Ph. D. Thesis

State-selective stabilization of adenosine A_{2A} receptor by de novo design of protein structures

Department of Structural Molecular Science
School of Physical Sciences
The Graduate University for Advanced Studies, SOKENDAI

Masaya Mitsumoto

January 7, 2022

Table of Contents

Abstract

Computational design technology to accurately create protein structures de novo with high thermal stability has greatly advanced in this decade. Based on the technology, I aimed to develop computational methods to stabilize proteins in a selected state.

In this thesis, I targeted on G-protein coupled receptors (GPCRs) for state-selective stabilization. GPCRs are the largest membrane protein family encoded by the human genome. Canonically, GPCRs are activated upon the binding of extracellular ligands, which induces a conformational change from the inactive state to the active state, leading to intracellular coupling of G-proteins that trigger downstream biochemical cascades. Because GPCRs control diverse physiological functions, they have been of major scientific interest and also among the main drug targets. However, despite their scientific and pharmacological importance, the innate instability of GPCRs has been problematic for sample preparation and functional assay. Moreover, GPCRs are in equilibrium between inactive and active states that exhibit large conformational changes upon state transitions. The low homogeneity caused by this feature has been problematic for structure determination and state-specific ligand/antibody screening. In this thesis, I aimed to develop methods that not only stabilize GPCRs but also stabilize them in a specific state.

One of the causes of GPCR instability is the intracellular loop 3 (ICL3), an intracellular loop with varying lengths among different GPCRs that shows high structural flexibility. In addition, ICL3 connects the two transmembrane helices, transmembrane helix 5 (TM5) and transmembrane helix 6 (TM6), which move outward upon the conformational change from the inactive to active state (the agonist-bound G-protein coupled state). I hypothesized that by redesigning the TM5-ICL3-TM6 region to stabilize TM5 and TM6 in either the inactive or active state conformation, state-selective stabilization of GPCRs might be achieved.

I applied two approaches for state-selective stabilization of GPCRs: 1) designing stable TM5-ICL3-TM6 region by redesigning ICL3 to be short and typical structures, and 2) designing extremely stable all α-helical proteins that are made to fix the conformation of TM5 and TM6 by being replaced to ICL3. In this study, I targeted adenosine A_{2A} receptor ($A_{2A}R$) and attempted to stabilize it in either the inactive or active state. $A_{2A}R$ is a prototypical GPCR belonging to class A that constitutes the largest GPCR subfamily covering approximately 90% of human GPCRs.

My thesis consists of five chapters: Chapter 1 provides a brief overview of GPCRs and protein design, Chapter 2 describes the redesign of ICL3 for state-selective stabilization of $A_{2A}R$, Chapter 3 and 4 focus on the de novo design of fusion partner proteins customized to stabilize the inactive and active states of $A_{2A}R$, respectively, and Chapter 5 summarizes the conclusions of the thesis.

In Chapter 2, the computational strategy for redesigning TM5-ICL3-TM6 region and experimental results of the redesigned $A_{2A}Rs$ are described. One reason that can explain the exceptionally high thermal stability of computationally de novo designed proteins is the use of short and typical loop structures. Therefore, I redesigned the TM5-ICL3-TM6 region of the $A_{2A}R$ to have short and typical loop structures. For state-selective stabilization, I redesigned the region of the inactive state structure and active state structure of $A_{2A}R$, respectively. First, I computationally rebuilt ICL3 between TM5 to TM6 as a short and typical structure. Then, sequence design was performed around the cytosolic area of TM5-ICL3-TM6. The redesigned $A_{2A}Rs$ with new ICL3 structures, exhibiting the lowest energy, were selected for experimental characterization. The redesigned $A_{2A}Rs$ were experimentally characterized by Murata group at Chiba University, and the results suggest that the stability of the redesigned $A_{2A}Rs$ was not improved compared to that of the wild-type. This indicate more stabilization for the redesigned region is required.

