Systematic analyses of the sequence conservation and ligand interaction patterns of purinergic P1 and P2Y receptors provide a structural basis for receptor selectivity

Ri Han, Hongryul Yoon, Jiho Yoo *, Yoonji Lee *

College of Pharmacy, Chung-Ang University, Seoul 06974, Republic of Korea

ARTICLE INFO

ABSTRACT

Purinergic receptors are membrane proteins that regulate numerous cellular functions by catalyzing reactions involving purine nucleotides or nucleosides. Among the three receptor families, i.e., P1, P2X, and P2Y, the P1 and P2Y receptors share common structural features of class A GPCR. Comprehensive sequence and structural analysis revealed that the P1 and P2Y receptors belong to two distinct groups. They exhibit different ligand-binding site features that can distinguish between specific activators. These specific amino acid residues in the binding cavity may be involved in the selectivity and unique pharmacological behavior of each subtype. In this study, we conducted a structure-based analysis of purinergic P1 and P2Y receptors to identify their evolutionary signature and obtain structural insights into ligand recognition and selectivity. The structural features of the P1 and P2Y receptor classes were compared based on sequence conservation and ligand interaction patterns. Orthologous protein sequences were collected for the P1 and P2Y receptors, and sequence conservation was calculated based on Shannon entropy to identify highly conserved residues. To analyze the ligand interaction patterns, we performed docking studies on the P1 and P2Y receptors using known ligand information extracted from the ChEMBL database. We analyzed how the conserved residues are related to ligand-binding sites and how the key interacting residues differ between P1 and P2Y receptors, or between agonists and antagonists. We extracted new similarities and differences between the receptor subtypes, and the results can be used for designing new ligands by predicting hotspot residues that are important for functional selectivity.

1. Introduction

Purinergic receptors, also referred to as purinoceptors, are membrane proteins that regulate numerous cellular functions by catalyzing reactions involving purine nucleotides or nucleosides [1]. They mediate relaxation of smooth muscles in response to the release of adenosine (P1 receptors) or ATP (P2 receptors). As one of the most widespread signaling systems in the body, the purinergic system exerts a great deal of influence on a wide range of tissues and cells, mediating numerous physiological reactions and contributing to various pathological responses [2,3]. For example, P1 receptor agonists are effective for treating supraventricular tachycardia, and A_{2A} receptor antagonists are promising for the treatment of Parkinson's disease [4]. The $P2Y_{12}$ antagonist Clopidogrel is used to treat thrombosis and stroke because it inhibits the activation of $P2Y_{12}$ receptors that are involved in platelet aggregation [5]. Several other studies on the use of purinergic agents to treat osteoporosis, myocardial infarction, epilepsy, atherosclerosis, depression, autism, diabetes, cancer, and other diseases are currently underway [3].

Purinergic receptors can be classified into three families, P1 (adenosine), P2X, and P2Y, based on endogenous ligands, molecular structures, and tissue distributions [1]. P1 receptors (A_1, A_{2A}, A_{2B}, and A_3) are stimulated by adenosine, whereas P2 receptors ($P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$, $P2Y_{11}$, $P2Y_{12}$, $P2Y_{13}$, and $P2Y_{14}$) respond to various purines and pyrimidine mono/dinucleotides (ATP, ADP, UTP, and UDP). Purinergic signaling can also be modulated by interactions between P1 and P2 receptors and by the formation of heterodimeric receptors A_{2A} and P2Y. All these receptors bind to purine-based substances; however, they have a long evolutionary distance, especially ATP-gated ionotropic P2X receptors, which are ligand-gated ion channels, whereas P1 and P2Y receptors belong to the G protein-

* Corresponding authors.
E-mail addresses: jyoo@cau.ac.kr (J. Yoo), yoonjilee@cau.ac.kr (Y. Lee).

coupled receptor (GPCR) family. As one of the most important and common receptors in mammals, P1 and P2Y receptors are widely distributed in the brain, heart, kidneys, and adipose tissues. Because they are involved in various physiological functions, several ligands for these receptors are actively under consideration for various pharmaceutical treatments [2,3]. Although the P1 and P2Y receptors share common structural features of class A GPCRs, a comprehensive sequence and structural analysis showed that the P1 and P2Y receptors belong to two distinct groups, α and δ, respectively [6]. They have different ligand-binding site features that can distinguish between specific activators. These specific amino acid residues in the binding cavity may be involved in the selectivity and unique pharmacological behavior of each subtype.

For rational drug design of GPCR ligands, it is essential to have an in-depth understanding of the binding properties of ligands, the structural complexities of the receptor, the interaction between ligands and receptors, and the interaction between receptors and downstream signaling complexes [7]. In this study, we conducted a structure-based analysis of purinergic P1 and P2Y receptors to identify their evolutionary signature and obtain structural insights into ligand recognition and selectivity (Fig. 1). Although P1 and P2Y receptors belong to different subgroups in class A GPCRs, they also belong to the same classification as "purinergic", especially in terms of recognizing ligands with similar scaffolds. We conducted this study to reveal where the shared features and functional selectivity come from. In the case of P2X receptors, since they belong to an ion channel class, it was challenging to analyze the sequence and structures consistently in terms of receptor selectivity. Therefore, we focused on the GPCR family purinergic receptors in this study.

The structural features of the P1 and P2Y receptor classes were compared based on sequence conservation and ligand interaction patterns. Orthologous protein sequences were collected for P1 and P2Y receptors and aligned to produce multiple sequence alignment (MSA) tables for different levels of classes: (i) family level: P1 and P2Y, and (ii) subtype level: A_1, A_{2A}, A_{2B}, A_3, $P2Y_1$, $P2Y_{12}$, and others. Sequence conservation scores were calculated based on Shannon entropy to identify the highly conserved residues at each level. To analyze the ligand interaction patterns, we performed docking studies on the P1 and P2Y receptors using known ligand information extracted from the Chemogenomic European Molecular Biology Laboratory (ChEMBL) database [8]. We analyzed how the conserved residues are related to ligand-binding sites and how they differ between P1 and P2Y receptors. Based on this work, we extracted new similarities and differences between receptor subtypes, and the results can be used for designing new ligands by predicting key interaction residues that are important for functional selectivity.

2. Experimental procedures

2.1. Sequence collection and conservation analysis

The purinergic receptor protein sequences were retrieved from the UniProt Knowledgebase (UniProtKB) [9]. We searched the sequences with the terms 'adenosine receptor,' 'P2Y,' 'purinoceptor,' or 'purinergic receptor' annotated in the 'protein name' field. The fragment sequences were excluded, and data processing was performed to reduce errors in MSAs. All reviewed sequences were included, whereas among the unreviewed sequences, the ones that contain 'X' or 'B' or that are too short or too long (upper and lower 5 % in length) were omitted. The sequences annotated as 'low quality' or 'putative' were also removed. When there were too many sequences (for the P1 or P2Y whole-class level), MMseq2 [10] was used to perform sequence clustering with a similarity value of 0.9. The number of final collected sequences is summarized in Tables S1-S3.

Based on the collected sequences, MSA was performed using the MUSCLE5 program [11,12] with default settings. For the alignment of the large class level (P1 or P2Y), the super5 option was added to reduce the computation time. The sequence conservation score was calculated based on Shannon entropy [13] and normalized for each subtype. Plots of the sequence conservation for each subtype are shown in Fig. S1-S2. Hierarchical clustering was performed using a heat map of sequence conservation based on residue numbers. The residue numbers in this study followed the Ballesteros-Weinstein (B&W) numbering system, a method for numbering residues in the seven transmembrane (TM) helices of GPCRs [14].

B&W numbering is based on the most highly conserved residues within each of the seven TM helices [15]. It consists of two numbers: i) the first number represents the seven helices and ii) the second number indicates the position of the residue relative to the most conserved residue in the specific helix. The number of the most conserved residues, that is, the reference residue, is defined as number 50 (in the case of class A GPCR conservation: N1.50: 98 %, D2.50: 90 %, R3.50: 95 %, W4.50: 97 %, P5.50: 78 %, P6.50: 99 %, and P7.50: 88 % [16]). For example, 4.55 indicates the residue located in TM4, and is the 5th residue after W4.50, the most conserved residue in TM4. Thus, the B&W numbering system enables intuitive comparison of the relative positions within GPCR architectures [15].

