup>10</sup> defined selectivity determinant residues using evolutionary concepts resulting in approximately 30–40% of amino acids defined as selectivity barcodes (Supplementary Fig. S3a). The present results were a marked improvement in selecting more specific positions for selectivity determinants.

Concerning the functions of sectors, when compared with previous findings<sup>10</sup>, the G-Blue sector was the most conserved sector in the Ga proteins (Fig. 4d,e). More specifically, the G-Blue sector comprised three subsectors: G-Blue1, G-Blue2, and G-Blue3. These subsectors contained many key residues for essential functions (Fig. 4a,d, and Supplementary Fig. S3b). Positions of G-Blue1 contain switch 1, 2, and 3 regions that regulate G protein activation status by changing their conformation depending on the binding GDP or GTP (Fig. 4d and Supplementary Fig. S5b)<sup>31</sup>. Residues of the G-Blue2 sector were mostly located around  $\alpha 1$  helix (Fig. 4d and Supplementary Fig. S5b), which is related to stabilization of GDP binding<sup>33</sup>. Residues from G-Blue3 also contained residues from the switch 1, 2, and 3 regions (Fig. 4d and Supplementary Fig. S5b). Moreover, the majority of G-Blue3 residues were located in the regions that stabilized the basal conformation of the Ga proteins, which include hydrophobic residues at G.H5.07 and G.H.5.08 and residues within the  $\beta$ -sheets, inter-domain interface, and TCAT motif (Fig. 4d and Supplementary Fig. S5b). Interestingly, G-Blue3 was also present at the C-terminus of  $\alpha 5$  helix (Fig. 4d and Supplementary Fig. S5b), where the Ga protein formed a major interface with a receptor

(Fig. 1a)<sup>8</sup> and G-Blue3 bridged to the G-Red sector in the sector matrix (Fig. 4a). Based on these data, it can be suggested that, among G-Blue sectors, G-Blue3 is the first to receive the recognition signal from G-Red.

Lastly, the G-Green sector seems to be an intermediate group between G-Red and G-Blue. The sector matrix revealed that G-Green sector was located between the G-Red and G-Blue sectors (Fig. 4a). The G-Green sector residues had G-Red-like and G-Blue-like properties (Fig. 4f, Supplementary Fig. S5c). More specifically, the residues that overlapped with the selectivity barcode or conserved residues were plotted on the snake map (Fig. 4g and Supplementary Fig. S5c). The conserved residues found in G-Green (Supplementary Fig. S5c, purple) were mostly involved in stabilizing the conformation of the G $\alpha$  protein. The selectivity barcode residues within G-Green (Supplementary Fig. S5c, green) were divided into two positions. Five of eight were located in the upper part of the G $\alpha$  protein, making them potential receptor binding residues. The remaining three were located near the nucleotide-binding pocket. Therefore, we presumed that the signals from GPCRs were somehow received in the G-Red sector and transduced to G-Blue to perform the actual functions of G protein, which is mediated by G-Green sectors.

**Red sectors mediate between GPCR and G protein signaling.** To integrate the analyses of GPCRs and G $\alpha$  proteins, we built a new multiple sequence alignment with coupling information. First, we used the sector positions of GPCRs and G $\alpha$  proteins to study the relationship between GPCR coevolution and G $\alpha$  protein coevolution. We concatenated the HTR sector to the G $\alpha$  protein sector with known selectivity information in GPCRdb (Fig. 5a). For example, the HTR1A family interacted with the Gi and Gq family. Therefore, we repeated the HTR1A sequence twice and integrated one with Gai and the other with Gaq. HTR1B coupled with only Gi. Therefore, we simply integrated HTR1B residues with Gai. After repeating this process for all subtypes of HTR, we ran the SCA process and obtained a positional coevolution matrix.

Integrated SCA revealed a strong correlation signal between the HTR-Red and G-Red sectors (Fig. 5b,c). This result is consistent with a previous interpretation of the sectors, because each is suggested to be involved in selective recognition between GPCRs and G proteins (Figs. 3, 4). In other words, the HTR-Red and G-Red sectors were strongly coevolved and played a role as selectivity determinants. Additionally, G-Red also weakly coevolved with some positions in HTR-Blue or HTR-Yellow, and these make minor correlation signals (Fig. 5b).

