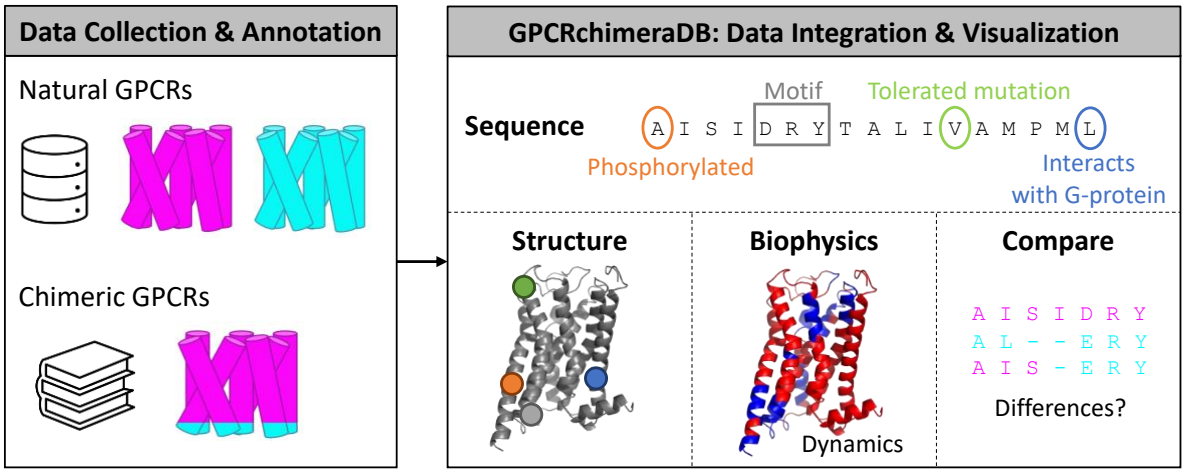
GPCRchimeraDB: A database of chimeric G-Protein Coupled Receptors (GPCRs) to assist their design

Charlotte Crauwels[1,2,3], Adrián Díaz[1,2,3], Wim Vranken[1,2,3,4,5*]

1. Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, Brussels, Belgium
2. Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium
3. AI Lab, Vrije Universiteit Brussel, Brussels, Belgium
4. Chemistry Department, Vrije Universiteit Brussel, Brussels, Belgium
5. Biomedical sciences, Vrije Universiteit Brussel, Brussels, Belgium.

* To whom correspondence should be addressed: wim.vranken@vub.be

Graphical Abstract



Abstract

G-Protein Coupled Receptors (GPCRs) are membrane proteins implicated in numerous diseases that have been studied for decades. However, despite their significance, many GPCRs remain poorly characterized and untargeted by drugs. Chimeric GPCRs have emerged as valuable tools for elucidating GPCR function by facilitating the identification of signaling pathways, resolving structures, and discovering novel ligands of poorly understood GPCRs. Such chimeric GPCRs are obtained by merging a well-characterized and a less-well-characterized GPCR at the intracellular loops, leveraging knowledge transfer from the well-characterized GPCR. However, despite the 170 chimeric GPCRs engineered to date, the design process remains largely trial-and-error and lacks a standardized approach. To address this gap, we introduce GPCRchimeraDB (https://www.bio2byte.be/gpcrchimeradb/), the first comprehensive database dedicated to chimeric GPCRs. It catalogs 170 chimeric receptors, identified through extensive literature review, and includes

1,758 natural GPCRs, enabling connections between chimeras and their parent receptors while facilitating the exploration of novel parent combinations. Both chimeric and natural GPCR entries are extensively described at the sequence, structural, and biophysical level through a range of visualization tools, with annotations from resources like UniProt and GPCRdb and predictions from AlphaFold, ESMFold and b2btools. Additionally, GPCRchimeraDB offers a GPCR sequence aligner and a feature comparator to investigate differences between natural and chimeric receptors as well as guidelines to support the design of novel chimeras. GPCRchimeraDB is therefore a resource to facilitate and optimize the design of new chimeras, so helping to gain insights into poorly characterized receptors and contributing to advances in GPCR therapeutic development.

Keywords:

G-Protein Coupled Receptor, GPCR, chimera, hybrid, protein design, database

Introduction

G-protein coupled receptors (GPCRs) are the largest superfamily of membrane proteins. With over 800 GPCRs identified in humans, these receptors make up more than ~4% of the human genome[1,2]. GPCRs serve as signal transducers, allowing cells to respond to a variety of external stimuli, including hormones, neurotransmitters, and sensory signals[1,2]. Consequently, mutations that affect their function can lead to a wide range of diseases, from neurodegenerative and psychiatric disorders to cardiovascular conditions and cancers[3,4]. However, despite their importance, 60 to 85% of the therapeutically relevant GPCRs remain undrugged[5]. Much of the difficulty in targeting GPCRs arises from their innate conformational flexibility, low shared sequence identity, undefined endogenous ligands, insolubility, cell-type specific expression, and off-target effects of their ligands[1,6,7].

Chimeric GPCRs (or hybrid GPCRs) are powerful tools for advancing our understanding of GPCRs. Their goal is to leverage what is known about one GPCR to gain deeper insights into the other, less understood, GPCR[6,8]. They so provide insights into the roles of specific residues in GPCR function[9,10], help decipher biological pathways[11,12], facilitate the elucidation of 3D structures in distinct activation states[13–15], and assist in identifying novel ligand binders[15]. Chimeric GPCRs are designed by merging regions from two parent GPCRs, typically a well- and not well-characterized GPCR (Figure 1A-B). The sites where the parent GPCRs merge into each other are usually situated at the limits of their intracellular (IC) loops, resulting in an extracellular (EC) parent that also provides the TMs and an IC parent.