After hierarchical clustering, a heatmap was generated using the pheatmap library (v.1.0.12) [17]. As the distance matrix computation method option, 'correlation' was used for a row, 'canberra' was used for a column, and 'ward.D2' was used as the clustering method.

2.2. Clustering of receptor structures and core module prediction

3D structural clustering was performed using Bio3D (R package) [18] based on the pairwise root-mean-square deviation (RMSD) values with the clustering method 'complete'. All available X-ray
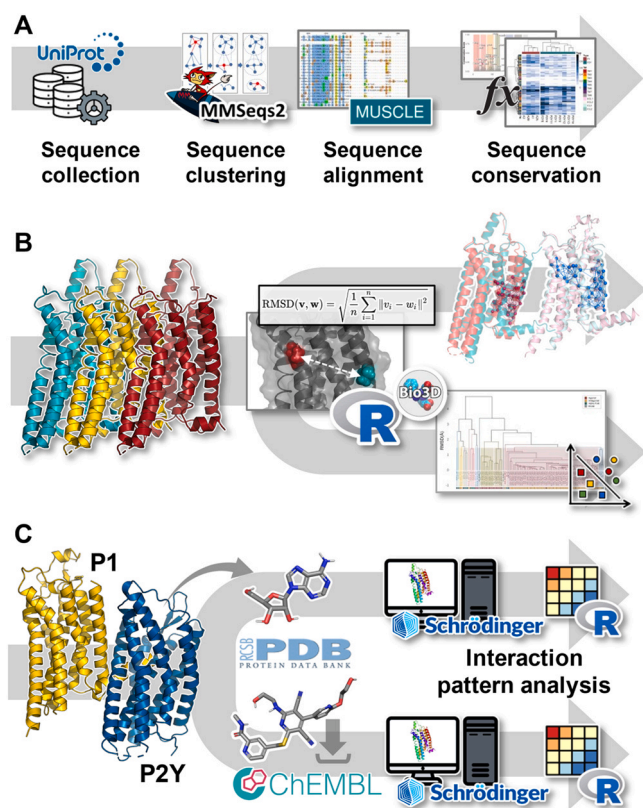

Fig. 1. Workflow of sequence- and structure-based clustering and interaction pattern analysis of purinergic receptors. (A) Sequence alignment and conservation score calculation. (B) Core superposition and clustering of 3D structures. (C) Ligand interaction pattern analysis using X-ray crystal structures and known ligands in ChEMBL.

Table 1
The experimental or predicted 3D structures used in this study.

Receptor subtype	PDB ids	ref.	Receptor subtype	PDB ids	ref.
$A_{2A}AR$	2YDO, 2YDV	[37]	$A_{2A}AR$	7RM5	[38]
	3EML	[36]		AF-P29274-F1	Alpha Fold DB
	3PWH, 3REY, 3RFM	[39]		6GT3	[40]
	3QAK	[41]		6JZH	[42]
	3UZA, 3UZC	[43]		6LPJ, 6LPK	[44]
	3VG9, 3VGA	[45]		6MH8	[46]
	4EIY	[47]		6PS7	[48]
	4UG2, 4UHR	[49]		6S0L, 6S0Q	[50]
	5G53	[51]	A_1AR	5N2S	[52]
	5IU4, 5IU7, 5IU8, 5IUA, 5IUB	[53]		5UEN	[54]
	5JTB	[55]		6D9H	[56]
	5K2A, 5K2B, 5K2C, 5K2D	[57]		7LD3, 7LD4	[58]
	5MZJ, 5MZP, 5N2R	[52]		AF-P30542-F1	Alpha Fold DB
	5NLX, 5NM2, 5NM4	[59]	$A_{2B}AR$	AF-P29275-F1	Alpha Fold DB
	5OLG, 5OLH, 5OLO, 5OLV, 5OLZ, 5OM1, 5OM4	[60]	A_3AR	Model structures	[21]
	5UIG	[61]		AF-P0DMS8-F1	Alpha Fold DB
	5UVI	[62]	$P2Y_1R$	4XNV, 4XNW	[63]
	5VRA	[64]		AF-P47900-F1	Alpha Fold DB
	5WF5, 5WF6	[65]	$P2Y_2R$	AF-P41231-F1	Alpha Fold DB
	6AQF	[66]	$P2Y_4R$	AF-P51582-F1	Alpha Fold DB
	6GDG	[67]	$P2Y_6R$	AF-Q15077-F1	Alpha Fold DB
	6WQA	[68]	$P2Y_{11}R$	AF-Q96G91-F1	Alpha Fold DB
	6ZDR, 6ZDV	[69]	$P2Y_{12}R$	4NTJ	[70]
	7ARO	[71]		4PXZ, 4PY0	[72]
	7EZC	[73]		AF-Q9H244-F1	Alpha Fold DB
	7PX4, 7PYR	[74]	$P2Y_{13}R$	AF-Q9BPV8-F1	Alpha Fold DB
			$P2Y_{14}R$	AF-Q15391-F1	Alpha Fold DB

Table 2
The number of searched ligands for the given target from the ChEMBL database.

Receptor subtype	UniProt ID	Ligand type	Standard type	Number of compounds
A_1	P30542	Agonist	EC_{50}	245
A_1	P30542	Antagonist	IC_{50}	215
A_{2A}	P29274	Agonist	EC_{50}	170
A_{2A}	P29274	Antagonist	IC_{50}	298
A_3	P0DMS8	Agonist	EC_{50}	138
A_3	P0DMS8	Antagonist	IC_{50}	369
$P2Y_1$	P47900	Antagonist	IC_{50}	180
$P2Y_{12}$	Q9H244	Agonist	EC_{50}	70
$P2Y_{12}$	Q9H244	Antagonist	IC_{50}	533

crystal or cryo-EM structures for adenosine receptors (A_{2A} and A_1) and P2Y receptors ($P2Y_1$ and $P2Y_{12}$) were retrieved from the Protein Data Bank (PDB) [19,20]. The experimental structures used in this study are summarized in Table 1. For A_3AR, previously constructed model structures [21] were used. For other subtypes whose experimental structures were not available, alpha-fold models [22,23] were also used. After removing unnecessary molecules (that is, engineered parts and buffer molecules), all structures were aligned based on their sequence homology.

During the clustering process, an ellipsoid was created based on the location of the alpha carbons of a specific residue in all aligned proteins. If the volume of the ellipsoid was < 1 Å^3, the residues were selected as the core region of the protein in that cluster (that is, the most overlapped region in the cluster). In the cluster dendrogram, the number of cluster groups was divided into 15 trees, with a cut-off RMSD value of 1.5 Å.

2.3. Data extraction of known ligands reported with their biological activities

Known ligands of purinergic P1 and P2 receptors were searched in the ChEMBL database [8]. Using a Python client library officially supported by ChEMBL [24], the search was conducted based on the target's UniProt ID to retrieve the reported ligand data with

biological activities on the target. For a given target, the criteria for ligand selection were as follows: i) assay type label was B-binding and F-function, ii) assay standard type was IC_{50} for the antagonist and EC_{50} for the agonist, iii) standard values of activity were < 10 μM, and iv) the molecular weight was < 600.

For the resulting ligand sets (summarized in Table 2), ligand similarity analysis was performed based on MACCSKeys fingerprints and Tanimoto similarity index. The scikit-learn library [25] and RDKit [26] were used for analysis. After creating a distance matrix, it was easily expressed in two dimensions by graphing it using the t-distributed Stochastic Neighbor Embedding (t-SNE) method with a random_state value of 0, an angle value of 0.5, and a perplexity value of 50.

2.4. Ligand binding mode prediction

The available X-ray crystal structures of P1 (A_1 and A_{2A}) and P2Y receptors ($P2Y_1$ and $P2Y_{12}$) were used for docking studies of known ligands retrieved from the ChEMBL database. For A_3AR, our previously constructed model structures were used [21]. Each receptor structure was prepared using "Protein Preparation Workflow" [27] in Schrödinger Maestro v.13.3 [28]. The structures and activity data of the known ligands of each receptor subtype were collected from the ChEMBL database. The ligand structures were prepared using the "LigPrep" module [29] in Schrödinger with ionization states at pH 7.4. The Optimum Potential for Liquid Simulation (OPLS) 4 force field [30] was used for the energy minimization.

