

RESEARCH ARTICLE

GPCR-SAS: A web application for statistical analyses on G protein-coupled receptors sequences

José Carlos Gómez Tamayo^{1‡}, Mireia Olivella², Santiago Ríos¹, Marlous Hoogstraat³, Angel Gonzalez¹, Eduardo Mayol¹, Xavier Deupi⁴, Mercedes Campillo¹, Arnau Cordermi^{1*}

1 Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain, **2** Grup de recerca de Bioinformàtica i Estadística Mèdica, Universitat de Vic, Barcelona, Spain, **3** Department of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands, **4** Laboratory of Biomolecular Research and Condensed Matter Theory Group, Paul Scherrer Institute, Villigen, Switzerland

‡ Current address: Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

* arnau.cordermi@uab.cat



OPEN ACCESS

Citation: Gómez Tamayo JC, Olivella M, Ríos S, Hoogstraat M, Gonzalez A, Mayol E, et al. (2018) GPCR-SAS: A web application for statistical analyses on G protein-coupled receptors sequences. PLoS ONE 13(7): e0199843. <https://doi.org/10.1371/journal.pone.0199843>

Editor: Alexey Porollo, Cincinnati Children's Hospital Medical Center, UNITED STATES

Received: March 2, 2018

Accepted: June 14, 2018

Published: July 25, 2018

Copyright: © 2018 Gómez Tamayo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data needed to replicate the results of this study are contained within the paper and its Supporting Information files. The GPCR-SAS database (<http://lmc.uab.cat/gpcrsas>) contains 1) a sequence alignment generated by the authors and 2) classification schemes that were collected from different third-party sources and curated as described in the Methods section.

Funding: This study was supported by the Spanish Ministerio de Economía y Competitividad and

Abstract

G protein-coupled receptors (GPCRs) are one of the largest protein families in mammals. They mediate signal transduction across cell membranes and are important targets for the pharmaceutical industry. The G Protein-Coupled Receptors—Sequence Analysis and Statistics (GPCR-SAS) web application provides a set of tools to perform comparative analysis of sequence positions between receptors, based on a curated structural-informed multiple sequence alignment. The analysis tools include: (i) percentage of occurrence of an amino acid or motif and entropy at a position or range of positions, (ii) covariance of two positions, (iii) correlation between two amino acids in two positions (or two sequence motifs in two ranges of positions), and (iv) snake-plot representation for a specific receptor or for the consensus sequence of a group of selected receptors. The analysis of conservation of residues and motifs across transmembrane (TM) segments may guide the design of more selective ligands or help to rationalize activation mechanisms, among others. As an example, here we analyze the amino acids of the “transmission switch”, that initiates receptor activation following ligand binding. The tool is freely accessible at <http://lmc.uab.cat/gpcrsas/>.

Introduction

G protein-coupled receptors (GPCRs) are one of the most prevailing protein families in mammalian genomes [1] and the major protein family as drug targets, with about one third of marketed compounds targeting them [2, 3]. They are involved in most signal transduction processes across membranes, including the response to hormones and neurotransmitters and the senses of sight, smell and taste. GPCRs transduce extracellular signals across the cell membrane through G protein dependent but also through G protein independent processes [4]. GPCRs are classified into six families or classes named A to F according to sequence

European Regional Development Fund (SAF2015-74627-JIN) (AC).

Competing interests: The authors have declared that no competing interests exist.

similarities [5–7]. Only classes A, B, C and F are present in humans and class A or rhodopsin-like comprises, by far, the largest number of members [8]. The available GPCRs crystal structures showed a conserved TM structure with a common fold formed by an extracellular N-terminus, seven transmembrane helices (TM1-7), connected by alternating intracellular (ICL1 to ICL3) and extracellular (ECL1 to ECL3) hydrophilic loops, and a cytoplasmic C-terminus [9–11] that starts with an α -helix (Hx8) parallel to the cell membrane in classes A, B and F. This structural similarity in the TM domain facilitates comparative analysis between members of the family. This was early recognized by Ballesteros and Weinstein when they developed the common residue numbering system [12], and has been exploited in the GPCRdb [13], a dedicated database for GPCRs, with sequence, structural and ligand information.

Here we present GPCR-SAS, a web application that permits easy comparison and statistical analysis of sequence positions or motifs within the TM helices and helix 8 across GPCRs of classes A, B, C and F. Our tool can be of help in identifying residues undergoing correlated evolution, and thus represents a useful instrument to rationalize ligand selectivity, G protein recognition or receptor activation, among others. As an example, here we illustrate the utility of GPCR-SAS by analyzing the amino acids that belong to the “transmission switch”, that initiates receptor activation following ligand binding [14–16].

Material and methods

GPCR-SAS web application

GPCR-SAS is a web application freely accessible at <http://lmc.uab.cat/gpcrsas/>. The main tool is written in Python (version 2.7; available at <http://www.python.org>) and employs Django framework (version 1.5; available at <https://djangoproject.com>). The application relies on a MySQL database (version 5.1.73; available at <https://www.mysql.com/>) that contains the sequence alignments of all GPCRs sequences and previously reported classification schemes [8, 13, 17–19]. Its design and implementation permit automatic incorporation of additional sequences as they are incorporated or edited in the UniProt [20].

The input of GPCR-SAS consists in two main sections: (i) *Positions and Sequence*: where the user can introduce a position/set of positions and a residue or sequence motif and (ii) *Classification*, that provides navigable multilevel hierarchical classification of GPCRs in families, branches and various levels of subfamilies according to different implemented schemes. The output provides conservation, covariance or correlation analysis for different classification sub-levels depending on the input provided, as schematically shown in Fig 1.

Database of aligned GPCR sequences

GPCRS-SAS database currently contains multiple sequence alignments of the TM helices and helix 8 of 2982 GPCRs sequences of all species: 2377 class A, 206 class B, 111 class C and 297 class F. Amino acid sequences for all GPCRs belonging to classes A, B, C and F were retrieved from the UniProtKB/Swiss-Prot database (<http://www.uniprot.org>) on May the 15th 2018 [20]. For each class, the sequences of receptors with an available crystal structure were aligned in a first step using MultiProt [21]. This structural alignment was used to define consensus boundaries for each TM helix on every GPCR class and to construct an initial sequence profile aware of the TM segments. Because the main purpose of GPCR-SAS is the comparison between sequences, we chose to be conservative in terms of gap introduction. Otherwise the alignment cannot be used for unambiguous predictions [22]. In the class A, irregularities observed between different structure patterns were handled using gaps on TMs 2 and 5 as previously described [23–25]. For each class, we aligned the sequences in three steps: first human sequences, next the rest of mammalian sequences and finally, the remaining vertebrate and