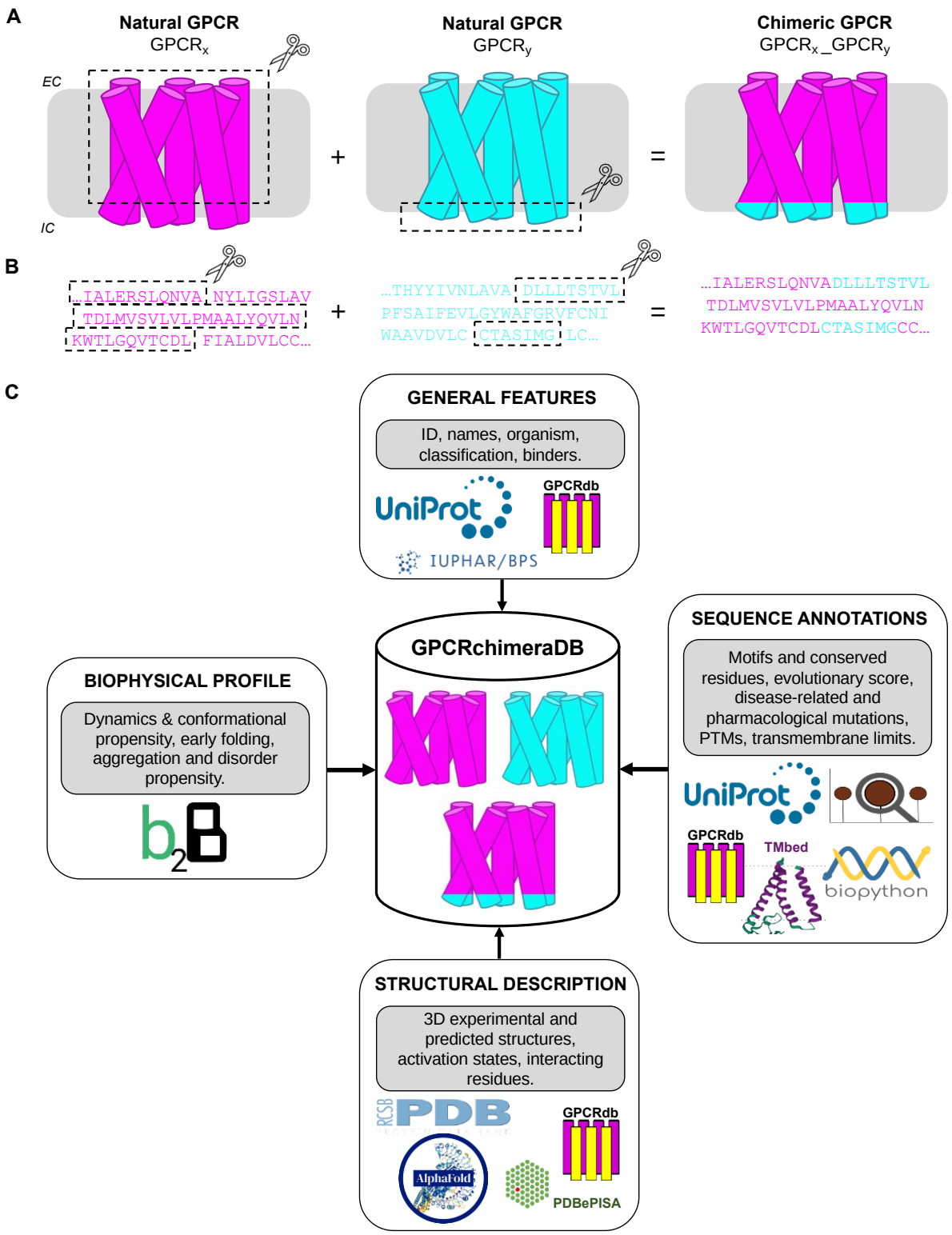


Figure 1: *GPCRchimeraDB framework.* ***A.*** *Schematic illustration of chimeric GPCR design.* ***B.*** *Introduction of the cutting points or sites, in the parent GPCRs to obtain the chimeric sequence.* ***C.*** *Features and annotations collected on GPCRchimeraDB to describe the natural and chimeric entries. The data was either extracted from a pre-existing database or generated with a predictor.*

Chimeric GPCRs capitalize on the common structure-function relationship among GPCRs[16–18]. Despite having less than 30% sequence identity, GPCRs share a similar activation mechanism, involving the displacement of transmembrane (TM) helices 2 and 6[19], and exhibit a highly conserved 3D structure[1,2]. This structure consists of seven TM helices that span the cell membrane and that are connected by three EC and IC loops. The key to successful chimeric design lies in selecting optimal cutting sites within the parent proteins to preserve the functional integrity of the EC+TM and IC fragments[8,16]: the binding of an agonist at the EC side needs to trigger conformational changes in the TMs to enable the binding of a G protein or β-arrestin at the IC side and allow signal transmission. This requires careful consideration of the sequence, structural, and biophysical (e.g. dynamic) features of the parent receptors. To date, 170 engineered chimeric GPCRs are available in the public domain. However, in spite of their relevance, no generic design approach to create chimeric GPCRs has been established yet, perhaps due to the lack of overview and annotations in the previous designs.

Here we present GPCRchimeraDB (https://www.bio2byte.be/gpcrchimeradb/), a comprehensive database designed to describe and visualize previously constructed chimeric GPCRs, so facilitating the creation of novel ones. As the first resource to standardize and consolidate data on chimeric GPCRs, it features 170 curated entries derived from extensive literature review. Additionally, it includes 1,758 natural class A mammalian GPCRs, enabling links between chimeric designs and their parent receptors. By integrating detailed annotations, advanced visualization tools, and predictive insights for both chimeric and natural GPCRs at the sequence, structural and biophysical level, GPCRchimeraDB builds upon and extends the descriptions provided by GPCRdb[20]. GPCRchimeraDB opens new opportunities for researchers to explore the field of GPCR chimeras and consider their potential as tools to investigate poorly characterized GPCRs of interest.

Material and methods

Collecting natural and chimeric GPCRs

The natural GPCR entries on GPCRchimeraDB consist of all reviewed class A (rhodopsin-like) mammalian GPCRs available on InterPro[21] (IPR017452). Based on models available from the AlphaFold Protein Structure Database[22] (AFDB), entries with more or fewer than 7 predicted transmembrane regions (TMs) were removed.

Chimeric GPCR entries were retrieved from the literature using a systematic search on databases such as PubMed and Google Scholar. The search included keywords such as ‘GPCR chimera’, ‘GPCR hybrid’, ‘GPCR design’, and ‘GPCR engineering’. Cross referencing was also critical to retrieve previously

designed chimeras. In addition, the Protein Data Bank[23] (PDB) was searched for GPCR chains connected to two different GPCR gene names, which were manually added to our dataset.

For each identified chimeric design, the following information was extracted from the related paper:

- Protein sequence: When the DNA sequence was provided, it was translated into a protein sequence using Python scripts[24].
- UniProt[25] IDs of the two parent GPCRs: If not provided, a BLAST[26] search against the SwissProt database[25] (E-value 0.001) was performed on the chimeric sequence to identify the parent GPCRs. Based on the section the parent GPCRs provided, we defined the EC and IC parent. The EC parent typically provides the EC loops and the TMs. Depending on the design, the EC parent can also provide some IC loops. The IC parent provides the (remaining) IC loop(s). More details about this terminology in the Results section 'Chimeric entries: definition'.
- Positions of the cutting sites: If not described, a pairwise local sequence alignment using Biopython was performed between the chimeric and the parent sequences ('pairwise2.align.localxx' function from biopython[27]).
- Mutations introduced into the chimeric sequence compared to the parent sequences.
- Experimental 3D structure(s) of the chimera (PDB id), if available.
- Functionality information: Functionality data is included as either a quantitative value (when available) or a qualitative value (yes/no). As detailed in the Results section 'Extensive description of chimeric and natural GPCRs', the reported functionality information is not normalized across different chimeric designs because of variability in experimental approaches and reporting practices in the source literature.
- Binding ligand(s) and G-protein(s), if available.
- Design type (Type 1/2/3 as defined in 'Chimeric entries: definition') and design purpose of the chimera.
- DOI of the paper.

Duplicate designs are considered as the same entry in the database but are referred to by multiple DOIs. Finally, as no naming convention exists for chimeric GPCRs, all chimeras were given a unique identifier and name: <*UniProt ID EC parent*>_<*UniProt ID IC parent* >_i and < *UniProt Protein Name EC parent*>_<*UniProt Protein Name IC parent* >_i, respectively and where i denotes the sequence in which they were identified in the literature (e.g, P43140_P18841_1 and ADA1A_RAT_ADA1B_MESAU_1, respectively). This data retrieval and formatting was handled by an in-house Python pipeline.

Description of the entries

For every natural and chimeric entry, a comprehensive set of features was collected and formatted using a Python pipeline divided into four sections: the general features (e.g., classification and ligands), the sequence annotations, the biophysical profile, and the structural description. These sections are described in more detail below, with the features and tools summarized in Figure 1C. The retrieved data was compiled into uniform JSON files, one per entry (see Zenodo for an example). These files were then uploaded individually to GPCRchimeraDB. This system simplifies updating entries when new information becomes available (more details in the section Technical architecture).