Ensemble docking studies for each receptor subtype were performed by using a "virtual screening process" procedure incorporating Glide docking [31,32]. As protein parts for docking simulation, multiple structures (PDB ids: 2YDO, 2YDV, 3EML, 3QAK, 4UG2, 4UHR, 5G53, 5IU7, 5IU8, 5MZJ, 5N2R, 5OLH, 5OLO, 5OLV, 5OLZ, 5UIG, 6GT3, 6WQA, and 6ZDR for $A_{2A}AR$; 5N2S, 5UEN, 6D9H, 7LD3, and 7LD4 for A_1AR; model structures for A_3AR; 4XNW for $P2Y_1R$; 4NTJ, 4PXZ, and 4PY0 for $P2Y_{12}R$) were used to incorporate the protein flexibility. Post-docking minimization using the OPLS3 [33] force field was followed, and the top-ranked ligand conformation in the multiple receptor structures was finally selected as the binding mode. The interaction energy (E_{int}) with the bound ligand was calculated for the binding site residues
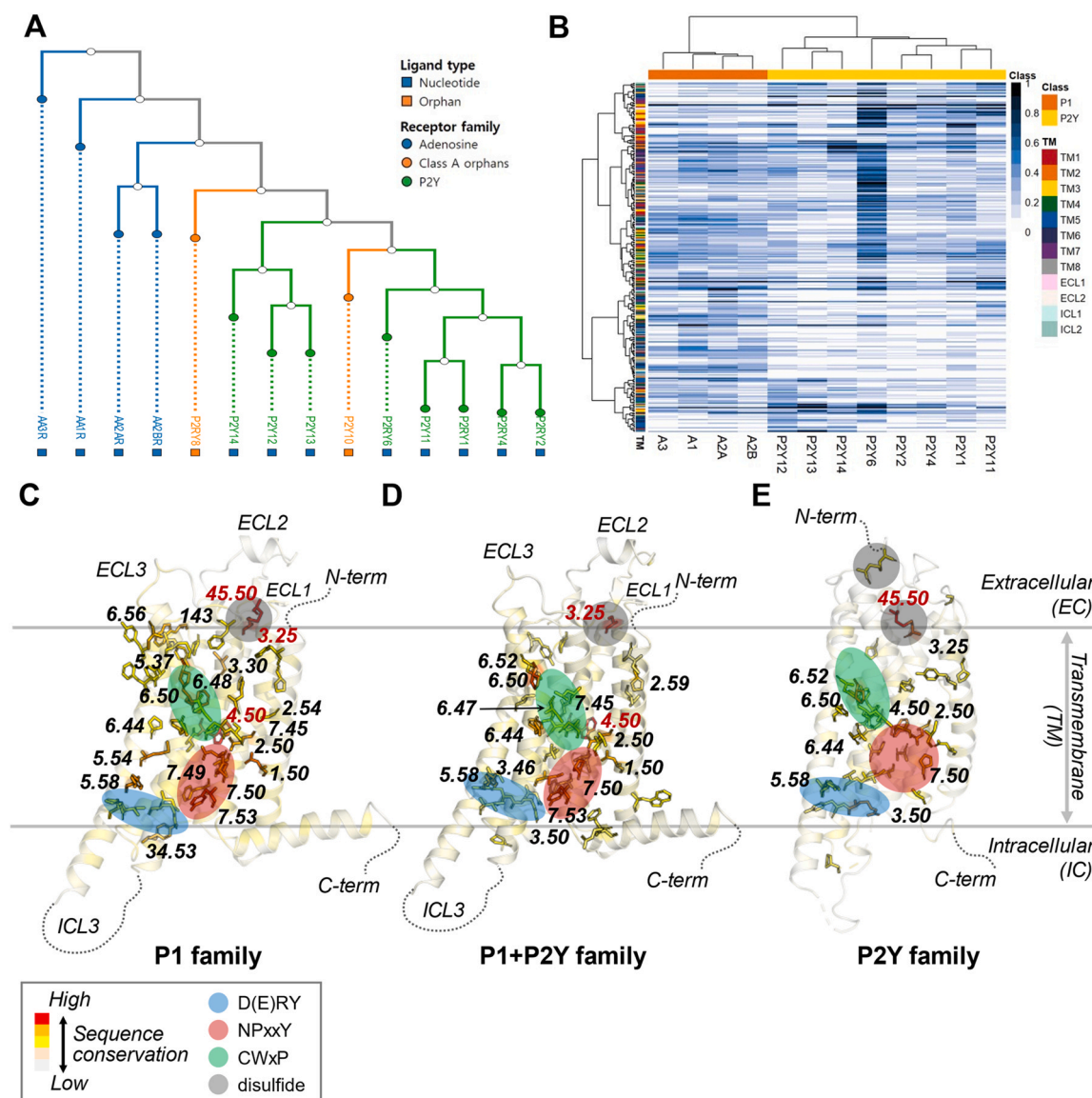
Fig. 2. Sequence conservation of the purinergic receptors. (A) Phylogenetic tree of P1 and P2Y receptors generated using the Receptor Similarity module in GPCRdb. (B) Sequence clustering based on the conservation patterns of the homologous protein family. Sequence conservation scores were calculated in different receptor family levels: (C) P1, (D) P1 and P2Y, and (E) P2Y receptor families. Highly conserved residues were mapped to the representative 3D structures of human $A_{2A}AR$ and $P2Y_{12}$ receptors (PDB id: 2YDO and 4PXZ, respectively).

within 12 Å of the grid center. All molecular graphic figures were generated using PyMOL v.2.5.4 software [34].

2.5. Hierarchical clustering of the interaction maps

Interaction maps were expressed through heatmaps using the receptor-ligand interaction energy (E_{int}) of the binding pocket residues (note that the residue numbers follow the B&W numbering system). Hierarchical cluster analysis of the interaction maps was performed with the pheatmap R package v.1.0.12 software [17] using 'ward.D2' (Ward's minimum variance method) as the clustering method.

3. Results and discussion

3.1. Identification of evolutionary features through protein sequence conservation analysis

The purinergic signaling system is thought to have appeared early in the evolution of life [35]. The first specific receptors are

observed in single-cell eukaryotic protozoa, and ATP-degrading enzymes and release mechanisms are present in bacteria [35]. Throughout further evolution of the purinergic signaling system, a wide range of purinoceptor classes have been developed, as well as pathways capable of releasing nucleotides and adenosine [35]. As class A GPCRs, P1 and P2Y receptors have common structural features of the 7-transmembrane (TM) system. They also share structural similarities in the recognition of purine scaffolds [1]. To compare the evolutionary features of purinergic receptor families, we conducted conservation analysis using the collected homologous sequences of purinergic P1 and P2Y receptors.