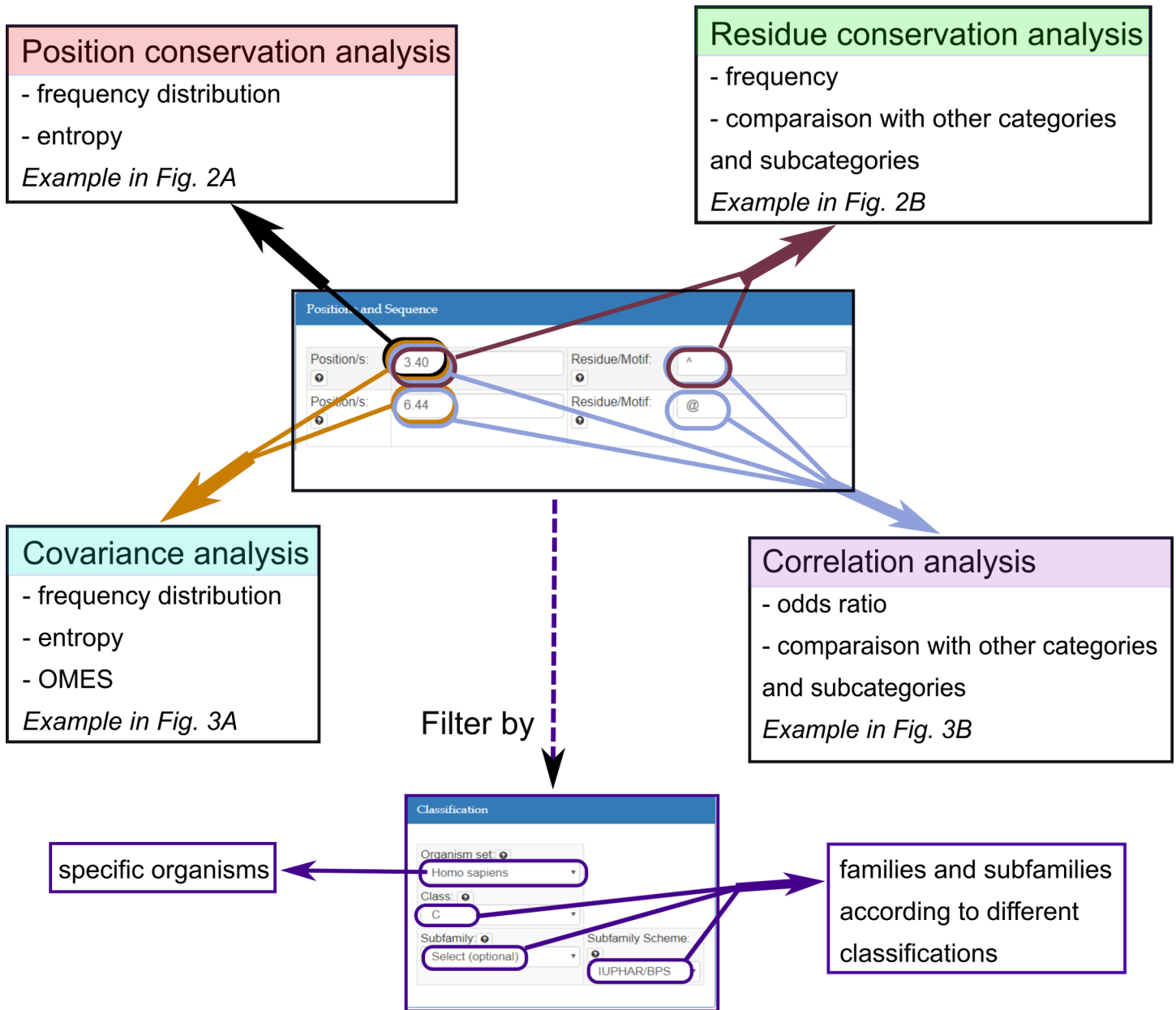


Fig 1. Schematic representation of the input forms of GPCR-SAS and the possible outcomes. The input of GPCR-SAS consists in two main sections: (i) Positions and Sequence: where the user can introduce a position/set of positions and a residue or sequence motif and (ii) Classification, that provides navigable multilevel hierarchical classification of GPCRs in families, branches and various levels of subfamilies according to different implemented schemes.

<https://doi.org/10.1371/journal.pone.0199843.g001>

invertebrate sequences. All multiple sequence alignments were performed with Clustal Omega [26]. At each step the alignment was manually curated ensuring lack of gaps on TM regions other than those at TMs 2 and 5 in class A receptors [23, 24] or those associated to trivial deletions. Finally, the alignments for each class were assembled based on the structural alignment between crystallized receptors from the different classes [27, 28]. The final alignment is similar to the one presented by Cvicsek and collaborators [22] except for the gaps in TMs 2 and 5 (positions 2x551 and 5x461 according to the generic residue numbering [25]), whereas it does not contain other gaps that appear in the GPCRdb alignment [25].

The updated list of GPCRs sequences and the alignment are available at the help panel (<http://lmc.uab.cat/gpcrsas/about/>). By default, the class A set in the GPCRS-SAS database consists of the 1824 non-olfactory GPCRs, since olfactory receptors are excluded to avoid biasing of results towards this subfamily—it accounts for almost two thirds (561 receptors) of the human class A GPCRs. The sequence alignment for TM helices and the short helix 8 perpendicular to de membrane for a selected group of receptors is shown in [S1 Fig](#). Regular updates are planned every 3 months. This step implies aligning new sequences to the previous alignment and assigning the different classification categories. We have scheduled a yearly full update where we will rebuild the complete alignment to account for possible changes in the structural alignment due to new structures.

Sequence numbering scheme

Each position in the TM segments is numbered according to the Ballesteros & Weinstein numbering scheme for class A GPCRs [12]. In this numbering, the position of each residue is described by two numbers: the helix in which the residue is located and the position relative to a conserved residue in that helix, arbitrarily assigned to 50, separated by a dot. In class A, these amino acids are: N1.50 in TM1 (97.6% conserved in human class A excluding olfactory receptors; data from GPCR-SAS), D2.50 in TM2 (92.1%), R3.50 in TM3 (94.8%), W4.50 in TM4 (95.8%), P5.50 in TM5 (76.0%), P6.50 in TM6 (98.3%), and P7.50 in TM7 (93.7%). Although GPCRs of classes other than A most often do not have such conserved amino acids at these positions [25], extrapolation of the class A numbering-scheme is now possible thanks to structure-based sequence alignment between classes [27, 28]. [S1 Fig](#) shows the sequence alignment for the TM helices of a selected group of receptors from different classes.