General features

For every natural GPCR, we gathered their UniProt ID, protein name (also called gene or abbreviated name), commonly used names and source organisms (from UniProt) as well as their pharmacological name (from GPCRdb). Moreover, based on information from GPCRdb, all natural entries were classified into families and subfamilies based on the endogenous ligands they bind. A phylogenetic-based subfamily classification was also added, where we extended the original version by Cvicek et al.[28] for a limited dataset to all mammals using an annotated phylogenetic tree generated with IQtree[29]. Additionally, information on the principal endogenous ligand(s) and, G-protein and β-arrestin subclass(es) the GPCRs bind was also extracted from GPCRdb and added to the database. Currently, there is no established naming convention or classification system for chimeric GPCRs. To address this in GPCRchimeraDB, we defined their pharmaceutical name and classification (including class, family, and ligand- and phylogenetic-based subclass) by concatenating the corresponding values of their EC and IC parents, in that order. For example, this approach results in 'Opsins & Adrenoceptors' for the receptor family of the chimera OPSD_BOVIN_ADA1A_HUMAN_1. More commonly this receptor will be called the Rhodopsin α1A-adrenoceptor (pharmaceutical name). Additionally, for the organism value for the chimeras, we concatenated the organisms from which the EC and IC parents originated to indicate the combinations of organisms explored in chimeric design (e.g. 'bovine & human' for OPSD_BOVIN_ADA1A_HUMAN_1).

Sequence annotations

Class A GPCRs are characterized by a set of well-conserved motifs, also known as microswitches, that are essential for the receptor's function. These motifs include the ionic lock switch (E/DRY/W), the transmission toggle switch (CWxP), the Tyr toggle switch (NPxxY), the PIF motif, and the hydrophobic and ionic locks[1]. Each TM in class A GPCRs also contains a highly conserved residue, identified as residue x.50 in Ballesteros and Weinstein numbering[30], which are in TM sequence order N, D, R, W, P,

P, and P (TM1-TM7). Each GPCRchimeraDB entry was annotated to indicate the presence of these conserved residues and motifs. A per residue evolutionary score indicating the conservation of the residues compared to the consensus sequence of their phylogenetic subclass is also included as well as the overall conservation score of the phylogenetic subclass the protein belongs to, calculated using Shannon entropy. The sequence of chimeric entries is compared to their parents' sequence and subclass, with especially of interest the conservation of the chimera sequence in the IC region compared to the EC parent sequence, and vice versa. For both natural and chimeric entries, secondary structure annotations (start and end TMs) were added to their description. For natural GPCRs, these were obtained from precomputed predictions (TMbed[31]), available for all UniProt entries. For the chimeric entries, the TMbed predictor (sequence-based) was run using TMbed's Google Colab notebook (default settings).

To extend the description of the sequence space of the natural entries, we included mutagenesis information. Disease-related mutations were collected from UniProt together with their predicted functional effects. Pharmacological mutations that are known to alter the response of a GPCR to specific ligand were retrieved from GPCRdb. Finally, post-translational modifications (PTMs) from UniProt and Scop3P[32] were added.

Biophysical profile

The predicted biophysical profile includes the protein's backbone and sidechain dynamic as well as the early folding, disorder and beta-sheet aggregation propensity and the conformational propensities (helix, sheet, coil and polyproline II helix). We obtained these features using sequence-based predictors that are part of the b2btools[33]. For the chimera entries, we also computed the difference between their biophysical profile and the one of their parents to enable comparison. To achieve this, we aligned the chimera's and parents' sequence and calculated the difference between the predicted values for each.

Structural description

Based on the PDB ID(s) related to every UniProt ID, 3D experimental structures were collected from the PDB. The GPCR chain was identified based on its annotation on the PDB, and a structure was retained only when its GPCR chain was longer than 200 residues (GPCRs are typically around 400 residues long). When available, the activation state (active, intermediate, inactive) of the 3D structure was also retrieved from GPCRdb. Furthermore, using PDBePISA[34], we identified in the full 3D structure complex, the GPCR residues interacting with a ligand or an effector (G-protein, nanobody/antibody,

or β-arrestin). Furthermore, based on the annotations in the PDB file, disulfide bonds within the GPCRs were determined. Sequence-to-structural position mapping was done with the aid of SIFTS[35]. Finally, the structural information was complemented with AlphaFold2[36] (AF2) models, retrieved from AFDB for the natural GPCRs. For chimeric GPCRs, we computed the AF2 models, as well as the ESMFold[37] (ESM) models, using the default parameters and the VSC Tier-2 general-purpose clusters provided by VUB-HPC. Precomputed human AlphaFold multistate[38] (AFms) models were also added to GPCRchimeraDB for natural human GPCRs and were generated for the chimeras using the VSC Tier-2 infrastructure. AFms is a variant of AF2 without MSA input features where activation state-specific GPCR structures are used as templates, so enabling users to generate state-specific models.

Sequence alignment tool

The sequence alignment tool on GPCRchimeraDB is based on a profile Hidden Markov Model (pHMM). The pHMM is generated on-demand with HMMER[39], based on a pre-computed master alignment created using the structure-informed MSA pipeline SIMSApiper[40]. More details about the alignment procedure and the MA are available in the Supplementary Material.

Accessibility and maintenance

GPCRchimeraDB is a freely accessible database under the FAIR principles[41]. The database is currently maintained and will be updated on a regular basis. We encourage contributions to our database through a dedicated submission page on GPCRchimeraDB (https://www.bio2byte.be/gpcrchimeradb/submitDesign). The database is supported by extensive documentation (https://gpcrchimeradb-docs.readthedocs.io/) and a FAQ (https://www.bio2byte.be/gpcrchimeradb/faqs).

Technical architecture

The GPCRchimeraDB is deployed inside a GNU/Linux web server running Ubuntu 22.04. The backend component written in Python3 is based on the Django framework to build the endpoints and the database middleware. Internally, this component relies on a SQLite3 file-based database. The database layer stores the JSON files as well as the interpretation data in tables. The backend component has a CLI mechanism to import entries allowing maintainers to update and add new entries. The frontend component built on React.js library fetches the data via HTTP requests to the backend in order to render pages and visual components. The frontend codebase is compiled to static HTML, CSS, and JS files, while the backend component runs as an uWSGI socket managed by

a supervisor service. Web server's HTTP requests are handled by a NGINX HTTP proxy instance which redirects GCPRchimeraDB requests to both frontend pages and backend API endpoints.

Results

Chimeric entries: definition

In the context of GPCRchimeraDB, we define a chimeric GPCR as a receptor obtained by fusing two natural GPCRs (the parents). In particular, the IC loops of one GPCR are exchanged, or swapped, with the equivalent regions from another GPCR (Figure 1A-B). As such, the GPCR on the EC side, the EC parent, provides the EC loops and the TM regions while the GPCR on the IC side, the IC parent, supplies the IC loops. The swapped regions must span at least three residues (otherwise the change in sequence is defined as a mutation) and the cutting points or, cutting sites, in the two natural parent GPCRs are typically located at the end of the TM helices/beginning of the IC loops and, at the end of the IC loops/beginning of the next TM helices. Depending on the design strategy, the EC parent can contribute one or more IC loops, i.e., not all IC loops are provided by the IC parent. On the contrary, the IC parent never contributes to the EC side.