In terms of phylogenetic relationships (Fig. 2A; calculated using the receptor similarity module in GPCRdb [36]), the P1 and P2Y families can be divided distinctively. Among P1 receptors, $A_{2A/2B}$ adenosine receptors (AR) are seemingly more similar to A_1AR than A_3AR. In the case of P2Y families, two distinct subgroups with a relatively high level of divergence were observed: the first ($P2Y_1$-like) subgroup contained $P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$, and $P2Y_{11}$ receptors, and the second ($P2Y_{12}$-like) subgroup contained $P2Y_{12}$, $P2Y_{13}$, and $P2Y_{14}$ receptors.
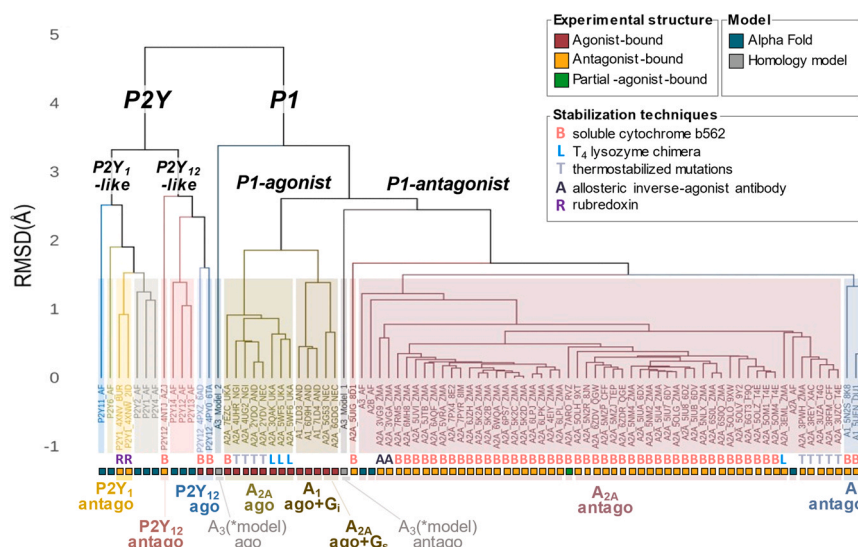
Fig. 3. Clustering of the available 3D structures of the P1 and P2Y receptors. Hierarchical clustering was performed using the aligned 3D structures of the pure receptors (the engineered portions were omitted). At the bottom, the labels for experimental or model structures, stabilization techniques for crystallization, and receptor or ligand subtypes are annotated.
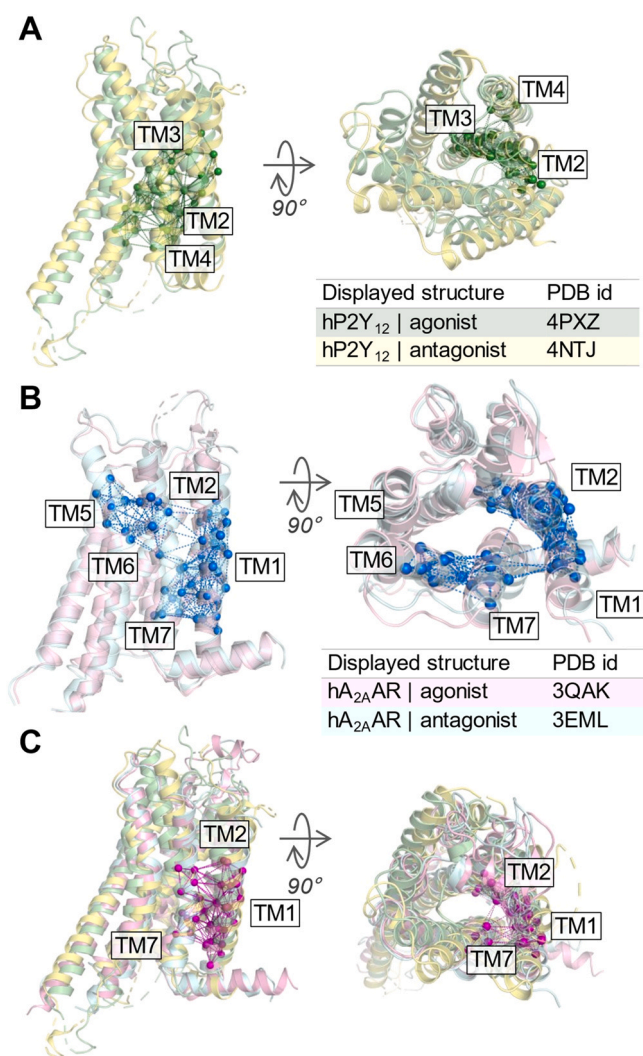

Fig. 4. Most overlapped core regions in purinergic receptor families. (A) P2Y receptors, (B) P1 receptors, and (C) both P1 and P2Y receptors. The core regions are represented by spheres and networks.

We also analyzed sequence conservation in each receptor subtype family by aligning ortholog sequences collected from UniProt (Fig. 2B; note that $P2Y_8$ and $P2Y_{10}$, which are classified as orphan receptors, were not included). Sequence conservation scores were calculated using the Shannon entropy-based method [13], and residue numbers were represented using the GPCR B&W numbering system [15]. Interestingly, the hierarchically clustered sequence conservation matrix (Fig. 2B) showed evolutionary differences among receptor families and relative positions in the 7TM topology of GPCRs. The loops and C-terminal regions showed greater variability than did the TM regions. As expected, the functional key residues, also known as micro-switches (for example, N24, D52, D101, R102, Y103, W129, P189, Y197, E228, C245, W246, P248, N280, S281, N284, P285, Y288, and F295 in the human $A_{2A}AR$), showed significantly high conservation scores in all receptor classes (Fig. 2C–E, Fig. S2).

In class A GPCRs, universally conserved positions, such as microswitches, allow receptors to bind G proteins in a similar orientation, thus modulating receptor activation. The subtype-specific residues surrounding the conserved core may serve as a selectivity filter to ensure that different receptors bind differently, that is, the P2 family responds to ATP, ADP, UTP, and UDP, whereas the P1 receptor family recognizes its natural agonist adenosine. Interestingly, the extracellular half of the TM region of the P2 receptor exhibited less sequence similarity than that of P1, reflecting the greater diversity of their natural ligands.

3.2. Clustering of 3D structures and common core regions of each class

For over few decades, active research and innovation have accumulated in the field of GPCR structures, rendering $A_{2A}AR$ one of the best structurally characterized GPCRs at the atomic level [41]. In particular, the experimental structures were determined in various pharmacological states, that is, antagonist-bound inactive, agonist-bound active, and fully active states, together with G-proteins. The structures of other subtypes such as A_1AR, $P2Y_1$, and $P2Y_{12}$ are also available. These structures have highlighted key structural features in the activation mechanism of the receptor and provided structural insight into the molecular determinants of ligand binding specificity in different receptor classes. Moreover, the Alpha Fold (AF), a successful artificial intelligence (AI) system that predicts the 3D structure of a protein from a given amino acid sequence, provides