To better understand the coevolution and selectivity between the red sector of GPCRs and the red sector of G $\alpha$  proteins, we analyzed the pattern of coevolving amino acids. The results of HTR-G $\alpha$  protein pattern analysis revealed the pairwise pattern of G $\alpha$  protein-specific amino acids (Fig. 5d). For example, the pairwise frequency of amino acids between residue 2.42 in HTR and G.H2.05 in G $\alpha$  protein are shown in the first row of Fig. 5d. Each 21  $\times$  21 matrix in Fig. 5d represents Gai, Gaq, Gas, and Ga12 coupled sequence patterns. Gai-coupling and Ga12-coupling had similar amino acid patterns. Likewise, Gaq-coupling and Gas-coupling showed similar amino acid patterns. These results are visually represented as sequence logos in Fig. 5d. Amino acid frequencies between HTR 3.52 and G protein G.S3.01 are shown in the middle of the rows (Fig. 5d). While Gas coupling and Ga12 coupling had similar amino acid patterns, the patterns of Gai-coupling and Gaq-coupling were different and distinct from the others. The last row shows that the patterns of GPCR-G $\alpha$  protein coupling are all different types between HTR 3.22 and G $\alpha$  protein G.S3.01 (Fig. 5d). We examined all possible pairs between GPCR and G protein sectors and represented patterns with color coding (Fig. 5e). Distinguishing patterns of amino acid pairs were evident between HTR-Red and G-Red (Fig. 5e). Similar results were also found at DAR (Supplementary Fig. S6a,b). DAR-Red and G-Red were strongly coevolved (Supplementary Fig. S6a). Moreover, the other coevolved sectors, DAR oranges, were also correlated to DAR-Red. Most of DAR residues had different amino acid compositions depending on their G protein selectivity (Supplementary Fig. S6b). Amino acid patterns of DAR-G protein were simpler than HTR cases, because DAR does not primarily couple to Ga12. According to the results, we could conclude that intra-coevolution units make inter-coevolution with other proteins in case of GPCR-G protein. These are the most relevant evidence for coevolutionary sectors as a selectivity determinant.

When we analyzed the GPCR-Blue and G-Blue interactions in both receptors, most of the residue pairs had unified amino acids in four different G $\alpha$  subtypes (Supplementary Fig. S7a,b). These results affirm that the blue sectors are conserved through subtypes of GPCRs and G $\alpha$ . Compared to the analysis of red sectors, its interactions were less related to GPCR-G protein selectivity. Furthermore, the results of pattern analysis showed differences between G-Blue1 and G-Blue3, which were not clearly distinct from the SCA results (Supplementary Fig. S7a). Residues of G-Blue1 distinguished amino acids in G12 subtypes and different selectivity patterns for GPCR coupling.

## Discussion

GPCR-G protein coupling selectivity has been an active research topic. Extensive studies on three-dimensional GPCR-G protein complex structures analyzed by X-ray crystallography or cryo-electron microscopy (cryo-EM)<sup>32,35</sup> have failed to deduce the precise coupling selectivity mechanism. This is probably because most of these structures are GDP-released final state of GPCR-G protein complex structures. Recently, it was suggested that step-wise conformational changes occur during GPCR-G protein coupling, and that the early-stage structure is different from the structures observed from the currently available X-ray or cryo-EM structures<sup>36,37</sup>. Therefore, there have been attempt to solve this selectivity question with evolutionary approach, experimental approach, and various computational simulations<sup>10,11,38</sup>. However, it remains still elusive to explain. Therefore, we tried to explain this issue and fill the missing parts of the selectivity mechanism using a coevolutionary approach.

SCA is a popular tool for coevolutionary analysis, but it is difficult to apply to proteins with low sequence identity, such as GPCR because multiple sequence alignments (MSA) with low sequence identity makes it difficult to separate coevolving positions from evolutionary random sequences. To overcome this problem, we selected a unit of the receptor sub-family. This approach enabled the analysis of model GPCRs and it could be possible

to extend further research about general GPCR family. In addition, we built the input alignments with coupling information to consider GPCR-G $\alpha$  protein binding selectivity (Fig. 5a and Supplementary Fig. S6a). By reflecting the selectivity of MSA, residue pairs for selective GPCR-G $\alpha$  coupling were shown through SCA analysis.