The most common chimeric GPCR designs in literature can be grouped in three types (Figure S1):

- Type 1 designs, also called Opsin-based or OptoXRs designs, combine a light-sensitive GPCR, often Rhodopsin, at the EC side of the chimera with a poorly characterized GPCR at the IC side. Hence, through light stimulation, the designed chimera is activated, enabling the study of the biological pathways associated with the IC parent, which could not be activated previously due to the absence of a known ligand (orphan GPCR). This design strategy also enables identification of the importance of specific regions in IC parent, for example for G-protein selectivity.
- Type 2 designs have a similar configuration as Type 1 chimeras, but the EC parent is a non-sensory GPCR, i.e., the chimera gets activated by a known endogenous or synthetic ligand instead of light. This again enables the study of the (activated) IC side GPCR or its related signaling pathway(s).
- In contrast, Type 3 contains the well characterized GPCR on the IC side of the chimera. For these IC parents, a stabilizing effector (nanobody/VHH) that can lock the GPCR in a specific conformation is known. This effector can so stabilize the chimera in, for example, a disease related conformation and drug screening for this disease state can then be performed to help deorphanize GPCRs. By stabilizing the chimeric GPCR in specific conformations, structure

elucidation can also be facilitated which can, in turn, aid structure-based drug design or bring new structural insights.

To summarize, the design strategy of the chimera depends on the research goal: elucidating GPCR signaling pathways or unravelling key regions for the protein function (Type 1 and 2), identifying new potential ligands (Type 3), or supporting 3D structure determination (Type 3).

Overview of previously designed chimeras

We classified the 170 chimeras on GPCRchimeraDB according to their design type as described above. As shown in Table 1 and Figure S2A, 94 of these designs are rhodopsin-based chimeras (Type 1) and were designed based on the pioneering work of Kim et al. in 2005[17]. The overall extensive use of rhodopsin underscores its utility as a reliable framework in the design of GPCR chimeras. In contrast, only a few Type 3 chimeras have been successfully constructed (14 in total). They all consist of the human κ opioid receptor (KOR) at the IC side, stabilized by a nanobody (Nb6), and combined with a GPCR of interest at the EC side[7]. A commercial patent[42] has been filed for the development of Type 3 chimeras which highlights their significant therapeutic and economic potential. Type 2 designs are referenced in the largest number of distinct publications (13 in total) and display the broadest applicability, ranging for example from understanding the role of the C terminal in functional desensitization[10] to exploring G-protein selectivity[9] and comparing the pharmacology of GPCR homologs[11]. Notably, DREADD-based GPCR chimeras[43] also fall under Type 2 designs. In these chimeras, the EC parent is a DREADD, a GPCR mutated to respond exclusively to a synthetic ligand with minimal biological activity[44], while the IC parent is the GPCR of interest. This configuration overcomes challenges such as off-target effects and the poor bioavailability of natural ligands, which often hinder studies of natural and some chimeric designs. Despite their significance, the total number of Type 2 chimeras remains relatively low, with only 62 designs reported, of which 29 feature unique parent combinations. This limited number likely reflects the inherent difficulties in producing constructs that are both well-expressed and functionally robust.

Table 1: *Data integrated into GPCRchimeraDB. On the left, the type of entries on the database and on the right, the type and average number of features available to describe the entries. Note that the natural GPCR dataset is relatively unbalanced with some GPCRs (rhodopsin, adrenergic receptors, …) or organisms (human, mouse, …) that are highly annotated compared to other GPCRs (e.g. orphan GPCRs).*

Data type	Total	Data type	Average per Natural GPCRs	Average per Chimeric GPCRs
Entries	1,928	*Features per entry*	226.5	16.5
Natural entries	1,758	Disease-related mutations	187.7	0.0
Human entries	712	Pharmacological mutations	17.1	0.0
Non-human entries	1,046	Conserved motifs and x.50 residues	8.4	8.9
Chimeric entries	170	PTMs	3.3	0.0
Type 1 chimeras	94	Disulfide bonds	0.1	0.0
Type 2 chimeras	62	Experimental 3D structures	0.6	0.1
Type 3 chimeras	14	Predicted 3D models	1.8	4.0
Unique EC parents	39	Activation states	1.3	2.1
Unique IC parents	94	Interacting residues EC side	2.5	0.3
		Interacting residues IC side	3.9	0.7
		Endogenous ligands	0.5	0.0
		G-protein and β-arrestin	0.1	0.5

Both successful and unsuccessful chimera designs are included in GPCRchimeraDB. While all chimeras reported in the literature are confirmed to express, some exhibit reduced functionality or are entirely non-functional compared to their parents. On GPCRchimeraDB, each chimera's functionality is described using either the quantitative or qualitative value provided in the publication describing them. However, in case users want to compare functionality values across different chimeras, we strongly advise them to refer to the original publications and compare the experimental conditions. This is because the assays used to assess functionality can vary (e.g., real-time versus endpoint assays), as do experimental conditions (e.g., reagent concentrations) and the way results are presented (relative versus absolute values). For this reason, it is impossible to standardize this information, and we do not report the total number of successful chimeras on GPCRchimeraDB.

Extensive description of chimeric and natural GPCRs

In total, 170 chimeric and 1,758 natural class A GPCRs were uploaded on GPCRchimeraDB (Figure S2B). For natural GPCRs entries, the UniProt ID allowed the retrieval of experimental and precomputed results from various databases (UniProt, GPCRdb, PDB, AFDB, Scop3P and, precomputed TMbed annotations and AFms models), which were then supplemented with predicted data (b2btools, Figure

1). On the contrary, the data for chimeric entries was exclusively extracted from the paper in which they are described and complemented with predicted data (b2btools, TMbed and, AF2, AFms and ESMF models), limiting their annotations compared to natural GPCRs. Nevertheless, GPCRchimeraDB offers a broad description of both the chimeric and natural entries (Table 1).

For every entry, a brief overview (classification, binding partners, …) is presented next to an interconnected interactive sequence-structure (1D-3D) viewer (Figure 2A-C). The 1D-3D viewer enables studying the location of pharmacological and disease-related mutations, conserved motifs, interacting residues or PTMs, directly on a user-selected 3D structure. This 3D structure can be either experimentally resolved (from the PDB, when available) or predicted (AF2, AFms and ESMF). Also, when available, the user can also choose the 3D structure based on its activation state (active or inactive, defined using GPCRdb's definition or by AFms). For each entry, an AF2 model was either retrieved or generated. However, AFms models are only available for chimeric and human natural GPCRs. Studies indicate that AFms models, particularly in the active state, are more accurate than the standard AF2 models. This is because AF2 is biased toward predicting the inactive state, likely due to an unbalanced training and template dataset. Consequently, computing the missing AFms models for non-human natural GPCRs will be considered in future work. For chimeric entries, their ESMF model is also provided. ESMF generates 3D models directly from single sequences, leveraging a language model[37]. In contrast, AF2 requires evolutionary information from a MSA and structural templates[36]. Therefore, the AF2 model of chimeric GPCRs might get biased by the evolutionary and structural information from the EC parent as it mainly contributes to the chimera's sequence compared to the IC parent. We leave it to the user to choose the most suitable model (Figure S3A). AF2 single-sequence mode (default parameters with and without relaxation on ColabFold[45]) was also tested. While it performs well on *de novo* designs[46], it produces highly inaccurate models for chimeric GPCRs (Figure S3B).

In addition, the biophysical profile of all GPCRs is provided. It represents emergent features that evolution conserves to maintain protein function, such as backbone dynamics or conformational propensities, that are not directly apparent from protein sequence or structure (see the Biophysical profile section for more details). This type of information can help detect global homology between proteins[47], as well as improve the pairwise alignment of proteins that are evolutionary distant[48], and is therefore particularly relevant for very low shared sequence identity protein families like GPCRs. These biophysical profiles can be directly mapped on a selected 3D structure to facilitate interpretation (Figure 2B) and are unique to GPCRchimeraDB.