A

B
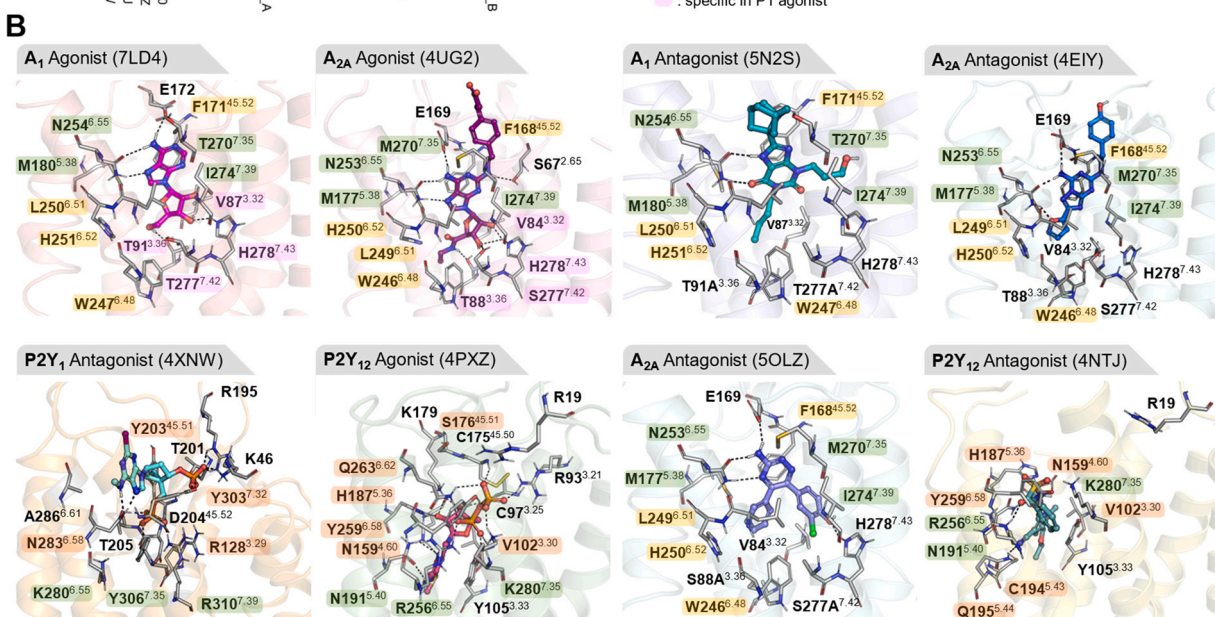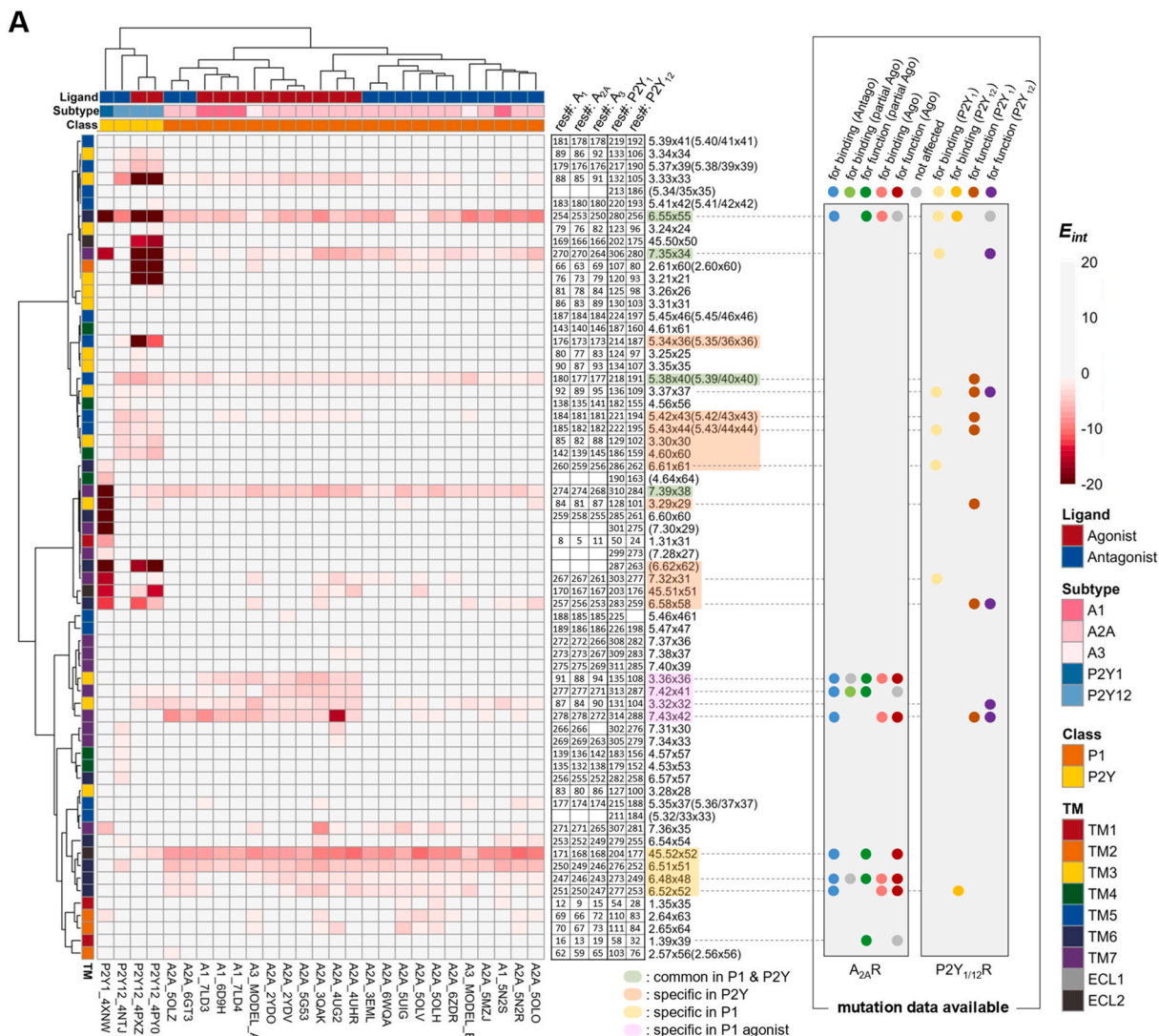
(caption on next page)

Fig. 5. Ligand interaction pattern analysis using the crystal structures of P1 (adenosine A₁, A₂A, and A₃) and P2Y (P2Y₁ and P2Y₁₂) receptors. (A) Hierarchically clustered interaction map with the distance matrix computation methods 'correlation' for a row and 'canberra' for a column. The color in the heatmap means the interaction energy (E_int) calculated by using the crystal structures. The residue numbers and B&W numbers are marked in the middle column. At the right side, the positions with the published mutation data (examined for ligand binding and activation) [76–78] are marked with colored circles. (B) Experimentally validated protein-ligand interactions of the representative receptors. The bound agonists or antagonists are displayed as ball-and-stick models, and the interacting residues are represented by gray sticks. The H-bond interactions are depicted in black dashed lines. The residue numbers significant for both P1/P2Y-binding are highlighted in green. The residues observed specifically in P1- and P2Y- bindings are marked in yellow and orange, respectively. The key interacting residues for P1 agonists are highlighted in pink.
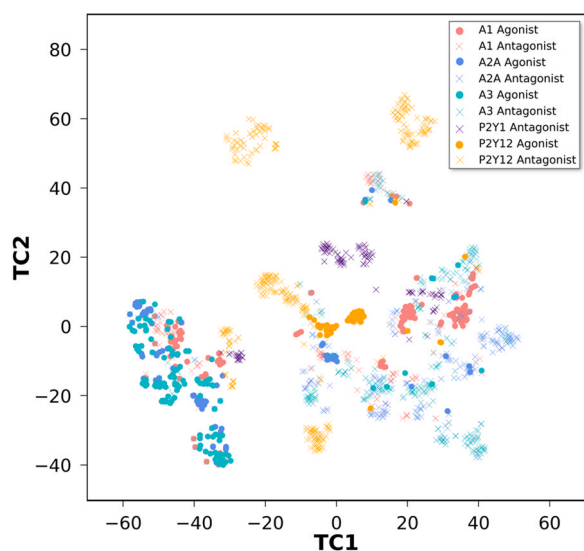

Fig. 6. Chemical space analysis of the known purinergic receptor ligands collected from the ChEMBL database. Chemical structures were clustered using t-SNE based on the MACCSKeys fingerprint. This analysis was performed only when the total number of ligands of the receptor subtype was > 200.

individual structures for the human proteome with high coverage [22,23]. These structural sources provide valuable information about receptor function, activation mechanisms, and ligand interactions that can be used in rational drug design processes.

We systematically analyzed the structural data reported for the purinergic P1 and P2Y receptors, that is, the X-ray crystal structures of P2Y₁R, P2Y₁₂R, A₁AR, and A₂AAR, our previously constructed homology models of A₃AR [21], and AF models of other subtypes (Table 1). After deleting the regions artificially engineered for crystallography, such as the nanobody or T4 lysozyme, we prepared the sole purinergic receptor structure with ligands bound at the orthosteric sites. All the structures were aligned based on their sequence homology. These structural ensembles were clustered based on their structural similarities as calculated using Cα-RMSDs. The 3D structural clustering itself also distinguished the receptor subtypes (Fig. 3).

Interestingly, in both the P1 and P2Y cases, the agonist-bound structures were clearly distinguishable from the antagonist-bound structures, reflecting the structural differences between the two groups. In the absence of G-protein binding, the agonist- and antagonist-bound A₂AAR structures show only microscopic changes in their helical arrangements; however, this fine-tuning of structures shows some systematic trends that can be rationally applied in drug discovery. Additionally, small clusters in the same receptor subtypes can be distinguished by differences in the bound ligand and also by the stabilization techniques during the crystallization. We can speculate that the most significantly affecting reason for the receptor conformational change is the ligands' pharmacological profile (agonist or antagonist), followed by the ligand scaffold, and then the crystallization techniques.

During the 3D structural clustering process, the most overlapped core region was calculated for the aligned structures in the same cluster (note that this core region is the most structurally similar and well-aligned section of the same cluster). The core regions are

displayed as spheres and networks in Fig. 4. Interestingly, the core regions in the P1, P2Y, and purinergic families were different. The core region in the P2Y family structures tended to be in the intracellular half of TM2, 3, and 4 (Fig. 4A), whereas the core region in the P1 family was observed in the middle of TM1, 2, 5, 6, and 7 (Fig. 4B). Regarding the entire purinergic family (P1 and P2Y combined), the core regions were located in TM1, 2, and 7 (Fig. 4C).