The most progress compared to the study of Flock et al.<sup>10</sup> is that the current study provided selectivity determinants within the receptors (HTR-Red and DAR-Red) and narrowed the selectivity determinants within the G $\alpha$  proteins (28 residues of G-Red sector). Red sector residues from GPCRs were subdivided according to the coupling G proteins (Supplementary Fig. S1). Strong coevolving signals were evident between the red sectors of GPCRs and G $\alpha$  proteins (Fig. 5b and Supplementary Fig. S6a). Furthermore, G $\alpha$  protein-specific amino acid patterns coevolved together (Fig. 5e and Supplementary Fig. S6b). Among them, some residues were not located in binding interface, but these residues may be related with allosteric effects. Similar results were obtained for HTR and DAR. Thus, although we did not analyze the class A GPCR family, defining the selectivity pattern related to coevolution may be generalized to class A GPCRs.

G-Red residues are located in both Ras-like and  $\alpha$ -helical domains (Supplementary Fig. S5b). Within Ras-like domain, G-Red residues are mainly localized at reported receptor-contact sites such as  $\beta$ 2/ $\beta$ 3,  $\alpha$ 2/ $\beta$ 4,  $\alpha$ 4/ $\beta$ 6 loops and  $\alpha$ 5<sup>33, 39–45</sup> and potential contact sites, such as  $\alpha$ 3,  $\alpha$ G, and  $\alpha$ 4. These sites may contact a receptor when the receptor has a long ICL3, although long ICL3 is often not resolved or is truncated in all reported GPCR-G protein structures. Other residues within the Ras-like domain are  $\alpha$ 2 (one residue) and  $\beta$ 4 (one residue). The potential involvement of these residues in GPCR-G protein coupling selectivity should be further investigated.

The identification of G-Red within  $\alpha$  helical domain was unexpected because it does not directly interact with receptors. The role of the  $\alpha$ -helical domain has not been well studied. The results from several studies indicated that the domain may influence GTP hydrolysis activities depending on G $\alpha$  subtypes<sup>30, 46, 47</sup> and might regulate effector<sup>30, 48</sup>. Generally, residues in  $\alpha$ -helical domains have more divergent sequences than the Ras-like domain. Therefore, the suggested ten G-Red residues within the  $\alpha$ -helical domain may be involved in GPCR-G protein selectivity and also in coevolution with other downstream effectors. These possibilities require further investigation.

In addition to the GPCR-G protein coupling selectivity determinants, SCA revealed other functional sectors. The blue sectors of each target protein (HTR-Blue, DAR-Blue, G-Blue) were highly conserved compared to other sectors (Figs. 3e,f and 4e). These sectors are critical role in their conserved functions (e.g., ligand-binding for GPCRs and GTP-mediated G protein activation for G $\alpha$  proteins) and conformational integrity (Supplementary Figs. S2, S5). The results of the sector analysis also implied that G-Green may be an intermediary between the G-Red and G-Blue sectors (Fig. 4a) and that HTR-Yellow and DAR-Orange sectors are assumed to regulate receptor subtype specific characteristics because of the features of the sequence alignments (Supplementary Fig. S1).

In conclusion, the significance of our study is the suggested clarification of the GPCR-G $\alpha$  selectivity mechanism and functional sectors using coevolutionary analysis. The actual roles of these sectors need to be confirmed with further experimental approaches. The coevolutionary analysis approach adopted in this study could be extended to class A GPCRs and for other molecules, such as arrestin, G protein-coupled receptor kinases (GRKs), or regulators of G protein signaling (RGSs). Moreover, the finding that subtype specific sequences can be extracted by SCA highlights the possibility of application in the pharmacological field for receptor subtype-selective targeting within the same subfamilies. The current research on coevolving sequences shows promise for systematic studies on GPCR-driven signaling.

## Materials and methods

**Statistical coupling analysis (SCA).** The SCA 5.0 toolbox used for SCA was implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA). SCA was originally developed to measure evolutionary covariance between protein positions<sup>13–15</sup>. All MSA data were obtained from GPCRdb (GPCRdb.org). We selected sequences from all species provided in the database. In the major processing step, the positional conservation level was obtained using the KLD at position  $i$  and covariance between positions  $i$  and  $j$ . After these calculations, we removed statistical and historical noise and defined clusters (sectors) that consisted of highly coevolved positions. The details of the measurement procedure are described in a previous study<sup>13</sup>. Details of methods are in Supplementary Appendix.