GPCR classifications

GPCR-SAS queries permit to filter receptor sequences based on previously reported classification systems including: (i) Fredriksson [8], based on a phylogenetic analysis of human GPCR sequences (branches and one subfamily level); (ii) GPCRdb [29], which uses a pharmacologic classification of the receptors (three subfamily levels); (iii) BIAS-PROF GDS [17], based on the comparison of the protein sequence using the physicochemical properties of the amino acids (two subfamily levels); (iv) Rognan [18], that relies on the phylogenetic analysis of 30 positions putative involved on the ligand binding site (one subfamily level), (v) Chabbert [19], which employs multidimensional scaling to cluster GPCRs (branches and one subfamily level), and (vi) GPCR SARfari (<https://www.ebi.ac.uk/chembl/sarfari/gpcrsarfari>), based on chemogenic data (three subfamily levels). As a complementary filter option, we defined five sequence sets based on different levels of taxonomic classification: Human, Mammals, Vertebrates, Eukaryotes and All (sequences for all species).

Analysis tools

The relevance of GPCR-SAS is its capability to analyze conservation of residues or sequence motifs across TM segments of GPCRs and to identify covariance between two positions or correlation between specific amino acids for various classification schemes at different classification levels (class, branch and subfamilies) and within five different taxonomic sets.

Conservation analysis

The entropy can be given for a single position (i.e. 3.50) or for a range of consecutive (i.e. 3.50–3.54) or non-consecutive (3.50, 4.50, 5.50) positions. When the input is a single position, a graph

with the amino acids counts is also displayed. For each position or range of positions i , the entropy of the information contained $H(i)$ is computed according to Shannon's theorem [30] as:

$$H(i) = - \sum_x p_x(i) \log_b p_x(i) \tag{1}$$

where $p_x(i)$ is the probability mass function for the amino acid(s) at position (or group of positions) i . The logarithm base b serves to scale the entropy in the range [0, 1] for one or more positions. Consequently, b is 20^n , with n being the number of positions used for the calculation. A position or group of positions with low variability (high conservation) has an entropy $H(i)$ close to 0, while high variability (low conservation) gives an entropy close to 1.

The percentage of occurrence of an amino acid or motif in a specific position(s) (i.e. N 7.49, P 7.50) can also be computed. In this case, the output allows comparison to other categories and subcategories. To represent residues or motifs with specific physicochemical properties the user can utilize one-character wildcards as a residue or as part of a motif with the following correspondences: 'X' (any amino acid), '@' (aromatic, W/Y/F/H), '~' (apolar, I/L/V/A/F/P), '+' (positively charged, R/H/K), '-' (negatively charged, D/E), '*' (charged, R/H/K/D/E) and '^' (polar, D/E/N/Q/K/R/H/S/T/C/W/Y).

Covariance analysis

To analyze the covariance of two positions, GPCR-SAS uses the Observed Minus Expected Squared (OMES) [31] that is based on a χ^2 test, and a corrected mutual information method (MIp) [32]. Both methods have previously been employed by Pelé et al. to identify evolutionary hubs between pairs of residues in GPCRs [33].

OMES calculates the difference between the observed and expected frequencies of each possible pair of amino acids or groups of amino acids (x, y), at positions i and j of the alignment:

$$OMES(i, j) = \frac{1}{N(i, j)} \sum_{x, y} (N_{x, y}^{obs}(i, j) - N_{x, y}^{exp}(i, j))^2 \tag{2}$$

with $N(i, j)$, $N_{x, y}^{obs}$ and $N_{x, y}^{exp}$ being the number of sequences in the alignment with non-gapped residues, the observed frequency and the expected frequency, respectively, at positions (or list of positions) i and j .

The MI content $MI(i, j)$ between two positions (or lists of positions) i and j on an alignment is based on the probability of joint occurrence of events and is defined as:

$$MI(i, j) = \sum_{x, y} p_{x, y}(i, j) \ln \frac{p_{x, y}(i, j)}{p_x(i) p_y(j)} \tag{3}$$

where $p_x(i)$, $p_y(j)$ and $p_{x, y}(i, j)$ are respectively the frequencies of amino acid x at position i , amino acid y at position j and the amino acid pair (x, y) at positions i and j .

The corrected MIp version is defined as:

$$MIp(i, j) = MI(i, j) - \frac{\frac{1}{n-1} \sum_{j \neq i} MI(i, j) \frac{1}{n-1} \sum_{i \neq j} MI(i, j)}{\frac{2}{n(n-1)} \sum_{i, j} MI(i, j)} \tag{4}$$

with n being the number of columns in the alignment.

To evaluate the statistical significance for the computed OMES and MIp values, GPCR-SAS provides the Z-scores and the associated p-values, which are computed by comparing with the mean value for all combinations of two positions.

Correlation analysis

To determine the correlation between two sequence positions, the occurrence of the amino acid or motif at the first position or range of positions is associated with the occurrence of the amino acid or motif at the second position or range of positions. The occurrences are used to compute an odds ratio (and the associated 95% confidence interval) that estimates how strongly the presence/absence of one of the first amino acids or motif is correlated with the presence/absence of the second amino acid or motif. To facilitate the comparison with the other categories at the chosen level of classification and in the subcategories, the output of a correlation analysis also returns the same analysis for these groups.

Snake-plot representations

GPCR-SAS can also provide snake-plot representations for the sequence of a specific receptor or for the consensus sequence of a group of receptors (S2 Fig). Each residue is represented by a circle with a letter in gray-gradient (representing the frequency of the residue on the class) and an outline in blue-gradient (representing the frequency of the residue on the selected group of receptors). In single-receptor snake-plots, residues colored in green are those that do not match the most conserved residue for the selected group of receptors (i.e. the selected subfamily).

Results

To illustrate the use of GPCR-SAS, we analyzed conservation, covariance and correlation of the residues of the “transmission switch”, which is one of the initial steps of receptor activation following ligand binding in class A GPCRs [14–16]. The “transmission switch” involves amino acids at positions 3.40, 5.50 and 6.44. Rearrangement of the packing between these residues following ligand binding at the extracellular side of the receptor weakens the interface between TM helices 5 and 6 and triggers local conformational changes that are transmitted towards the cytoplasmic side, where G proteins and β -arrestins bind [34, 35] (Fig 2A). By using GPCR-SAS we will determine if the “transmission switch” is likely to exist in receptors belonging to classes B and C of GPCRs.