3.3. Systematic analysis of the receptor-ligand interaction patterns observed in the crystal structures

Although sequence conservation and the topology of receptor structures provide valuable information for protein function, it is not sufficient to extract more detailed evidence on which interactions are crucial for ligand binding or receptor modulation. Since receptor-ligand binding is determined by the physicochemical interactions between the ligands with various chemical scaffolds and the surrounding amino acids, it is necessary to analyze the interactions of individual structures. Based on currently available experimental structures, we systematically investigated the interactions between purinergic receptors and various types of ligands.

By examining a dataset of 25 representative crystal structures (for A₁, A₂A, P2Y₁, and P2Y₁₂ receptors) and two model structures (for A₃AR), we calculated the interaction energy map between the receptor and bound ligands. First, the crystal structures were refined using the Schrödinger Glide program with the 'Refine only' option (note that the bound conformations of the bound ligands were nearly the same as the crystallized ones, only refined to calculate reasonable energy values). The interaction energy (E_int) values were extracted for the binding site residues (within 12 Å from the grid center), and the residue numbers were replaced with the corresponding B&W numbers. We then performed hierarchical clustering analysis to extract common or specific interaction patterns in the receptor families.

As illustrated in Fig. 5, the calculated interaction map shows some shared binding features among the receptor families. First, nucleoside- or nucleotide-like agonist scaffolds appear to be recognized by similar residues when they bind to the orthosteric sites of P1 and P2Y receptors. In both P1 and P2Y receptors, common interaction patterns with the residues 5.38, 6.55, 7.35, and 7.39 appear important for purinergic ligand recognition (note that all the numbers in this section follow B&W numbering). Interestingly, these residues may function as recognition points for purinergic ligands but play different roles in ligand binding in the P1 and P2Y classes. In P1, residues 5.38 and 6.55 form the interaction with the adenine moiety of the natural substrate adenosine. In contrast, in P2Y, these residues capture the interaction with the phosphate group of the natural substrate, such as ATP or AMP (see the green highlighted residues in Fig. 5).

The map also shows some interaction patterns common to P1 receptors, but not observed in P2Y receptors (see the yellow-highlighted residues in Fig. 5). The adenine or aromatic rings in the ligands almost always form π-π stacking with the hydrophobic residue 45.52 in ECL2 (upper part of the binding site). Many crystal structures of A₂AAR also confirmed that most of the ligands occupy the same binding pocket, maintaining H-bonds with the conserved residue 6.55 in all AR subtypes. In the middle of the binding site, hydrophobic interactions with residues 6.51 and 6.52 may be crucial for P1. The residue 6.48, one of the most highly conserved micro-
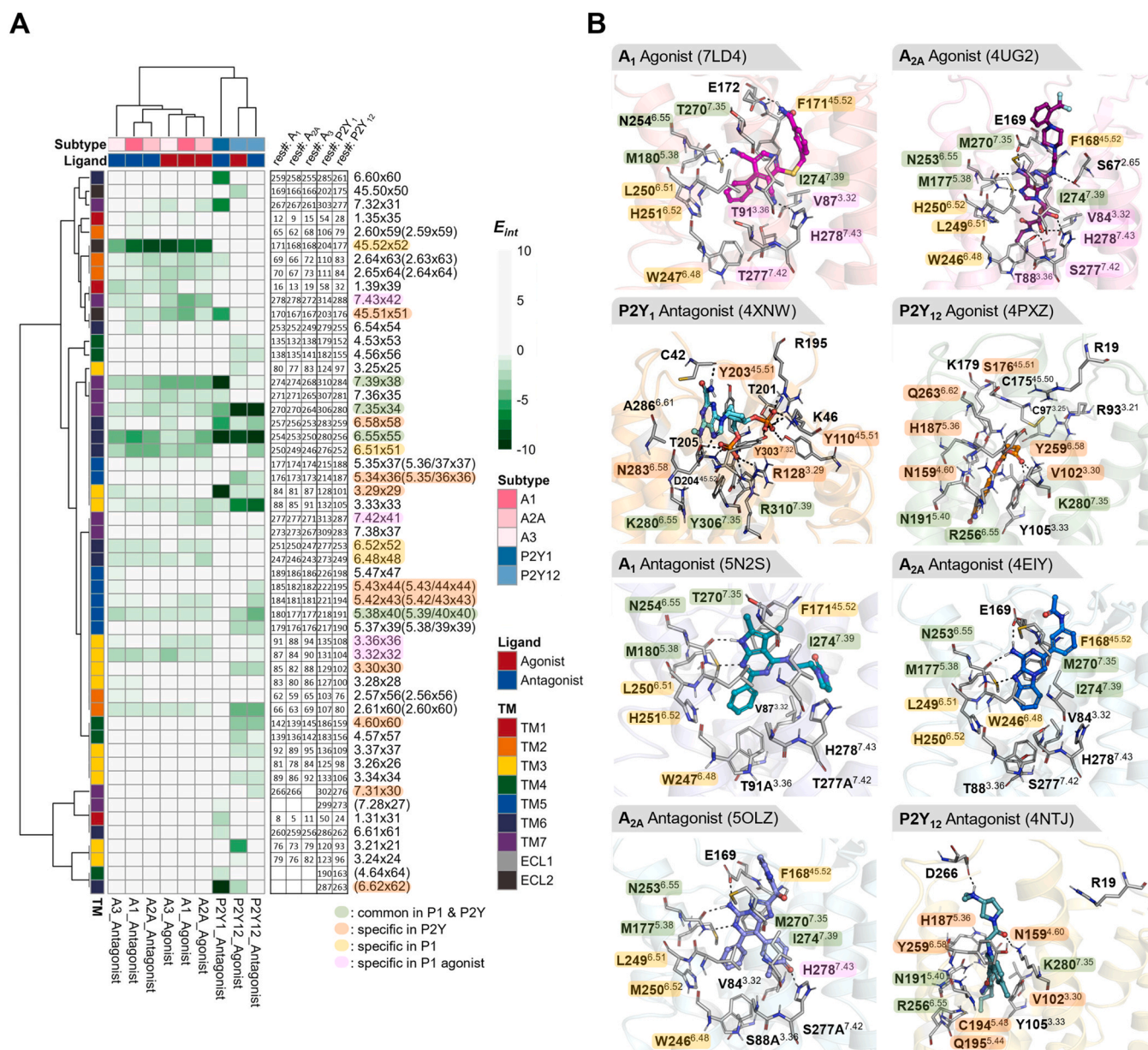
Fig. 7. Interaction pattern map of the P1 & P2Y receptor ligands collected from the ChEMBL database. (A) Hierarchically clustered interaction map with the distance matrix computation methods 'binary' for a row and 'canberra' for a column. The color in the heatmap means the average interaction energy (E_{int}) calculated by the docked conformations of the known ligands from the ChEMBL database. (B) Structural image of interaction pattern map of the exemplary P1 & P2Y receptor ligands.

switch residues, contributes to a typical hydrophobic interaction at the bottom of the binding site, exhibiting a common pattern of P1 ligand binding.

Interestingly, key interaction patterns for agonist binding were also detected in P1 receptors (see the pink-highlighted residues in Fig. 5). Interactions with 3.30, 3.32, 7.42, and 7.43 were commonly observed in the agonist-bound structures. Significantly, the H-bonds of the polar groups in the agonist ligands with residues 7.42 and 7.43 can be critical for agonist binding. These residues form H-bonds with the antagonist in some instances, but several studies have shown that these interactions with the antagonist can be transient and less persistent during molecular dynamics simulations [75]. As shown in the right side in Fig. 5A, some of these residues were also confirmed by site-directed mutational studies [76–78] for their roles in ligand binding or receptor function.

In the case of P2Y receptors, several interacting residues, such as 45.51, 6.58, and 7.32, may be necessary for P2Y ligand recognition (orange-highlighted residues in Fig. 5). Taken together with the published mutagenesis data [76,78], it appears that the essential amino acid residues for P2Y receptor activation are located at the extracellular side of TM5 and TM7 through their interaction with TM3. The residues in TM6 and ECL3 are observed as more crucial for receptor binding than receptor activation. Since the available X-ray crystal structures of P2Y receptors are limited, more reliable analysis of P2Y ligand binding is planned postponed for future study.