For the chimeric entries, the EC and IC parents, design type, functionality value (qualitative or quantitative), application, and the location of the cutting sites are specified on their entry page. A schematic representation of the chimera is also generated (Figure 2E), and the location of the cutting sites can be visualized directly on the 3D structure (Figure 2F). Additionally, their biophysical profile and sequence can be compared to those of their parents (Figure 2F-G) and a table listing other designs with at least one parent in common is available. Finally, for the natural entries involved in chimeric designs, all the previously used cutting points are highlighted (Figure 2D) and a table listing all the chimeric designs they are involved in is provided.

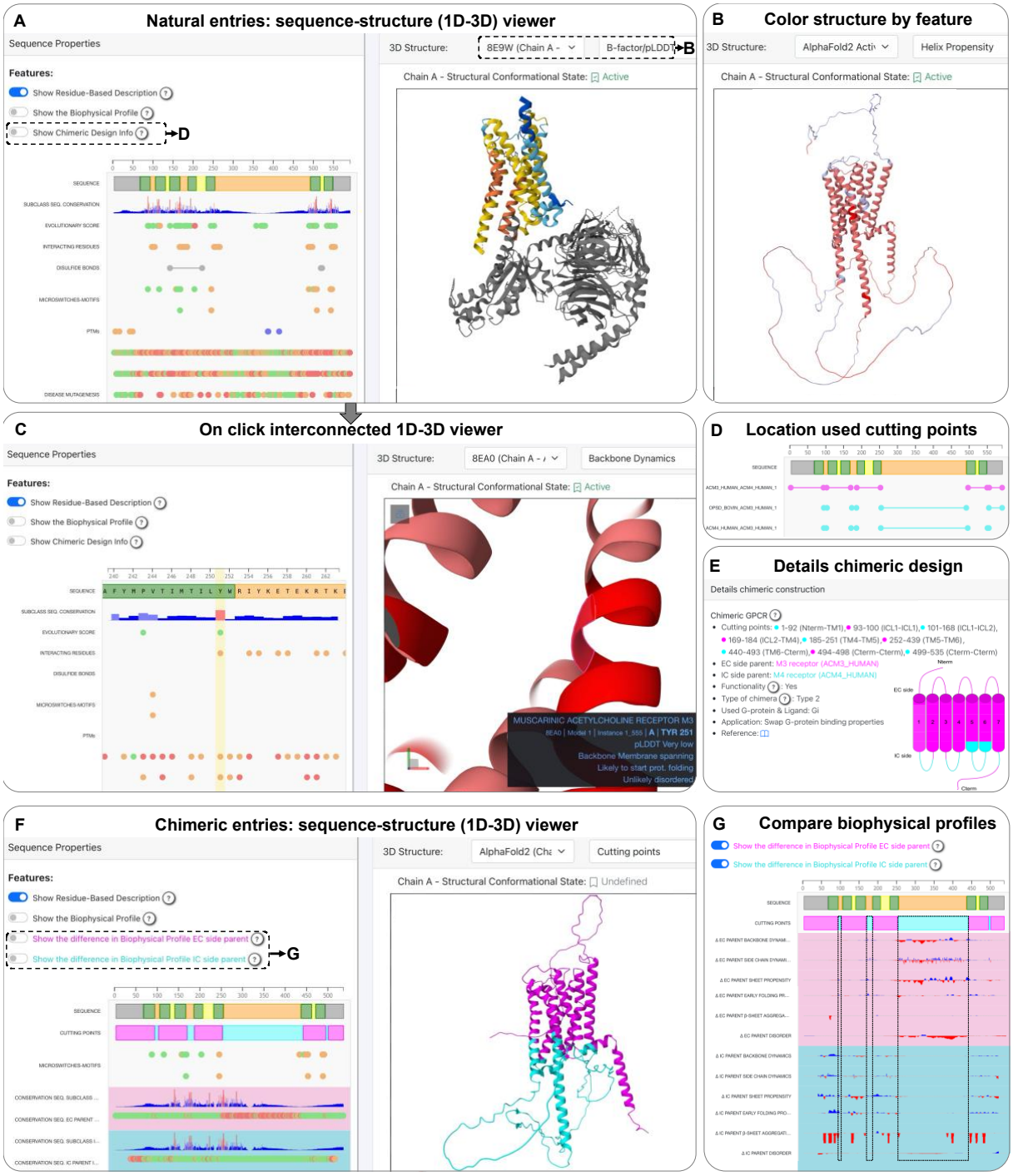


Figure 2: *Overview of GPCRchimeraDB's interactive visualization features.* ***A.*** *Default view of the interconnected sequence-structure viewer on each natural entry page.* ***B.*** *A 3D structure can be selected from a list of experimental and predicted models, with color schemes based on B-factor/pLDDT or biophysical features.* ***C.*** *Clicking a residue in the sequence or structure viewer zooms in for detailed analysis.* ***D.*** *Location of other cutting sites in designed chimeras involving the selected natural entry.* ***E.*** *Detailed description of the chimera on its entry page.* ***F.*** *Default sequence-structure viewer for each chimera entry page.* ***G.*** *Feature to compare the biophysical profile of the chimera with its parents.*

Leveraging GPCRchimeraDB: How to design a novel chimeric GPCR?

In this section, we describe how to utilize GPCRchimeraDB to guide the design of a novel chimera involving GPCRx, the GPCR under investigation. The design process begins by determining the appropriate design type (Type 1, 2 or 3, see 'Chimeric entries: definition'). This dictates whether GPCRx will serve as the EC or IC parent of the chimera:

- Type 1 or 2: For studying the functionality or signaling pathway of GPCRx. GPCRx will be the IC parent.
- Type 3: For uncovering the 3D structure or conducting ligand screening of GPCRx. GPCRx will be the EC parent.

The next step involves selecting GPCRy, the second parent. GPCRy should be a well-characterized receptor that complements the lack of knowledge for GPCRx. Depending on the design type, it is important to choose a GPCRy for which either a ligand (Type 1 or 2) or a stabilizing effector is known (Type 3). Finally, the cutting sites in both GPCRx and GPCRy need to be determined to ensure that the resulting chimera can get activated and successfully transmit the signal from the outside to the inside of the cell. GPCRchimeraDB can come into play to facilitate every stage of this design process, as outlined in the following steps:

1. **Explore the entry page of GPCRx:** On every entry page, GPCRchimeraDB integrates annotations of major databases in one location, so providing an overview for further analysis. By reviewing known mutations and their impact, the evolutionary score of the protein compared to its subclass, and the conservation of its motifs, a comprehensive picture of crucial residues emerges, so helping to define the sequence space of GPCRx and identifying regions sensitive to mutations. This is critical information, as chimeric design can be seen as introducing stretches of mutations in specific regions. Additional information, such as PTM sites and interacting residues, offer insights on why certain residues might be essential for the protein's function. The integrated sequence-structure (1D-3D) viewer enables direct examination of these key residues in their 3D context.
2. **Define potential GPCRy candidates:** The next step is to identify a suitable GPCRy to work with the GPCR under study (GPCRx). For Type 1 and Type 3 chimeras, GPCRy is usually respectively Rhodopsin and the κ-Opioid receptor. However, for Type 2 chimeras, many candidates may be appropriate. The following guidelines can help select the GPCRy:
    - Chimeras typically involve combining a well-characterized GPCR with one that is less understood. Therefore, GPCRy should not be an orphan GPCR (this can be established using the ligand-based subclass filter), nor a receptor with limited information on its entry page.