The transmission switch in Class A GPCRs

In class A GPCRs, the comparison between crystal structures of inactive versus active states showed a rearrangement of residues 3.40, 5.50 and 6.44 of the “transmission switch” [34, 36]. The changes observed in the β_2 -adrenergic receptor (β_2 -AR) are displayed in Fig 2B and 2C. The percentage of occurrence of each amino acid at position 3.40 for human non-olfactory class A GPCRs computed with GPCR-SAS is shown in Fig 3A. The information displayed includes also the entropy of the position (as a measure of variability, see Methods) and a histogram with the counts of each amino acid at this position grouped according to subfamily categories. The most prevalent residue in this position is Ile (39.2% conserved), followed by Val (24.1%) and Leu (11.3%), all of them sharing hydrophobic properties. The entropy value of 0.6 indicates a moderate variability (entropy values ranges between 0 and 1, see Methods). The histogram shows that most amine receptors, including the β_2 -AR, feature Ile at this position. Indeed, the frequency of Ile raises to 76.2% when restricting the query to the amine subfamily (according to Fredriksson’s classification scheme [8]). The output also provides (“Click to show receptors” button) the list of UniProt entry names for the receptors that matched the query. Clicking on an amino acid allows easy comparison with the same position in the other categories at the same level of classification (Same-level button) and within the categories of

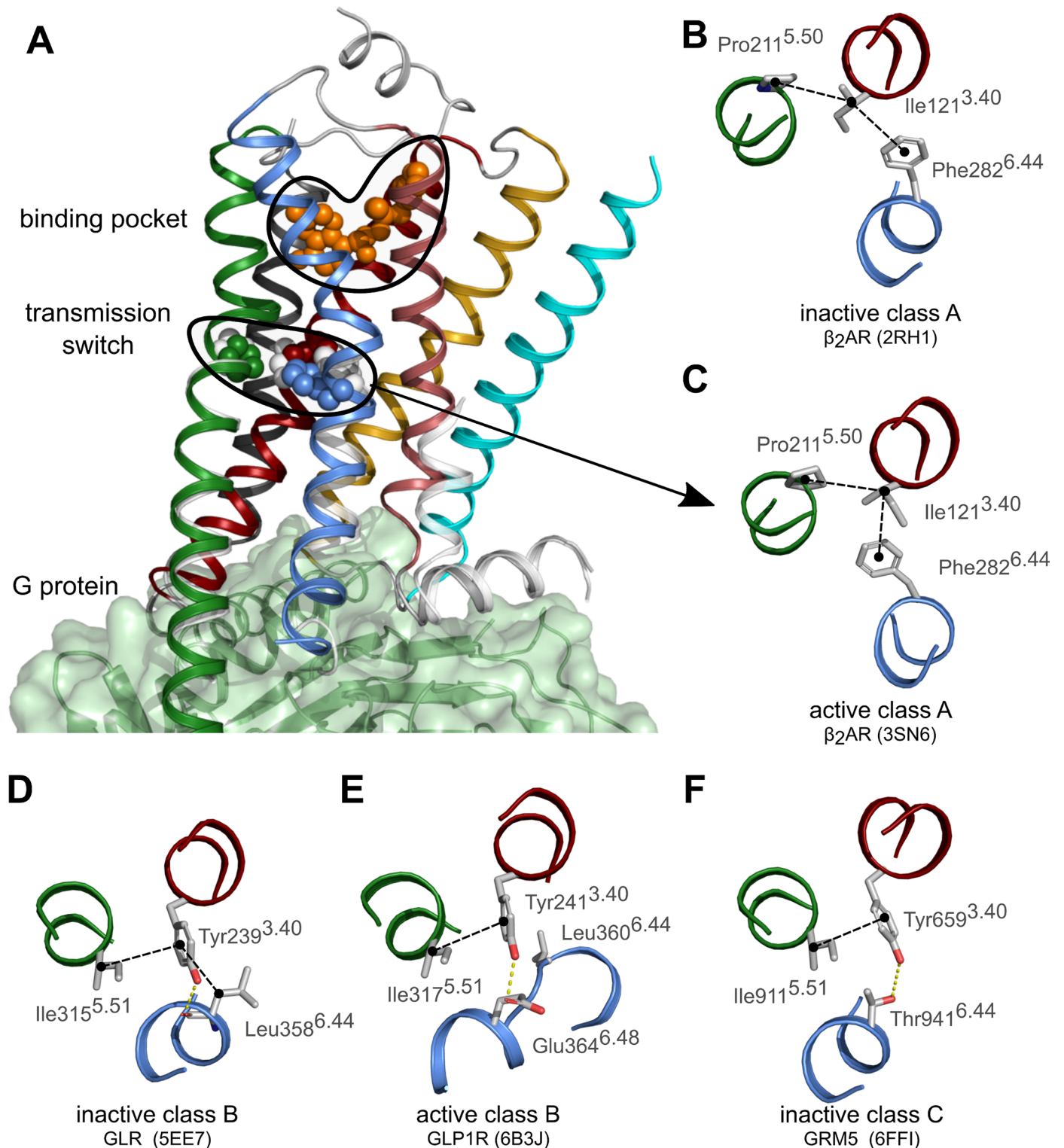


Fig 2. The transmission switch in the crystal structures of GPCRs. (A) Cartoon representation of the crystal structure of the active β_2 -adrenergic receptor (PDB id 3SN6) in complex with the G protein (green surface) illustrating the localization of the transmission switch just below the orthosteric binding pocket. The sidechains of residues of the transmission switch (3.40, 5.50 and 6.44) and the ligand (in orange) are shown as spheres. The color-code for the TM helices is 1:cyan, 2:gold, 3:red; 4:dark-gray, 5:green, 6_blue, 7:pale-red. Helix 8 and loops are shown in light-gray. Superposed (in white) the cytoplasmic ends of TMs 5, 6 and 7 and the residues of the transmission switch in the inactive structure (PDB id 2RH1). (B-F) Detail of the “transmission switch” in the crystal structures of inactive (B) and active (C) β_2 -adrenergic receptor (ADRB2), inactive glucagon receptor (GLR) (D), active glucagon-like peptide 1 receptor (GLP1R) (E), and inactive

metabotropic glutamate receptor (GRM5) (F). In receptors of classes B and C, residue 3.40 does not contact the residue at position 5.50 but rather the residue at position 5.51. (E-F) Dashed black lines indicate hydrophobic contacts whereas dashed yellow lines indicate hydrogen bonds.

<https://doi.org/10.1371/journal.pone.0199843.g002>

the child subfamilies (Sub-level button). In this particular example, the categories at the same level mean the other receptor classes (B, C and F) and the child categories are Fredriksson's branches [8]. The output of the query for Ile (shown in Fig 3D) indicates that this residue is rarely/never found at position 3.40 in classes B (0%), C (13.6%) and F (5.6%). Regarding the sublevels, GPCR-SAS tells that Ile is preferentially found at the α (48.0%) and δ branches