Among the key interaction residues (with good E_{int} values; note that lower E_{int} values indicate better binding), some are consistent with the sequence conservation profiles, whereas others show unique features regardless of sequence conservation (Fig. S3). Considering that protein structures are more conserved than protein sequences, it is possible to make evolutionary inferences through comparative structure analyses, also without considering the sequence information. Particularly with respect to ligand binding or

selectivity, structural conservation is seemingly more important than the sequence level.

3.4. Receptor-ligand interaction patterns calculated using the known ligands in ChEMBL

The observed binding patterns in the crystal structures will help us gain more insights into how these molecules bind with other molecules and how their specificity differs. Although these structures provide valuable information on receptor-ligand interactions, the types of bound ligand scaffolds are limited. To determine whether the resulting interaction patterns are consistent with various types of ligands, we collected validated ligand sets for each subtype in the ChEMBL database (Table 2) and predicted their binding poses and interaction energies.

As shown in Fig. 6, the chemical spaces of the known ligands overlap among the different receptor subtypes. It may differentiate between P1 and P2Y receptor ligands, but there are also some shared chemical spaces with other receptors. For P1 receptors, the known ligand scaffolds are in a widely shared chemical space, reflecting their chemical similarity. Therefore, selectivity among receptor subtypes can be achieved through detailed individual interactions with the surrounding receptor environment.

To predict the binding modes of these known ligands, docking studies were performed using Schrödinger Glide (SP mode). To consider receptor flexibility, the representative protein structures shown in Fig. 5 were used to perform ensemble docking. Interaction energy (E_{int}) values were extracted for the binding site residues (within 12 Å from the grid center) and averaged for each receptor subtype. Therefore, the good E_{int} value of a specific residue indicates the importance of the receptor-ligand interactions commonly observed in various types of ligands. As shown in Fig. 7, we confirmed similar interaction patterns in our previous analysis of the X-ray crystal structures. For subtypes with unknown X-ray structures, ligand docking and interaction pattern analysis were also attempted using the reported alphafold (AF) models; however, we failed to obtain reliable poses to calculate E_{int}. This may be because many AF models were predicted based on apo receptor structures; the coordinates of the binding site cavity might not be suitable for detailed analysis of ligand-binding interactions and require further optimization.

4. Conclusion

In summary, sequence conservation and ligand interaction pattern analyses of purinergic receptors allowed the identification of several common or distinct features in the P1 and P2Y receptor families. Using the information of well-characterized purinergic ligands, we proposed their binding modes and potential key residues for selectivity against receptor subtypes or pharmacological profiles. The conservation score of a residue is a useful measure of its importance in protein structure and function; however, there may be less association with individual ligand interactions. Receptor-ligand information should be systematically analyzed for various cases to extract the knowledge necessary for molecular design. The P1 and P2Y receptors present different binding pockets that share only a small portion of the site. Systematic analysis showed that some of these patterns can help classify receptor-ligand interactions. Our study can be used to improve GPCR drug design by systematically analyzing valuable information to our understanding of GPCR receptor-ligand interactions and their ligand selectivity and functionality.

Funding sources

CRediT authorship contributions statement

Ri Han: Methodology, Software, Data curation, Validation, Writing – Original draft; **Hongryul Yoon**: Data curation, Validation; **Jiho Yoo**: Investigation, Supervision; **Yoonji Lee**: Conceptualization, Methodology, Writing – Original Draft, Writing-Reviewing and Editing, Supervision, Project administration.

Declarations of Competing Interest

The authors declare no competing financial or other interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.01.010.

References

[1] Huang Z, Xie N, Illes P, Di Virgilio F, Ulrich H, Semyanov A, et al. From purines to purinergic signalling: molecular functions and human diseases. Signal Transduct Tar 2021;6:162.
[2] Mahmood A, Iqbal J. Purinergic receptors modulators: an emerging pharmacological tool for disease management. Med Res Rev 2022;42:1661–703.
[3] Burnstock G. Purinergic signalling: therapeutic developments. Front Pharmacol 2017;8:661.
[4] Jacobson KA, Gao ZG. Adenosine receptors as therapeutic targets. Nat Rev Drug Discov 2006;5:247–64.
[5] Dorsam RT, Murugappan S, Ding Z, Kunapuli SP. Clopidogrel: interactions with the P2Y₁₂ receptor and clinical relevance. Hematology. 2003;8:359–65.
[6] Burnstock G. Purine and purinergic receptors. Brain Neurosci Adv 2018;2. 2398212818817494.
[7] Lee Y, Basith S, Choi S. Recent advances in structure-based drug design targeting class A G protein-coupled receptors utilizing crystal structures and computational simulations. J Med Chem 2018;61:1–46.
[8] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 2018;47:D930–40.
[9] Consortium TU. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. 2020;49:D480-D9.
[10] Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. Nature Commun 2018;9:2542.
[11] Edgar R.C. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. BioRxiv. 2021:2021.06.20.449169.
[12] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 2004;5:113.
[13] Lee Y, Choi S, Hyeon C. Mapping the intramolecular signal transduction of G-protein coupled receptors. Proteins. 2014;82:727–43.
[14] Ballesteros JA, Weinstein H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Methods in Neurosciences. Elsevier; 1995. p. 366–428.
[15] Isberg V, de Graaf C, Bortolato A, Cherezov V, Katritch V, Marshall FH, et al. Generic GPCR residue numbers - aligning topology maps while minding the gaps. Trends Pharmacol Sci 2015;36:22–31.
[16] Isberg V, Vroling B, van der Kant R, Li K, Vriend G, Gloriam D. GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 2014;42:D422–5.
[17] Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0. 12. 2019.
[18] Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22:2695–6.
[19] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res 2000;28:235–42.
[20] Berman H, Henrick K, Nakamura H. Announcing the worldwide protein data bank. Nature Struct Mol Biol 2003;10:980.
[21] Lee Y, Hou X, Lee JH, Nayak A, Alexander V, Sharma PK, et al. Subtle chemical changes cross the boundary between agonist and antagonist: new A₃ adenosine receptor homology models and structural network analysis can predict this boundary. J Med Chem 2021;64:12525–36.
[22] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.
[23] Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 2022;50:D439–44.
[24] Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, et al. ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res 2015;43:W612–20.
[25] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Machine Learn Rese 2011;12:2825–30.

[26] Landrum G. RDKit: open-source cheminformatics. 2016.

[27] Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput-Aided Mol Design 2013;27:221–34.

[28] Maestro. Schrödinger Release 2021–3. Maestro, Schrödinger, LLC New York, NY; 2021.

[29] LigPrep. Schrödinger Release 2022–3. LigPrep, Schrödinger, LLC, New York, NY, 2021.; 2021.

[30] Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, et al. OPLS4: improving force field accuracy on challenging regimes of chemical space. J Cheml Theory Comput 2021;17:4291–300.

[31] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–49.

[32] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein− ligand complexes. J Med Chem 2006;49:6177–96.

[33] Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J Chem Theory Comput 2016;12:281–96.

[34] Schrödinger L., DeLano W. The PyMOL molecular graphics system, version 2.0 Schrödinger, LLC (2017).

[35] Verkhratsky A, Burnstock G. Biology of purinergic signalling: its ancient evolutionary roots, its omnipresence and its multiple functional significance. Bioessays. 2014;36:697–705.

[36] Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, et al. The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science 2008;322:1211–7.

[37] Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AG, et al. Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature 2011;474:521–5.

[38] Martynowycz MW, Shiriaeva A, Ge X, Hattne J, Nannenga BL, Cherezov V, et al. MicroED structure of the human adenosine receptor determined from a single nanocrystal in LCP. Proc Natl Acad Sci 2021;118:e2106041118.

[39] Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, et al. Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. Structure 2011;19:1283–93.

[40] Borodovsky A, Barbon CM, Wang Y, Ye M, Prickett L, Chandra D, et al. Small molecule AZD4635 inhibitor of A2AR signaling rescues immune cell function including CD103+ dendritic cells enhancing anti-tumor immunity. J Immunother Cancer 2020;8:e000417.

[41] Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao Z-G, et al. Structure of an agonist-bound human A_{2A} adenosine receptor. Science 2011;332:322–7.

[42] Shimazu Y, Tono K, Tanaka T, Yamanaka Y, Nakane T, Mori C, et al. High-viscosity sample-injection device for serial femtosecond crystallography at atmospheric pressure. J Appl Crystallogr 2019;52:1280–8.