- Since preserving key functional connections in the chimera is crucial, designing chimeras between related GPCRs is in principle more likely to succeed than between distant GPCRs. The filtering system on the GPCRchimeraDB entries list page can be used to identify GPCRs from the same subclass as GPCRx.
- Depending on the research objective, it may be important to select a GPCRy that either binds the same or a different G-protein compared to GPCRx. The integrated filtering system can also assist in identifying GPCRs based on their G-protein coupling profile.
- GPCRy candidates can also be identified by reviewing past chimeric designs involving GPCRx (see bottom of entry page GPCRx). Related GPCRs of past parents could also be considered.
- GPCRy should be a receptor that is easy to work with in the lab. For instance, opioid related GPCRs may pose practical challenges due to regulatory or experimental constraints.

Once suitable GPCRy candidates have been identified, their entry pages should be explored in the same manner as outlined in Step 1 to gain insights into the key residues for GPCRy's function.

3. **Respect the key rules for determining cutting sites:** The cutting sites for GPCRx and GPCRy chimeras need to be carefully determined. While no standardized procedure exists, based on the studied literature we recommend the following guidelines to enhance the success rate of chimera designs:
   - Preserve the microswitches essential for the function of the EC parent within the chimera's sequence.
   - Conserve the secondary structure from the parent GPCRs within the chimera by introducing the cutting sites within the parents' regions with the same secondary structure.
   - Favor the EC parent: the more closely the chimeric sequence resembles the EC parent sequence, the better. For example, studies have shown that exchanging IC loop 1 is not critical for successful chimera design[11,49].
   - Ensure that the IC parent contributes its residues responsible for the binding of the G-protein of interest to the chimeric sequence[18].

   GPCRchimeraDB supports these rules by providing information on conserved motifs, secondary structures, and interacting residues.

4. **Improve the cutting point selection by analyzing related chimeras:** Although the above rules are essential for successful chimera design, they may not be sufficient. We suggest reviewing previously designed chimeras involving GPCRx, GPCRy, or related GPCRs, or chimeras designed for the same research purposes. GPCRchimeraDB offers the following assisting features:
    - Compare biophysical profiles of previously designed chimeras with their parent GPCRs using the 1D-3D viewer (Figure 2F-G). Does the biophysical profile of the chimera match with the profile of the parents?
    - Analyze residue conservation in exchanged regions using the 'Conservation parent vs chimera' track or GPCRchimeraDB's alignment tool. How much do EC parent residues need to be conserved in the provided IC parent region for a chimera to be successful? Is conserving the motifs enough? GPCRchimeraDB also displays the parent subclass conservation: are residues that are conserved throughout the protein subclass required in the chimeric sequence?
    - Review the location of the previous cutting sites in the parents and investigate how far they extend into the TM helices.
    - Download and align the 3D structures of the parents to identify equivalent regions suitable for positioning the cutting sites.
5. **Design the chimera:** Once cutting sites have been defined, align GPCRx and GPCRy using the GPCRchimeraDB alignment tool and design your chimera.

To increase the likelihood of success, *in silico* validation is recommended. For example, the AF2 or ESMF model of the designed chimera can be generated and superimposed with the 3D structure of GPCRx and GPCRy to study how much they differ. The model's confidence metrics (pLDDT and PAE) could also be compared with those of the parent models. Finally, we also suggest computing the biophysical profile of the designed chimera with the b2btools to ensure its properties align with natural GPCR behavior. Molecular Dynamics (MD) simulations could also be considered. The most promising designs can then be tested experimentally before pursuing the research objectives that motivated the chimera's creation.

Discussion and perspectives

We present GPCRchimeraDB, a comprehensive database dedicated to both natural and chimeric GPCRs. As the first centralized repository for chimeric GPCRs, it offers extensive, standardized descriptions at the sequence, structural, and biophysical level, complemented by visualization tools and links to their natural, parent GPCRs. This standardization facilitates comparisons between

chimeric designs and the parent receptors, providing insights into the requirements for successful designs. To support this, GPCRchimeraDB introduces a naming convention and a classification system based on the type of design (Type 1, 2 or 3). Furthermore, expanding on resources like GPCRdb and UniProt, GPCRchimeraDB enriches the description of natural GPCRs by including their biophysical profile and integrating an interactive sequence-structure visualization tool. By covering all natural GPCRs, it also supports the exploration of novel chimeric designs through new parent combinations. Additionally, the database incorporates a sequence alignment tool, a feature comparator, and a step-by-step workflow to guide users through the chimera design process, making it a valuable resource for both analyzing past designs and developing new ones. Finally, assuming that chimeric GPCRs are functional only if they retain key sequence, structural, and biophysical properties of their parent GPCRs, analyzing previous chimeric designs also brings insights into the key regions of natural parent GPCRs and helps understanding their function.

Given the vast amount of information within GPCRchimeraDB, an AI tool could further streamline the design process. Such a tool could assist researchers by highlighting relevant data and suggesting potential constructs based on their GPCR of interest and research goals. However, to ensure the success of this AI-driven approach, it will be essential to standardize the functionality values of the chimeras or, to provide complete experimental contexts next to each functionality value in order to create a robust dataset for AI training and testing. High-throughput studies under uniform conditions would help to obtain such a dataset.

The strength of GPCRchimeraDB lies in its extensive overview of the GPCR chimera field, providing key insights to guide the design of novel chimeras. We believe this resource will promote a more systematic and informed approach to chimera development, ultimately increasing the success rate of these designs, lowering experimental costs, and deepening our understanding of GPCRs.

Data availability

The entries and their annotations are directly available on the database (https://www.bio2byte.be/gpcrchimeradb/) or on Zenodo (https://zenodo.org/records/14282679). The MA is also available on Zenodo. The Python scripts to retrieve the annotations to describe the entries as well as to format a chimera's data based on a sequence are accessible on GitHub (https://github.com/Bio2Byte/GPCRchimeraDB).

Conflict of interest

None declared.

Author contributions

Charlotte Crauwels: Data Curation, Formal analysis, Software, Validation, Writing - Original Draft, Funding acquisition. **Adrián Díaz:** Software, Validation. **Wim Vranken:** Conceptualization, Validation, Writing - Review & Editing, Supervision, Funding acquisition.

Acknowledgments

We thank the Bio2Byte research group for their testing and feedback to help us improve GPCRchimeraDB. We also highly appreciated the feedback provided by Prof. Julien Hanson and his group, as well as by Dr. Sarah Triest. Finally, we would like to express our gratitude to Dr. Zara Sands for her ideas, feedback, and enthusiasm during the early stages of development of this database.

Funding

This work was supported by the Research Foundation Flanders (FWO) SB PhD fellowship [1SE5923N to C.C.]; and the FWO International Research Infrastructure [I000323N to W.V. and G.0328.16N to A.D.]. The resources used in this work were provided in part by the VSC (Flemish Supercomputer Center), funded by the FWO and the Flemish Government.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI) in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.sh

References

(1) Zhang, M.; Chen, T.; Lu, X.; Lan, X.; Chen, Z.; Lu, S. G Protein-Coupled Receptors (GPCRs): Advances in Structures, Mechanisms and Drug Discovery. *Sig Transduct Target Ther* **2024**, *9* (1), 1–43. https://doi.org/10.1038/s41392-024-01803-6.

(2) Fredriksson, R.; Lagerström, M. C.; Lundin, L.-G.; Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* **2003**, *63* (6), 1256–1272. https://doi.org/10.1124/mol.63.6.1256.