**Sector refinements.** SCA could exclude high coevolved positions or involve relatively weak positions while defining sectors with eigenvectors. Therefore, we added sector refinement steps to permit a similarity comparison. Some residues in non-sector positions with pronounced similarity with the sector were newly defined as a sector. In addition, some residues in the sector that displayed low similarity were discarded. The detailed process is described in Supplementary Fig. S8.

**GPCR-G protein integrated SCA.** We constructed an integrated multiple sequence alignment to measure the coevolution between GPCR and G protein with previously defined sector positions. To reflect actual biological relationships, we considered the GPCR-G protein coupling pattern during MSA construction. For instance, HTR1A primarily interacts with Gi/Go and Gq/G11, generating HTR1A-Gi/o and HTR1A-Gq/o. We repeated this for every subtype of the target receptor family. The reference for coupling patterns was taken from GPCRdb, especially the merged one from two different data sources (Guide To Pharmacology DB, Inoue et al.<sup>11</sup>) We only considered primary coupling.

SCA processes were run with integrated alignments. However, we did not consider the noise cleaning step and did not drop any positions for clustering because we wanted to observe covariances between the already defined sector. Validation process for coevolution between GPCR and G proteins is in Supplementary information.



**GPCR-G protein pattern analysis.** The HTR-Red and G-Red interaction represented in Fig. 5d,e is considered as an example. First, we constructed four MSA according to the G protein coupling pattern (for example, Gi-coupled HTR sequences combined with Gi sequences). The alignments only involved sector positions of HTR (HTR-Red, 41 positions) and G proteins (G-Red, 28 positions).

For each pair of HTR-Red and G-Red ( $41 \times 28$ ), four sets of pairwise amino acid frequencies were counted and  $21 \times 21$  matrices (20 amino acid plus gaps) were generated. Plots in Fig. 5d are examples of  $21 \times 21$  matrices. Then, we converted the  $21 \times 21$  matrices to  $1 \times 441$  vectors to generate, four sets of  $1 \times 441$  vectors representing each of the HTR-G protein subtype coupling. The inner products between the four sets of vector groups were calculated to measure their similarities. There were four G protein subtypes and six possible combinations of inner products between them. We finally obtained six inner product values for each pair of HTR and G proteins (for example, group 1 consisted of [HTR1-Gi1]·[HTR1-Gq1], [HTR1-Gi1]·[HTR1-Gs1], [HTR1-Gi1]·[HTR1-G121], [HTR1-Gq1]·[HTR1-Gs1], [HTR1-Gq1]·[HTR1-G121], [HTR1-Gs1]·[HTR1-G121]). The threshold value to assess whether the patterns were similar or not was 0.6, and classified into 15 patterns. For instance, if all inner products values are under the threshold, four subtypes of HTR-G protein are having different patterns. We repeated this process to all pairs of HTR-Red and G-Red ( $41 \times 28$ ). (Supplementary Fig. S9).

**Structure and graphic representation.** The structures used for the GPCR-G protein combined form and GPCR structure were downloaded from the GPCRdb modeled structure list. We selected a 5HT2B-Gq coupled structure for HTR. For DAR, we used the drd2 modeled structure from GPCRdb. However, the drd2 structure from GPCRdb lost many parts of their ICL3. Therefore, we constructed the ICL3 of the DAR using MODELLER<sup>49</sup>. The G protein structure shown in Figs. 1 and 3 was provided is by Alhadeff et al. (active-closed-in form)<sup>34</sup>. To map sectors to structure, we first needed to align the sequence used for SCA and the sequence of the structure. MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle>) alignment web server<sup>50</sup> was used for the alignment. All protein structure representations were visualized using UCSF chimera 1.14 software<sup>51</sup>.

**Visualization and statistics.** Visualization of KLD (Fig. 3e,f) and the data in Fig. 5d,e were visualized using the Python package (matplotlib<sup>52</sup> and seaborn<sup>53</sup>). Statistical validations were performed using the Python package (statannot). The sequence logo plots in Fig. 5 were drawn using the Logomaker package. The remaining plots and matrices were generated using the SCA MATLAB code.