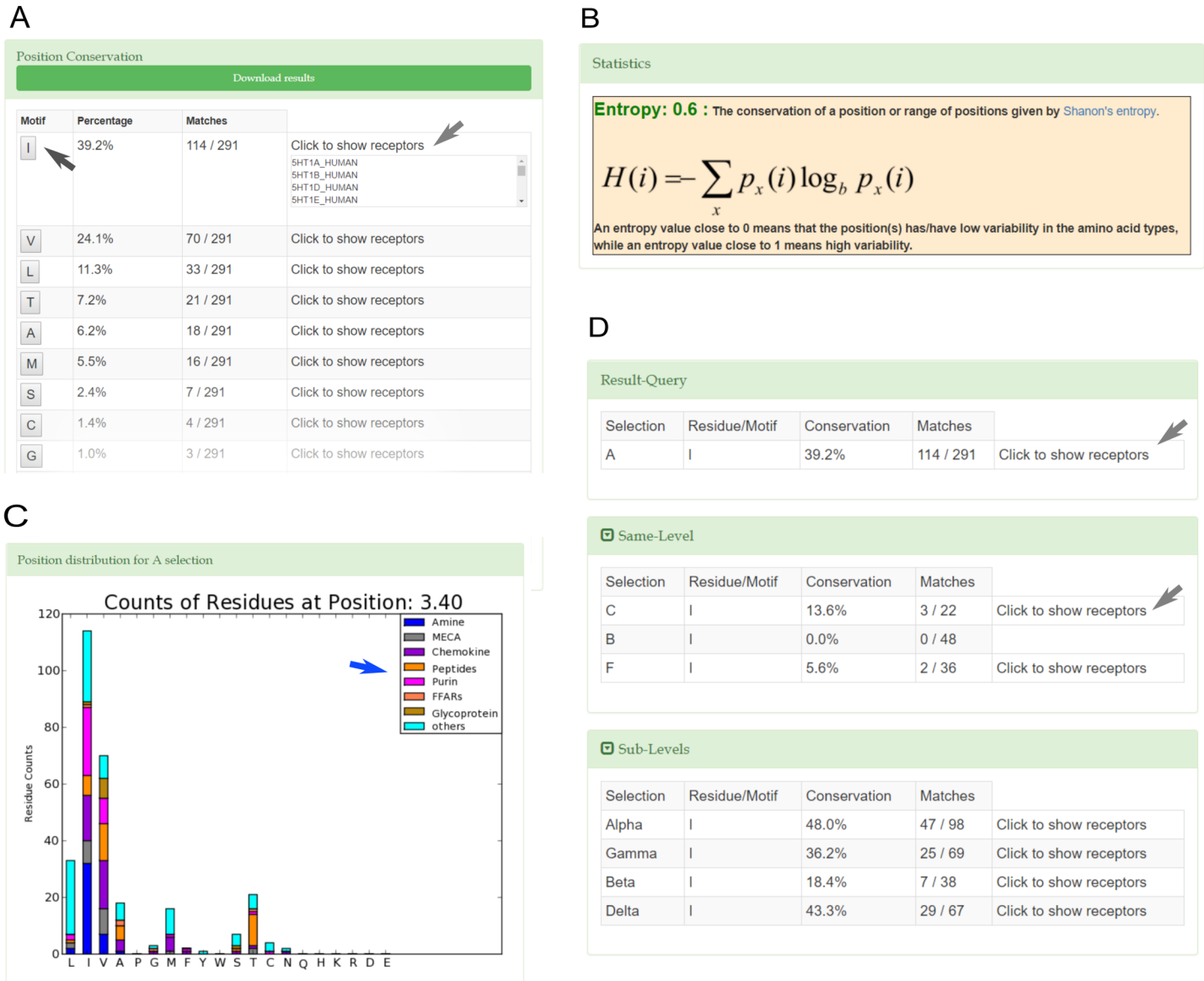


Fig 3. Conservation analysis queries for position 3.40 (A-C) and for Ile at position 3.40 (D) in human non-olfactory class A GPCRs. (A) the frequencies of amino acids at position 3.40; (B) the histogram of the amino acid frequencies at position 3.40 for the most the major subcategories (blue arrow); (C) the entropy of position 3.40; (D) the frequencies of Ile at position 3.40 for class A GPCRs compared the other classes (B, C and F; Same Level panel) and for the different class A branches (Sub-Level panel). Gray arrows indicate that it is possible to get the list of receptors that contain a certain residue or motif; the black arrow indicates that a click on the residue type in (A) provides the output displayed in (D).

<https://doi.org/10.1371/journal.pone.0199843.g003>

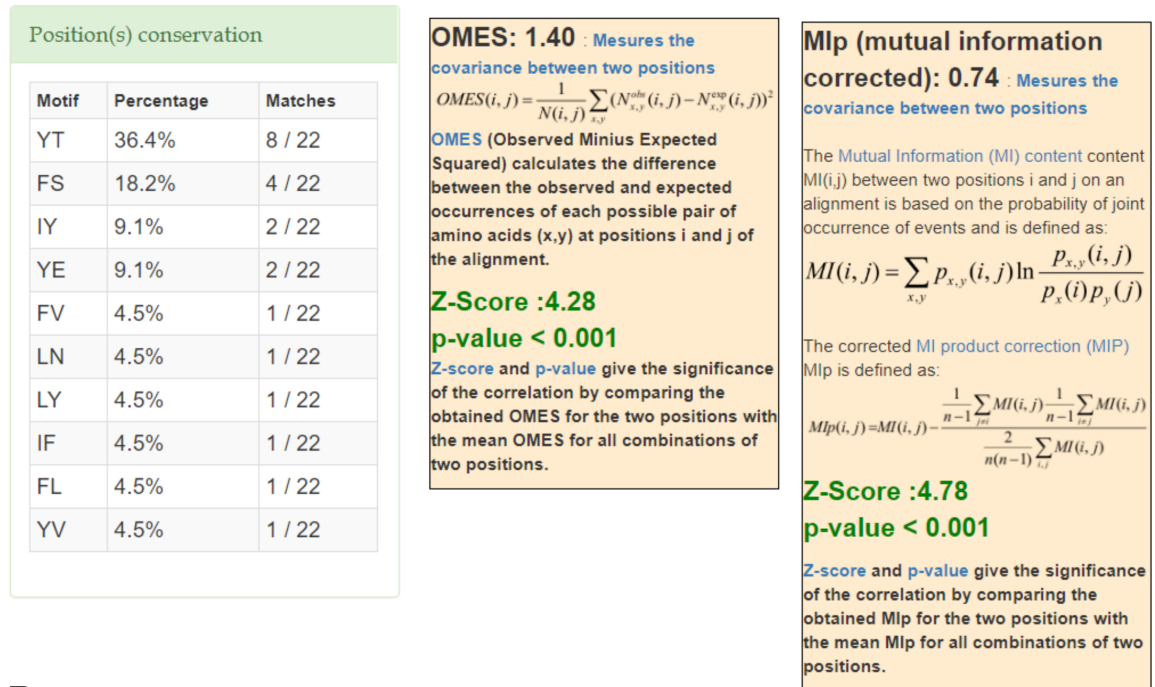
(43.3%). To expand the initial search, we used the wildcard for apolar amino acids (see [Methods](#)) which also accounts for Leu, Val, Ala, Phe and Pro. The frequencies for apolar residues add to 86.9% in class A, 79.2% in class B, 50% in class C, 69.4% in class F GPCRs. Thus, this analysis shows that despite Ile3.40 is not conserved in GPCRs of classes B, C and F, these receptors have mostly kept hydrophobic amino acids. Similar queries for the content of positions 5.50 and 6.44 in class A GPCRs reveals that Pro (78.4%) and aromatic residues (Phe:80.4%, Tyr:6.9%), respectively, are the most prevalent amino acids. The β_2 -AR has both Pro5.50 and Phe6.44 (see [Fig 2B and 2C](#)). To identify the most common residue triad, the three positions for human non-olfactory class A GPCRs were used as query (comma-separated in the box position). GPCR-SAS returns that the most frequent “transmission switch” residues are Ile3.40-Pro5.50-Phe6.44 as in the β_2 -AR (32.3%), Val3.40-Pro5.50-Phe6.44 (16.5%), and other triads with a percentage of occurrence smaller than 5% each. Overall, the triad Ile/Val3.40-Pro5.50-Phe6.44 comprises nearly half of the class A GPCRs.