(3) 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* **2017**, *16* (12), 829–842. https://doi.org/10.1038/nrd.2017.178.

(4) Schöneberg, T.; Schulz, A.; Biebermann, H.; Hermsdorf, T.; Römpler, H.; Sangkuhl, K. Mutant G-Protein-Coupled Receptors as a Cause of Human Diseases. *Pharmacol Ther* **2004**, *104* (3), 173–206. https://doi.org/10.1016/j.pharmthera.2004.08.008.
(5) Addis, P.; Bali, U.; Baron, F.; Campbell, A.; Harborne, S.; Jagger, L.; Milne, G.; Pearce, M.; Rosethorne, E. M.; Satchell, R.; Swift, D.; Young, B.; Unitt, J. F. Key Aspects of Modern GPCR Drug Discovery. *SLAS Discovery* **2024**, *29* (1), 1–22. https://doi.org/10.1016/j.slasd.2023.08.007.
(6) Keri, D.; Barth, P. Reprogramming G Protein Coupled Receptor Structure and Function. *Curr Opin Struct Biol* **2018**, *51*, 187–194. https://doi.org/10.1016/j.sbi.2018.07.008.
(7) Laeremans, T.; Sands, Z. A.; Claes, P.; De Blieck, A.; De Cesco, S.; Triest, S.; Busch, A.; Felix, D.; Kumar, A.; Jaakola, V.-P.; Menet, C. Accelerating GPCR Drug Discovery With Conformation-Stabilizing VHHs. *Front Mol Biosci* **2022**, *9*, 863099. https://doi.org/10.3389/fmolb.2022.863099.
(8) Yin, D.; Gavi, S.; Wang, H.; Malbon, C. C. Probing Receptor Structure/Function with Chimeric G-Protein-Coupled Receptors. *Mol Pharmacol* **2004**, *65* (6), 1323–1332. https://doi.org/10.1124/mol.65.6.1323.
(9) Masuho, I.; Kise, R.; Gainza, P.; Von Moo, E.; Li, X.; Tany, R.; Wakasugi-Masuho, H.; Correia, B. E.; Martemyanov, K. A. Rules and Mechanisms Governing G Protein Coupling Selectivity of GPCRs. *Cell Rep* **2023**, *42* (10), 113173. https://doi.org/10.1016/j.celrep.2023.113173.
(10) Liggett, S. B.; Freedman, N. J.; Schwinn, D. A.; Lefkowitz, R. J. Structural Basis for Receptor Subtype-Specific Regulation Revealed by a Chimeric Beta 3/Beta 2-Adrenergic Receptor. *Proceedings of the National Academy of Sciences* **1993**, *90* (8), 3665–3669. https://doi.org/10.1073/pnas.90.8.3665.
(11) Ilani, T.; Fishburn, C. S.; Levavi-Sivan, B.; Carmon, S.; Raveh, L.; Fuchs, S. Coupling of Dopamine Receptors to G Proteins: Studies with Chimeric D2/D3 Dopamine Receptors. *Cell Mol Neurobiol* **2002**, *22* (1), 47–56. https://doi.org/10.1023/A:1015341712166.
(12) Wang, L.-D.; Gantz, I.; Butler, K.; Hoeltzel, M.; Del Valle, J. Histamine H2 Receptor Mediated Dual Signaling: Mapping of Structural Requirements Using B2 Adrenergic Chimeric Receptors. *Biochemical and Biophysical Research Communications* **2000**, *276* (2), 539–545. https://doi.org/10.1006/bbrc.2000.3507.
(13) Toyoda, Y.; Zhu, A.; Kong, F.; Shan, S.; Zhao, J.; Wang, N.; Sun, X.; Zhang, L.; Yan, C.; Kobilka, B. K.; Liu, X. Structural Basis of α1A-Adrenergic Receptor Activation and Recognition by an Extracellular Nanobody. *Nat Commun* **2023**, *14* (1), 3655. https://doi.org/10.1038/s41467-023-39310-x.
(14) Robertson, M. J.; Papasergi-Scott, M. M.; He, F.; Seven, A. B.; Meyerowitz, J. G.; Panova, O.; Peroto, M. C.; Che, T.; Skiniotis, G. Structure Determination of Inactive-State GPCRs with a Universal Nanobody. *Nat Struct Mol Biol* **2022**, *29* (12), 1188–1195. https://doi.org/10.1038/s41594-022-00859-8.
(15) Fontaine, T.; Busch, A.; Laeremans, T.; De Cesco, S.; Liang, Y.-L.; Jaakola, V.-P.; Sands, Z.; Triest, S.; Masiulis, S.; Dekeyzer, L.; Samyn, N.; Loeys, N.; Perneel, L.; Debaere, M.; Martini, M.; Vantieghem, C.; Virmani, R.; Skieterska, K.; Staelens, S.; Barroco, R.; Van Roy, M.; Menet, C. Structure Elucidation of a Human Melanocortin-4 Receptor Specific Orthosteric Nanobody Agonist. *Nat Commun* **2024**, *15* (1), 7029. https://doi.org/10.1038/s41467-024-50827-7.
(16) Fauser, J.; Leschinsky, N.; Szynal, B. N.; Karginov, A. V. Engineered Allosteric Regulation of Protein Function. *Journal of Molecular Biology* **2022**, *434* (17), 167620. https://doi.org/10.1016/j.jmb.2022.167620.
(17) Kim, J.-M.; Hwa, J.; Garriga, P.; Reeves, P. J.; RajBhandary, U. L.; Khorana, H. G. Light-Driven Activation of B2-Adrenergic Receptor Signaling by a Chimeric Rhodopsin Containing the B2-Adrenergic Receptor Cytoplasmic Loops. *Biochemistry* **2005**, *44* (7), 2284–2292. https://doi.org/10.1021/bi048328i.
(18) Tichy, A.-M.; So, W. L.; Gerrard, E. J.; Janovjak, H. Structure-Guided Optimization of Light-Activated Chimeric G-Protein-Coupled Receptors. *Structure* **2022**, *30* (8), 1075-1087.e4. https://doi.org/10.1016/j.str.2022.04.012.