Received: 27 January 2021; Accepted: 25 March 2021

Published online: 12 April 2021

## References

- Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
- Thompson, M. D., Burnham, W. M. & Cole, D. E. C. The G protein-coupled receptors: Pharmacogenetics and disease. *Crit. Rev. Clin. Lab. Sci.* **42**, 311–389 (2005).
- Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
- Sriram, K. & Insel, P. A. G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs?. *Mol. Pharmacol.* **93**, 251–258 (2018).
- Rasmussen, S. G. F. et al. Crystal structure of the  $\beta$  2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
- Granier, S. & Kobilka, B. A new era of GPCR structural and chemical biology. *Nat. Chem. Biol.* **8**, 670–673 (2012).
- Dror, R. O. et al. Activation mechanism of the  $\beta$ 2-adrenergic receptor. *Proc. Natl. Acad. Sci.* **108**, 18684–18689 (2011).
- Flock, T. et al. Universal allosteric mechanism for G $\alpha$  activation by GPCRs. *Nature* **524**, 173–179 (2015).
- Woehler, A. & Ponimaskin, E. G. G protein-mediated signaling: Same receptor, multiple effectors. *Curr. Mol. Pharmacol.* **2**, 237–248 (2009).
- Flock, T. et al. Selectivity determinants of GPCR–G-protein binding. *Nature* **545**, 317–322 (2017).
- Inoue, A. et al. Illuminating G-protein-coupling selectivity of GPCRs. *Cell* **177**, 1933–1947 (2019).
- de Juan, D., Pazos, F. & Valencia, A. Emerging methods in protein co-evolution. *Nat. Rev. Genet.* **14**, 249–261 (2013).
- Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
- Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).
- Rivoire, O., Reynolds, K. A. & Ranganathan, R. Evolution-based functional decomposition of proteins. *PLoS Comput. Biol.* **12**, e1004817 (2016).
- Isberg, V. et al. GPCRdb: An information system for G protein-coupled receptors. *Nucleic Acids Res.* **44**, D356–D364 (2016).
- Probst, W. C., Snyder, L. A., Schuster, D. I., Brosius, J. & Sealfon, S. C. Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol.* **11**, 1–20 (1992).
- Rovati, G. E., Capra, V. & Neubig, R. R. The highly conserved DRY motif of class AG protein-coupled receptors: Beyond the ground state. *Mol. Pharmacol.* **71**, 959–964 (2007).
- Wescott, M. P. et al. Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *Proc. Natl. Acad. Sci.* **113**, 9928–9933 (2016).
- Tehan, B. G., Bortolato, A., Blaney, F. E., Weir, M. P. & Mason, J. S. Unifying family A GPCR theories of activation. *Pharmacol. Ther.* **143**, 51–60 (2014).
- Zhou, Q. et al. Common activation mechanism of class A GPCRs. *Elife* **8**, e50279 (2019).
- Schönegge, A.-M. et al. Evolutionary action and structural basis of the allosteric switch controlling  $\beta$  2 AR functional selectivity. *Nat. Commun.* **8**, 1–12 (2017).
- Filipek, S. Molecular switches in GPCRs. *Curr. Opin. Struct. Biol.* **55**, 114–120 (2019).