[43] Congreve M, Andrews SP, Doré AS, Hollenstein K, Hurrell E, Langmead CJ, et al. Discovery of 1, 2, 4-triazine derivatives as adenosine A_{2A} antagonists using structure based drug design. J Med Chem 2012;55:1898–903.

[44] Ihara K, Hato M, Nakane T, Yamashita K, Kimura-Someya T, Hosaka T, et al. Isoprenoid-chained lipid EROCOC17+ 4: A new matrix for membrane protein crystallization and a crystal delivery medium in serial femtosecond crystallography. Sci Rep 2020;10:1–12.

[45] Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, et al. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. Nature 2012;482:237–40.

[46] Martin-Garcia JM, Zhu L, Mendez D, Lee M-Y, Chun E, Li C, et al. High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. IUCrJ 2019;6:412–25.

[47] Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, et al. Structural basis for allosteric regulation of GPCRs by sodium ions. Science 2012;337:232–6.

[48] Ishchenko A, Stauch B, Han GW, Batyuk A, Shiriaeva A, Li C, et al. Toward G protein-coupled receptor structure-based drug design using X-ray lasers. IUCrJ 2019;6:1106–19.

[49] Lebon G, Edwards PC, Leslie AG, Tate CG. Molecular determinants of CGS21680 binding to the human adenosine A_{2A} receptor. Mol Pharmacol 2015;87:907–15.

[50] Nass K, Cheng R, Vera L, Mozzanica A, Redford S, Ozerov D, et al. Advances in long-wavelength native phasing at X-ray free-electron lasers. IUCrJ. 2020;7:965–75.

[51] Carpenter B, Nehmé R, Warne T, Leslie AG, Tate CG. Structure of the adenosine A_{2A} receptor bound to an engineered G protein. Nature 2016;536:104–7.

[52] Cheng RK, Segala E, Robertson N, Deflorian F, Doré AS, Errey JC, et al. Structures of human A_1 and A_{2A} adenosine receptors with xanthines reveal determinants of selectivity. Structure 2017;25:1275–85. e4.

[53] Segala E, Guo D, Cheng RK, Bortolato A, Deflorian F, Doré AS, et al. Controlling the dissociation of ligands from the adenosine A_{2A} receptor through modulation of salt bridge strength. J Med Chem 2016;59:6470–9.

[54] Glukhova A, Thal DM, Nguyen AT, Vecchio EA, Jörg M, Scammells PJ, et al. Structure of the adenosine A_1 receptor reveals the basis for subtype selectivity. Cell 2017;168:867–77. e13.

[55] Melnikov I, Polovinkin V, Kovalev K, Gushchin I, Shevtsov M, Shevchenko V, et al. Fast iodide-SAD phasing for high-throughput membrane protein structure determination. Sci Adv 2017;3:e1602952.

[56] Draper-Joyce CJ, Khoshouei M, Thal DM, Liang Y-L, Nguyen AT, Furness SG, et al. Structure of the adenosine-bound human adenosine A_1 receptor–Gi complex. Nature 2018;558:559–63.

[57] Batyuk A, Galli L, Ishchenko A, Han GW, Gati C, Popov PA, et al. Native phasing of x-ray free-electron laser data for a G protein–coupled receptor. Sci Adv 2016;2:e1600292.

[58] Draper-Joyce CJ, Bhola R, Wang J, Bhattarai A, Nguyen AT, O'Sullivan K, et al. Positive allosteric mechanisms of adenosine A_1 receptor-mediated analgesia. Nature 2021;597:571–6.

[59] Weinert T, Olieric N, Cheng R, Brünle S, James D, Ozerov D, et al. Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. Nature Commun 2017;8:1–11.

[60] Rucktooa P, Cheng RK, Segala E, Geng T, Errey JC, Brown GA, et al. Towards high throughput GPCR crystallography: in Meso soaking of Adenosine A_{2A} Receptor crystals. Sci Rep 2018;8:1–7.

[61] Sun B, Bachhawat P, Chu ML-H, Wood M, Ceska T, Sands ZA, et al. Crystal structure of the adenosine A_{2A} receptor bound to an antagonist reveals a potential allosteric pocket. Proc Natl Acad Sci 2017;114:2066–71.

[62] Martin-Garcia JM, Conrad CE, Nelson G, Stander N, Zatsepin NA, Zook J, et al. Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. IUCrJ 2017;4:439–54.

[63] Zhang D, Gao Z-G, Zhang K, Kiselev E, Crane S, Wang J, et al. Two disparate ligand-binding sites in the human $P2Y_1$ receptor. Nature 2015;520:317–21.

[64] Broecker J, Morizumi T, Ou W-L, Klingel V, Kuo A, Kissick DJ, et al. High-throughput in situ X-ray screening of and data collection from protein crystals at room temperature and under cryogenic conditions. Nature Protocols 2018;13:260–92.

[65] White KL, Eddy MT, Gao Z-G, Han GW, Lian T, Deary A, et al. Structural connection between activation microswitch and allosteric sodium site in GPCR signaling. Structure 2018;26:259–69. e5.

[66] Eddy MT, Lee M-Y, Gao Z-G, White KL, Didenko T, Horst R, et al. Allosteric coupling of drug binding and intracellular signaling in the A_{2A} adenosine receptor. Cell 2018;172:68–80. e12.

[67] García-Nafría J, Lee Y, Bai X, Carpenter B, Tate CG. Cryo-EM structure of the adenosine A_{2A} receptor coupled to an engineered heterotrimeric G protein. Elife 2018;7:e35946.

[68] Lee M-Y, Geiger J, Ishchenko A, Han GW, Barty A, White TA, et al. Harnessing the power of an X-ray laser for serial crystallography of membrane proteins crystallized in lipidic cubic phase. IUCrJ 2020;7:976–84.

[69] Jespers W, Verdon G, Azuaje J, Majellaro M, Keränen H, García-Mera X, et al. X-ray crystallography and free energy calculations reveal the binding mechanism of A_{2A} adenosine receptor antagonists. Angew Chem Int Ed 2020;59:16536–43.

[70] Zhang K, Zhang J, Gao Z-G, Zhang D, Zhu L, Han GW, et al. Structure of the human $P2Y_{12}$ receptor in complex with an antithrombotic drug. Nature 2014;509:115–8.

[71] Amelia T, van Veldhoven JP, Falsini M, Liu R, Heitman LH, van Westen GJ, et al. Crystal structure and subsequent ligand design of a nonriboside partial agonist bound to the adenosine A_{2A} receptor. J Medi Chem 2021;64:3827–42.

[72] Zhang J, Zhang K, Gao Z-G, Paoletta S, Zhang D, Han GW, et al. Agonist-bound structure of the human $P2Y_{12}$ receptor. Nature 2014;509:119–22.

[73] Cui M, Zhou Q, Xu Y, Weng Y, Yao D, Zhao S, et al. Crystal structure of a constitutive active mutant of adenosine A_{2A} receptor. IUCrJ. 2022;9:333–41.

[74] Claff T, Klapschinski TA, Tiruttani Subhramanyam UK, Vaaßen VJ, Schlegel JG, Vielmuth C, et al. Single stabilizing point mutation enables high-resolution co-crystal structures of the adenosine A_{2A} receptor with preladenant conjugates. Angew Chem Inte Ed 2022;61:e202115545.

[75] Shiriaeva A, Park D, Kim G, Lee Y, Hou X, Jarhad DB, et al. GPCR agonist-to-antagonist conversion: enabling the design of nucleoside functional switches for the A_{2A} adenosine receptor. J Med Chem 2022;65:11648–57.

[76] Neumann A, Muller CE, Namasivayam V. $P2Y_1$-like nucleotide receptors-Structures, molecular modeling, mutagenesis, and oligomerization. Wires Comput Mol Sci 2020;10:e1464.

[77] Jespers W, Schiedel AC, Heitman LH, Cooke RM, Kleene L, van Westen GJP, et al. Structural mapping of adenosine receptor mutations: ligand binding and signaling mechanisms. Trends Pharmacol Sci 2018;39:75–89.

[78] Jacobson KA, Delicado EG, Gachet C, Kennedy C, von Kugelgen I, Li B, et al. Update of P2Y receptor pharmacology: IUPHAR Review 27. Br J Pharmacol 2020;177:2413–33.