(19) Weis, W. I.; Kobilka, B. K. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu Rev Biochem* **2018**, *87*, 897–919. https://doi.org/10.1146/annurev-biochem-060614-033910.
(20) Herrera, L. P. T.; Andreassen, S. N.; Caroli, J.; Rodríguez-Espigares, I.; Kermani, A. A.; Keserű, G. M.; Kooistra, A. J.; Pándy-Szekeres, G.; Gloriam, D. E. GPCRdb in 2025: Adding Odorant Receptors, Data Mapper, Structure Similarity Search and Models of Physiological Ligand Complexes. *Nucleic Acids Research* **2024**, gkae1065. https://doi.org/10.1093/nar/gkae1065.
(21) Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B. L.; Salazar, G. A.; Bileschi, M. L.; Bork, P.; Bridge, A.; Colwell, L.; Gough, J.; Haft, D. H.; Letunić, I.; Marchler-Bauer, A.; Mi, H.; Natale, D. A.; Orengo, C. A.; Pandurangan, A. P.; Rivoire, C.; Sigrist, C. J. A.; Sillitoe, I.; Thanki, N.; Thomas, P. D.; Tosatto, S. C. E.; Wu, C. H.; Bateman, A. InterPro in 2022. *Nucleic Acids Research* **2023**, *51* (D1), D418–D427. https://doi.org/10.1093/nar/gkac993.
(22) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Research* **2022**, *50* (D1), D439–D444. https://doi.org/10.1093/nar/gkab1061.
(23) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(24) Estévez Soto, P. R.; Vehstedt, E.; Fischer, J.; Kumar, M.; Calderón, A.; Koleva, M.; Zimmer, F. DNA2protein Python Script, 2015. https://doi.org/10.5281/zenodo.34023.
(25) The UniProt Consortium; Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Adesina, A.; Ahmad, S.; Bowler-Barnett, E. H.; Bye-A-Jee, H.; Carpentier, D.; Denny, P.; Fan, J.; Garmiri, P.; Gonzales, L. J. D. C.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Joshi, V.; Jyothi, D.; Kandasaamy, S.; Lock, A.; Luciani, A.; Luo, J.; Lussi, Y.; Marin, J. S. M.; Raposo, P.; Rice, D. L.; Santos, R.; Speretta, E.; Stephenson, J.; Totoo, P.; Tyagi, N.; Urakova, N.; Vasudev, P.; Warner, K.; Wijerathne, S.; Yu, C. W.-H.; Zaru, R.; Bridge, A. J.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A. H.; Axelsen, K. B.; Bansal, P.; Baratin, D.; Batista Neto, T. M.; Blatter, M.-C.; Bolleman, J. T.; Boutet, E.; Breuza, L.; Gil, B. C.; Casals-Casas, C.; Echioukh, K. C.; Coudert, E.; Cuche, B.; De Castro, E.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gaudet, P.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz, N.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Kerhornou, A.; Mercier, P. L.; Lieberherr, D.; Masson, P.; Morgat, A.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Poux, S.; Pozzato, M.; Pruess, M.; Redaschi, N.; Rivoire, C.; Sigrist, C. J. A.; Sonesson, K.; Sundaram, S.; Sveshnikova, A.; Wu, C. H.; Arighi, C. N.; Chen, C.; Chen, Y.; Huang, H.; Laiho, K.; Lehvaslaiho, M.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Y.; Zhang, J. UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Research* **2024**, gkae1010. https://doi.org/10.1093/nar/gkae1010.
(26) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10* (1), 421. https://doi.org/10.1186/1471-2105-10-421.
(27) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25* (11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.
(28) Cvicek, V.; Goddard, W. A.; Abrol, R. Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLoS Comput Biol* **2016**, *12* (3), e1004805. https://doi.org/10.1371/journal.pcbi.1004805.

(29) Nguyen, L.-T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **2015**, *32* (1), 268–274. https://doi.org/10.1093/molbev/msu300.
(30) Ballesteros, J. A.; Weinstein, H. [19] Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors; 1995; pp 366–428. https://doi.org/10.1016/S1043-9471(05)80049-7.
(31) Bernhofer, M.; Rost, B. TMbed: Transmembrane Proteins Predicted through Language Model Embeddings. *BMC Bioinformatics* **2022**, *23* (1), 326. https://doi.org/10.1186/s12859-022-04873-x.
(32) Ramasamy, P.; Turan, D.; Tichshenko, N.; Hulstaert, N.; Vandermarliere, E.; Vranken, W.; Martens, L. Scop3P: A Comprehensive Resource of Human Phosphosites within Their Full Context. *J. Proteome Res.* **2020**, *19* (8), 3478–3486. https://doi.org/10.1021/acs.jproteome.0c00306.
(33) Kagami, L. P.; Orlando, G.; Raimondi, D.; Ancien, F.; Dixit, B.; Gavaldá-García, J.; Ramasamy, P.; Roca-Martínez, J.; Tzavella, K.; Vranken, W. b2bTools: Online Predictions for Protein Biophysical Features and Their Conservation. *Nucleic Acids Research* **2021**, *49* (W1), W52–W59. https://doi.org/10.1093/nar/gkab425.
(34) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* **2007**, *372* (3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.
(35) Dana, J. M.; Gutmanas, A.; Tyagi, N.; Qi, G.; O'Donovan, C.; Martin, M.; Velankar, S. SIFTS: Updated Structure Integration with Function, Taxonomy and Sequences Resource Allows 40-Fold Increase in Coverage of Structure-Based Annotations for Proteins. *Nucleic Acids Research* **2019**, *47* (D1), D482–D489. https://doi.org/10.1093/nar/gky1114.
(36) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(37) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379* (6637), 1123–1130. https://doi.org/10.1126/science.ade2574.
(38) Heo, L.; Feig, M. Multi-State Modeling of G-Protein Coupled Receptors at Experimental Accuracy. *Proteins: Structure, Function, and Bioinformatics* **2022**, *90* (11), 1873–1885. https://doi.org/10.1002/prot.26382.
(39) Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **2011**, *7* (10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
(40) Crauwels, C.; Heidig, S.-L.; Díaz, A.; Vranken, W. F. Large-Scale Structure-Informed Multiple Sequence Alignment of Proteins with SIMSApiper. *Bioinformatics* **2024**, btae276. https://doi.org/10.1093/bioinformatics/btae276.
(41) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* **2016**, *3* (1), 160018. https://doi.org/10.1038/sdata.2016.18.

(42) Blieck, A. D.; CLAES, P.; VERVERKEN, C.; MENET, C.; DEKEYZER, L. Chimeric Proteins and Methods to Screen for Compounds and Ligands Binding to Gpcrs. WO2020221768A1, November 5, 2020. https://patents.google.com/patent/WO2020221768A1/en (accessed 2022-02-25).
(43) Schulz, R.; Korkut-Demirbaş, M.; Venturino, A.; Colombo, G.; Siegert, S. Chimeric GPCRs Mimic Distinct Signaling Pathways and Modulate Microglia Responses. *Nat Commun* **2022**, *13* (1), 4728. https://doi.org/10.1038/s41467-022-32390-1.
(44) Urban, D. J.; Roth, B. L. DREADDs (Designer Receptors Exclusively Activated by Designer Drugs): Chemogenetic Tools with Therapeutic Utility. *Annual Review of Pharmacology and Toxicology* **2015**, *55* (Volume 55, 2015), 399–417. https://doi.org/10.1146/annurev-pharmtox-010814-124803.
(45) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat Methods* **2022**, *19* (6), 679–682. https://doi.org/10.1038/s41592-022-01488-1.
(46) Goverde, C. A.; Wolf, B.; Khakzad, H.; Rosset, S.; Correia, B. E. De Novo Protein Design by Inversion of the AlphaFold Structure Prediction Network. *Protein Sci* **2023**, *32* (6), e4653. https://doi.org/10.1002/pro.4653.
(47) Raimondi, D.; Orlando, G.; Moreau, Y.; Vranken, W. F. Ultra-Fast Global Homology Detection with Discrete Cosine Transform and Dynamic Time Warping. *Bioinformatics* **2018**, *34* (18), 3118–3125. https://doi.org/10.1093/bioinformatics/bty309.
(48) Orlando, G.; Raimondi, D.; Khan, T.; Lenaerts, T.; Vranken, W. F. SVM-Dependent Pairwise HMM: An Application to Protein Pairwise Alignments. *Bioinformatics* **2017**, *33* (24), 3902–3908. https://doi.org/10.1093/bioinformatics/btx391.
(49) Hickey, D. G.; Davies, W. I. L.; Hughes, S.; Rodgers, J.; Thavanesan, N.; MacLaren, R. E.; Hankins, M. W. Chimeric Human Opsins as Optogenetic Light Sensitisers. *Journal of Experimental Biology* **2021**, *224* (14), jeb240580. https://doi.org/10.1242/jeb.240580.