Transmission switch in classes B and C GPCRs

Next, we analyzed positions 3.40, 5.50 and 6.44 in GPCRs of classes B and C. GPCR-SAS shows that human class B GPCRs mostly feature aromatic residues at position 3.40 (Tyr:66.7% and Phe:20.8%), aliphatic residues at position 5.50 (39.6% Ile, 20.8% Val, 14.6% Ala) and Leu (64.6%) or Phe (20.8%) at position 6.44. Compared to class A GPCRs, class B receptors have switched residue types at positions 3.40 and 6.44 (from aliphatic-aromatic in class A to aromatic-aliphatic in class B), but still exhibit conserved residue types in these positions in this class. The results are compatible with our recent proposal, based on mutagenesis studies, that residues Phe3.40 and Leu6.44 form the “transmission switch” in the corticotropin-releasing factor 1 receptor [28]. The crystal structures of class B GPCRs show that the side chain of Tyr/Phe3.40 interacts with Leu/Phe6.44 in the inactive state but that this interaction is lost in the active state ([Fig 2D and 2E](#)). In addition, in the glucagon family of receptors (which contain Tyr3.40) the hydroxyl group of Tyr forms a hydrogen bond with the backbone of residue 6.44 in the inactive state and with the side-chain of Glu6.48 in the active state. GPCR-SAS tells that Glu6.48 is exclusive of the glucagon family of receptors. The most common pair for positions 3.40–6.44 is Phe/Leu. Class B receptors also lack the characteristic Pro at position 5.50 present in class A GPCRs.

For human class C receptors, GPCR-SAS shows that position 3.40 mostly contains an aromatic residue (Tyr:50%, Phe:27.3%; entropy 0.4), but position 6.44 exhibits more variability (Thr:36.4%, Ser:18.2%, Tyr:13.6%, Glu:9.1%, Val:9.1%; entropy of 0.6). The most common pairs of residues at these positions in the class C receptors are aromatic-polar (Tyr-Thr:36.4%, Phe-Ser:18.2%, Tyr-Glu:9.1%) and aliphatic-aromatic (Ile-Tyr: 9.1%). The subfamily of metabotropic glutamate receptors (mGluRs) all contain Tyr3.40 and Thr6.44 (see mGluR5 in [Fig 2F](#)). Like class B receptors, class C receptors lack Pro5.50. We next analyzed the covariance of positions 3.40 and 6.44, that is, if changes in position 3.40 occur together with changes in position 6.44. Both GPCR-SAS revealed statistically significant covariance using OMES and MIP analyses (p-values <0.001 in both cases) with Z-scores of 4.28 and 4.78 respectively ([Fig 4A](#)). Furthermore, A GPCRs-SAS correlation analysis for aromatic residues at position 3.40 and polar residues at position 6.44 provided an odds ratio of 18.67 (with a 95% confidence interval of 1.50 to 232.3). For better statistics, we increased the number of sequences by using the “All organism” set. In this case we got a stronger association (odds ratio of 40.89) and a narrower confidence interval (13.48 to 123.99), ([Fig 4B](#)). Similarly, (with the “All organism set”) we obtain an odds ratio of 28.27 for an apolar residue at position 3.40 and an aromatic residue at position 6.44 (with a 95% confidence interval between 3.67 and 217.83), supporting a coordinated role for both amino acids. This covariance of positions 3.40 and 6.44, together with

A



B

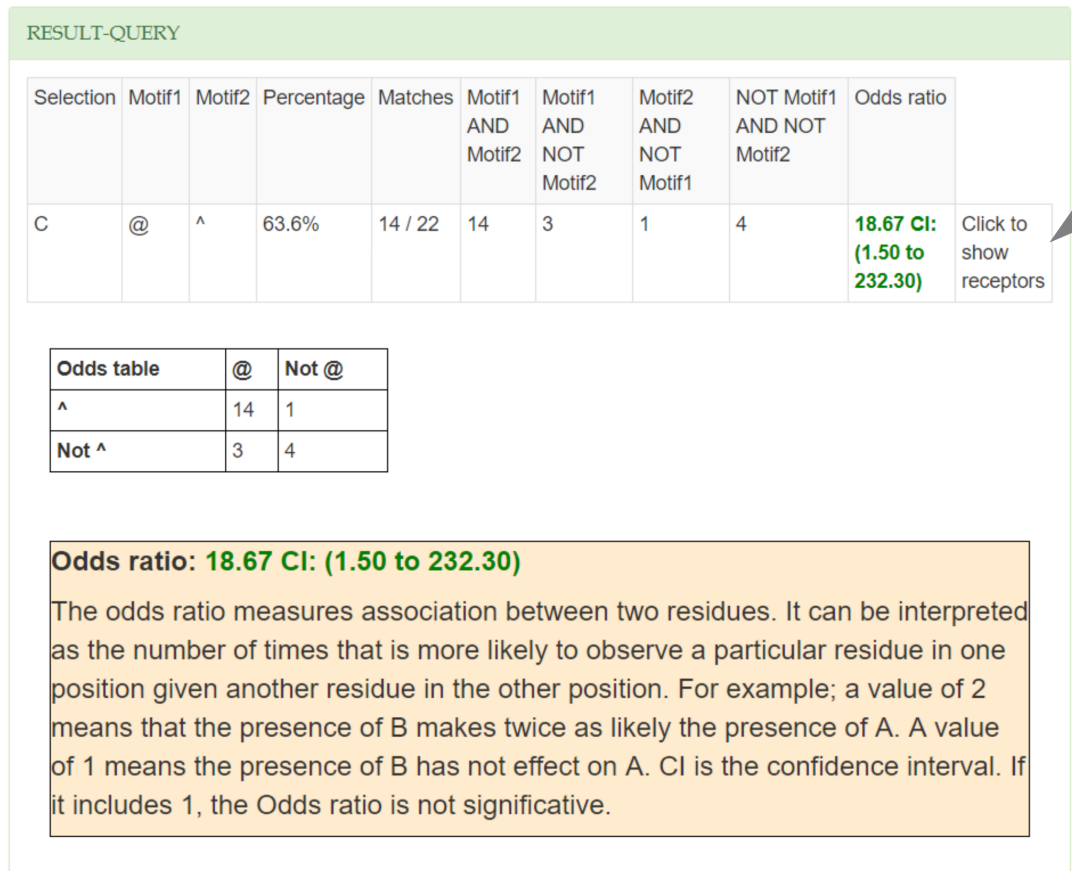


Fig 4. Covariance (A) and correlation (B) analysis for positions 3.40 and 6.44 in class C GPCRs (A). In (A) the left panel shows a list of all motifs and its frequencies and the right panel shows the statistical tests: the OMES and Mip analyses indicates statistical significant

covariance between both positions. In (B) the odds ratio for an aromatic residue at position 3.40 and a polar residue at position 6.44, which indicates statistical significant correlation. The gray arrow indicates that it is possible to get the list of receptors that contain a certain residue or motif.