24. Trzaskowski, B. *et al.* Action of molecular switches in GPCRs—theoretical and experimental studies. *Curr. Med. Chem.* **19**, 1090–1109 (2012).
25. Doré, A. S. *et al.* Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* **511**, 557–562 (2014).
26. Zhang, H. *et al.* Structural basis for selectivity and diversity in angiotensin II receptors. *Nature* **544**, 327–332 (2017).
27. Gether, U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr. Rev.* **21**, 90–113 (2000).
28. Eddy, M. T. *et al.* Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor. *Cell* **172**, 68–80 (2018).
29. Milligan, G. & Kostenis, E. Heterotrimeric G-proteins: A short history. *Br. J. Pharmacol.* **147**, S46–S55 (2006).
30. Oldham, W. M. & Hamm, H. E. Structural basis of function in heterotrimeric G proteins. *Q. Rev. Biophys.* **39**, 117 (2006).
31. Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. Structural determinants for activation of the  $\alpha$ -subunit of a heterotrimeric G protein. *Nature* **369**, 621–628 (1994).
32. Glukhova, A. *et al.* Rules of engagement: GPCRs and G proteins. *ACS Pharmacol. Transl. Sci.* **1**, 73–83 (2018).
33. Conklin, B. R., Farfel, Z., Lustig, K. D., Julius, D. & Bourne, H. R. Substitution of three amino acids switches receptor specificity of Gq $\alpha$  to that of Gi $\alpha$ . *Nature* **363**, 274–276 (1993).
34. Alhadeff, R., Vorobyov, I., Yoon, H. W. & Warshel, A. Exploring the free-energy landscape of GPCR activation. *Proc. Natl. Acad. Sci.* **115**, 10327–10332 (2018).
35. Wang, J., Hua, T. & Liu, Z.-J. Structural features of activated GPCR signaling complexes. *Curr. Opin. Struct. Biol.* **63**, 82–89 (2020).
36. Liu, X. *et al.* Structural insights into the process of GPCR-G protein complex formation. *Cell* **177**, 1243–1251 (2019).
37. Du, Y. *et al.* Assembly of a GPCR-G protein complex. *Cell* **177**, 1232–1242 (2019).
38. Sreekumar, K. R., Huang, Y., Pausch, M. H. & Gulukota, K. Predicting GPCR-G-protein coupling using hidden Markov models. *Bioinformatics* **20**, 3490–3499 (2004).
39. Kling, R. C., Lanig, H., Clark, T. & Gmeiner, P. Active-state models of ternary GPCR complexes: Determinants of selective receptor-G-protein coupling. *PLoS ONE* **8**, e67244 (2013).
40. Semack, A., Sandhu, M., Malik, R. U., Vaidehi, N. & Sivaramakrishnan, S. Structural elements in the Gas and Gq $\alpha$  C termini that mediate selective G protein-coupled receptor (GPCR) signaling. *J. Biol. Chem.* **291**, 17929–17940 (2016).
41. Johnston, C. A. & Siderovski, D. P. Structural basis for nucleotide exchange on Gai subunits and receptor coupling specificity. *Proc. Natl. Acad. Sci.* **104**, 2001–2006 (2007).
42. Natochin, M., Granovsky, A. E., Muradov, K. G. & Artemyev, N. O. Roles of the transducin  $\alpha$ -subunit  $\alpha 4$ -helix/ $\alpha 4$ - $\beta 6$  loop in the receptor and effector interactions. *J. Biol. Chem.* **274**, 7865–7869 (1999).
43. Hu, J. *et al.* Structural basis of G protein-coupled receptor–G protein interactions. *Nat. Chem. Biol.* **6**, 541–548 (2010).
44. Bae, H. *et al.* Molecular determinants of selectivity in 5-hydroxytryptamine1B receptor–G protein interactions. *J. Biol. Chem.* **272**, 32071–32077 (1997).
45. Bae, H., Cabrera-Vera, T. M., Depree, K. M., Graber, S. G. & Hamm, H. E. Two amino acids within the  $\alpha 4$  helix of Gai1 mediate coupling with 5-Hydroxytryptamine1B Receptors. *J. Biol. Chem.* **274**, 14963–14971 (1999).
46. Fields, T. A. & Casey, P. J. Signalling functions and biochemical properties of pertussis toxin-resistant G-proteins. *Biochem. J.* **321**, 561–571 (1997).
47. Cabrera-Vera, T. M. *et al.* Insights into G protein structure, function, and regulation. *Endocr. Rev.* **24**, 765–781 (2003).
48. Liu, W. & Northup, J. K. The helical domain of a G protein  $\alpha$  subunit is a regulator of its effector. *Proc. Natl. Acad. Sci.* **95**, 12878–12883 (1998).
49. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinform.* **15**, 5–6 (2006).
50. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
51. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
52. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
53. Waskom, M. *et al.* mwwaskom/seaborn: v0.11.0 (September 2020). (2020) <https://doi.org/10.5281/ZENODO.4019146>.

## Acknowledgements

This work was supported by NRF-2019R1C1C1003933 (MJS, JH, KK and WY) and R&D Programs of DGIST (20-CoE-BT-01) funded by the Ministry of Science and ICT of Korea (MJS and WY) and NRF-2019R1A5A2027340 (KYC).

## Author contributions

M.J.S. and W.Y. designed the research, performed the research, interpreted the data, and wrote the paper (with input from all authors). J.H., and K.K. performed the research, and interpreted the data. K.Y.C. interpreted the data and wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87251-6>.

**Correspondence** and requests for materials should be addressed to W.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021