<https://doi.org/10.1371/journal.pone.0199843.g004>

correlation of specific amino acids at these positions clearly suggest that both residues have a functional role in class C receptors as part of the same molecular switch. In fact, this is in accordance with our recent proposal, based on mutagenesis experiments and molecular modelling, that Tyr3.40 and Thr6.44 are part of the “transmission switch” in the GRM2 [37].

Conclusions

Here we present GPCRS-SAS, a web application that allows to perform frequency, covariance and correlation analyses for sequence positions or motifs in GPCRs. The tool takes advantage of the structural similarity in the TM domain of GPCRs and allows performing comparisons and statistical analyses of sequence positions or motifs within the TM helices and helix 8 for receptors of classes A, B, C and F. As an example of use, we here show its utility in the successful extrapolation of the “transmission switch” of class A GPCRs to classes B and C. GPCRS-SAS tells that nearly half the class A GPCRs have as triads for positions 3.40–5.50–6.44 either Ile-Pro-Phe (32.3%) or Val-Pro-Phe (16.5%). Class B and Class C receptors lack Pro5.50 and contain alternative, but also conserved, pairs at positions 3.40–6.44. The preferred pairs in class B receptors are Phe-Leu (45.8%) and Tyr-Leu (12.5%). This implies a switch between residue types compared to the same positions in class A. The preferred pairs in class C receptors are Tyr-Thr (36.4%) and Phe-Ser (18.2.1%). Overall, as supported by previous mutagenesis data [28, 37], this analysis suggests that different “transmission switches” involving the same positions but different residue types exist in classes A, B and C, and are likely to represent common activation pieces within the whole superfamily. The type of statistical analyses that GPCR-SAS performs can be used to find functionally important residues or groups of residues undergoing correlated evolution, and thus represents a useful instrument to rationalize ligand selectivity, G protein recognition or receptor activation, among others.

Supporting information

S1 Fig. The sequence alignment of the TM segments and helix 8 of selected GPCRs from classes A, B, C and F. The alignment is colored using Clustal scheme. Receptor abbreviations are as in UniProt. Greek letters α , β , γ and δ represent the branches described by Fredriksson and collaborators for the class A. Receptors of the class C do not have helix 8 according to the presently available structures.

(TIF)

S2 Fig. Snake-plot representation of the human Class C GPCRs. Each residue is represented with circle with a blue outline the one-letter amino acid code and the Ballesteros-and-Weinstein numbering scheme. The intensity of the blue outline represents the conservation of the position and the intensity of the gray-black letter represents the conservation of the specific amino acid relative to the consensus sequence.

(TIF)

Acknowledgments

This study was supported by the Spanish Ministerio de Economía y Competitividad and European Regional Development Fund (SAF2015-74627-JIN).

Author Contributions

Conceptualization: José Carlos Gómez Tamayo, Eduardo Mayol, Xavier Deupi, Arnau Cordoní.

Data curation: Santiago Ríos, Eduardo Mayol, Arnau Cordoní.

Formal analysis: Eduardo Mayol, Mercedes Campillo, Arnau Cordoní.

Funding acquisition: Arnau Cordoní.

Investigation: José Carlos Gómez Tamayo.

Methodology: José Carlos Gómez Tamayo, Santiago Ríos, Marlous Hoogstraat, Eduardo Mayol, Xavier Deupi.

Software: José Carlos Gómez Tamayo, Marlous Hoogstraat, Eduardo Mayol.

Supervision: Arnau Cordoní.

Validation: Mireia Olivella, Santiago Ríos, Angel Gonzalez, Eduardo Mayol, Mercedes Campillo, Arnau Cordoní.

Visualization: Mercedes Campillo.

Writing – original draft: Arnau Cordoní.

Writing – review & editing: Mireia Olivella, Angel Gonzalez, Mercedes Campillo, Arnau Cordoní.

References

1. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nature reviews Drug discovery*. 2006; 5(12):993–6. <https://doi.org/10.1038/nrd2199> PMID: 17139284
2. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nature reviews Drug discovery*. 2017; 16(1):19–34. <https://doi.org/10.1038/nrd.2016.230> PMID: 27910877.
3. Hauser AS, Attwood MM, Rask-Andersen M, Schioth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. *Nature reviews Drug discovery*. 2017; 16(12):829–42. <https://doi.org/10.1038/nrd.2017.178> PMID: 29075003.
4. Sun Y, McGarrigle D, Huang XY. When a G protein-coupled receptor does not couple to a G protein. *Mol Biosyst*. 2007; 3(12):849–54. Epub 2007/11/15. <https://doi.org/10.1039/b706343a> PMID: 18000562.
5. Kolakowski LF Jr. GCRDb: a G-protein-coupled receptor database. *Receptors & channels*. 1994; 2(1):1–7. PMID: 8081729.
6. Attwood TK, Findlay JB. Fingerprinting G-protein-coupled receptors. *Protein Eng*. 1994; 7(2):195–203. PMID: 8170923.
7. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G. GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res*. 2003; 31(1):294–7. PMID: 12520006; PubMed Central PMCID: PMC165550.
8. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*. 2003; 63(6):1256–72. Epub 2003/05/23. <https://doi.org/10.1124/mol.63.6.1256> PMID: 12761335.
9. Liapakis G, Cordoní A, Pardo L. The G-protein coupled receptor family: actors with many faces. *Curr Pharm Des*. 2012; 18(2):175–85. PMID: 22229577.
10. Katritch V, Cherezov V, Stevens RC. Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol*. 2013; 53:531–56. Epub 2012/11/13. <https://doi.org/10.1146/annurev-pharmtox-032112-135923> PMID: 23140243; PubMed Central PMCID: PMC3540149.
11. Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM. Molecular signatures of G-protein-coupled receptors. *Nature*. 2013; 494(7436):185–94. Epub 2013/02/15. <https://doi.org/10.1038/nature11896> PMID: 23407534.

12. Ballesteros JA, Weinstein H. Integrated Methods for Modeling G-Protein Coupled Receptors. *Methods Neurosci.* 1995; 25:366–428.
13. Isberg V, Mordalski S, Munk C, Rataj K, Harpsoe K, Hauser AS, et al. GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 2016; 44(D1):D356–D64. <https://doi.org/10.1093/nar/gkv1178> WOS:000371261700050. PMID: 26582914
14. Deupi X, Standfuss J, Schertler G. Conserved activation pathways in G-protein-coupled receptors. *Biochem Soc Trans.* 2012; 40(2):383–8. Epub 2012/03/23. <https://doi.org/10.1042/BST20120001> PMID: 22435816.
15. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S. Action of Molecular Switches in GPCRs—Theoretical and Experimental Studies. *Curr Med Chem.* 2012; 19(8):1090–109. WOS:000300644000002. <https://doi.org/10.2174/092986712799320556> PMID: 22300046
16. Sansuk K, Deupi X, Torrecillas IR, Jongejan A, Nijmeijer S, Bakker RA, et al. A structural insight into the reorientation of transmembrane domains 3 and 5 during family A G protein-coupled receptor activation. *Mol Pharmacol.* 2011; 79(2):262–9. Epub 2010/11/18. <https://doi.org/10.1124/mol.110.066068> PMID: 21081645.
17. Davies MN, Secker A, Freitas AA, Mendao M, Timmis J, Flower DR. On the hierarchical classification of G protein-coupled receptors. *Bioinformatics.* 2007; 23(23):3113–8. Epub 2007/10/25. <https://doi.org/10.1093/bioinformatics/btm506> PMID: 17956878.
18. Surgand JS, Rodrigo J, Kellenberger E, Rognan D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins.* 2006; 62(2):509–38. Epub 2005/11/19. <https://doi.org/10.1002/prot.20768> PMID: 16294340.
19. Deville J, Rey J, Chabbert M. An indel in transmembrane helix 2 helps to trace the molecular evolution of class A G-protein-coupled receptors. *J Mol Evol.* 2009; 68(5):475–89. Epub 2009/04/10. <https://doi.org/10.1007/s00239-009-9214-9> PMID: 19357801.
20. The UniProt C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2017; 45(D1):D158–D69. <https://doi.org/10.1093/nar/gkw1099> PMID: 27899622; PubMed Central PMCID: PMC5210571.
21. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins.* 2004; 56(1):143–56. <https://doi.org/10.1002/prot.10628> PMID: 15162494.
22. Cvicsek V, Goddard WA 3rd, Abrol R. Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLoS computational biology.* 2016; 12(3):e1004805. <https://doi.org/10.1371/journal.pcbi.1004805> PMID: 27028541; PubMed Central PMCID: PMC4814114.
23. Gonzalez A, Cordini A, Caltabiano G, Pardo L. Impact of helix irregularities on sequence alignment and homology modeling of G protein-coupled receptors. *Chembiochem: a European journal of chemical biology.* 2012; 13(10):1393–9. Epub 2012/07/05. <https://doi.org/10.1002/cbic.201200189> PMID: 22761034.
24. Becu JM, Pele J, Rodien P, Abdi H, Chabbert M. Structural evolution of G-protein-coupled receptors: a sequence space approach. *Methods Enzymol.* 2013; 520:49–66. Epub 2013/01/22. <https://doi.org/10.1016/B978-0-12-391861-1.00003-4> PMID: 23332695.
25. Isberg V, de Graaf C, Bortolato A, Cherezov V, Katritch V, Marshal FH, et al. Generic GPCR residue numbers—aligning topology maps while minding the gaps. *Trends Pharmacol Sci.* 2015; 36(1):22–31. <https://doi.org/10.1016/j.tips.2014.11.001> WOS:000347863100005. PMID: 25541108
26. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology.* 2011; 7:539. <https://doi.org/10.1038/msb.2011.75> PMID: 21988835; PubMed Central PMCID: PMC3261699.
27. Siu FY, He M, de Graaf C, Han GW, Yang D, Zhang Z, et al. Structure of the human glucagon class B G-protein-coupled receptor. *Nature.* 2013; 499(7459):444–9. Epub 2013/07/19. <https://doi.org/10.1038/nature12393> PMID: 23863937; PubMed Central PMCID: PMC3820480.
28. Spyridaki K, Matsoukas MT, Cordini A, Gkoutelias K, Papadokostaki M, Mavromoustakos T, et al. Structural-functional analysis of the third transmembrane domain of the corticotropin-releasing factor type 1 receptor: role in activation and allosteric antagonism. *J Biol Chem.* 2014; 289(27):18966–77. Epub 2014/05/20. <https://doi.org/10.1074/jbc.M113.544460> PMID: 24838244; PubMed Central PMCID: PMC3820480.
29. Vroiling B, Sanders M, Baakman C, Borrmann A, Verhoeven S, Klomp J, et al. GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res.* 2011; 39(Database issue):D309–19. Epub 2010/11/04. <https://doi.org/10.1093/nar/gkq1009> PMID: 21045054; PubMed Central PMCID: PMC3013641.
30. Shannon CE. A Mathematical Theory of Communication. *At&T Tech J.* 1948; 27(4):623–56. WOS: A1948UH04000002.

31. Fodor AA, Aldrich RW. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins*. 2004; 56(2):211–21. <https://doi.org/10.1002/prot.20098> PMID: 15211506.
32. Dunn SD, Wahl LM, Gloor GB. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*. 2008; 24(3):333–40. <https://doi.org/10.1093/bioinformatics/btm604> PMID: 18057019.
33. Pele J, Moreau M, Abdi H, Rodien P, Castel H, Chabbert M. Comparative analysis of sequence covariation methods to mine evolutionary hubs: examples from selected GPCR families. *Proteins*. 2014; 82(9):2141–56. Epub 2014/03/29. <https://doi.org/10.1002/prot.24570> PMID: 24677372.
34. Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*. 2011; 477(7366):549–55. Epub 2011/07/21. <https://doi.org/10.1038/nature10361> PMID: 21772288; PubMed Central PMCID: PMC3184188.
35. Kang Y, Zhou XE, Gao X, He Y, Liu W, Ishchenko A, et al. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature*. 2015; 523(7562):561–7. Epub 2015/07/23. <https://doi.org/10.1038/nature14656> PMID: 26200343; PubMed Central PMCID: PMC4521999.
36. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*. 2007; 318(5854):1258–65. <https://doi.org/10.1126/science.1150577> PMID: 17962520; PubMed Central PMCID: PMC2583103.
37. Pérez-Benito L, Doornbos MLJ, Cordero A, Peeters L, Lavreysen H, Pardo L, et al. Molecular Switches of Allosteric Modulation of the Metabotropic Glutamate 2 Receptor. *Structure*. 2017; 25(7):1153–62. <https://doi.org/10.1016/j.str.2017.05.021> PMID: 28648611