To improve the method, I computationally de novo designed super-stable protein structures to be replaced to ICL3. In Chapter 3, the computational strategy for the de novo design of extremely stable all

α-helical proteins customized for fusion to $A_{2A}R$ and the experimental results are described. The fusion partner strategy, which replaces ICL3 with soluble protein domains (fusion partner proteins), has been widely employed for stabilizing GPCRs. However, this method requires numerous experimental trials and errors. I aimed to develop a computational method to design fusion partner proteins that stabilize GPCRs in a specific state. First, I prepared a set of hundreds of all α-protein backbone models, and selected the backbone models that can be fused into $A_{2A}R$ through straight helical connections. Next, I designed amino acid sequences for the selected backbones. Then, I performed experimental characterizations of the designed proteins. The designed fusion partner protein, named FiX1, folded into a monomeric structure with high thermal stability (the melting temperature is more than 98 ˚C), and the chimeric $A_{2A}R$ with FiX1 ($A_{2A}R$–FiX1) exhibited higher thermal stability than the wild-type $A_{2A}R$. The ligand-binding affinity of $A_{2A}R$–FiX1 to the inverse agonist was similar to that of the wild-type $A_{2A}R$, while the affinity to the agonist was drastically reduced. These results indicate that the rational stabilization of the $A_{2A}R$ inactive conformational state was successful.

To investigate the potential of the developed method, in Chapter 4, I computationally de novo designed super-stable fusion partner proteins to stabilize $A_{2A}R$ in an active state. The de novo designed fusion partner proteins showed extreme thermostability, and the chimeric $A_{2A}R$ with a fusion partner protein, $A_{2A}R$–FaX3, exhibited higher stability than the wild-type $A_{2A}R$. However, $A_{2A}R$–FaX3 did not show ligand-binding affinities against the inverse agonist and agonist. These results indicate that $A_{2A}R$–FaX3 was stabilized in an unexpected state other than the inactive or active state. For stabilizing GPCRs in an active state, the TM5 and TM6 region needs to be more stabilized. The experimental characterization of the chimeric $A_{2A}R$s described in Chapters 3 and 4 was performed by Murata group at Chiba University.

I succeeded in stabilizing the inactive state of $A_{2A}R$ by the mainchain-level protein engineering: custom-made design of super-stable fusion partner proteins. The developed method still needs to be

improved to stabilize an active state. However, the method is expected to be a basic technology for designing rationally stabilized GPCRs in a desired state, which is thought to be applied in solving structures in specific states or state-dependent ligand/antibody screening. In addition, this is the first example of rational state-selective stabilization of proteins only by mainchain-level engineering, not limited in GPCRs. The method is expected to be utilized to stabilize other proteins in desired states.

Chapter 1: General introduction

G-protein coupled receptors

G-protein coupled receptors (GPCRs) are the largest membrane protein family in human genome[1]. Canonically, GPCRs receive extracellular ligands, such as neurotransmitters, hormones, autacoids and photons and transduce signals into cell [2].

There are approximately 800 GPCRs encoded by the human genome and they are classified into 5 subfamilies, class A, B1, B2, C and F [3, 4]. GPCRs of all the classes share transmembrane domain that consists of bundle of seven α-helices. Class A GPCR (Rhodopsin family) is the largest subfamily of GPCRs; 719 (≈ 90%) of the receptors belong to the class. Over 500 drugs act through class A GPCRs and the targeted diseases are diverse, including hypertension, depression, schizophrenia, glaucoma, migraine, etc. Class B1, B2, C and F receptors have extracellular domains at the N-terminal. Class B1 GPCRs (Secretin family) have 16 members and receptors belong to the subfamily bind peptides like glucagon, parathyroid hormone, calcitonin gene-related hormone, etc. Class B2 GPCRs (Adhesion family) consist of 33 receptors. Receptors in the class cleave the N-terminal domain as a fragment through autoproteolytic reaction. Class C GPCRs (Glutamate family) include 22 receptors. The class C GPCRs are constitutively dimerized through the large extracellular domain and the conformational change occur via binding of ligands to the large extracellular domain, not the transmembrane domain. Class F GPCRs (Frizzled family) comprise one smoothened receptor and ten Frizzled receptors. The receptors have conserved cysteine-rich domain as the extracellular cellular domain.[5] (**Figure 1-1**) In this thesis, because class A is the largest subfamily of GPCRs and there are lots of available structural and experimental information, I targeted on class A GPCRs.

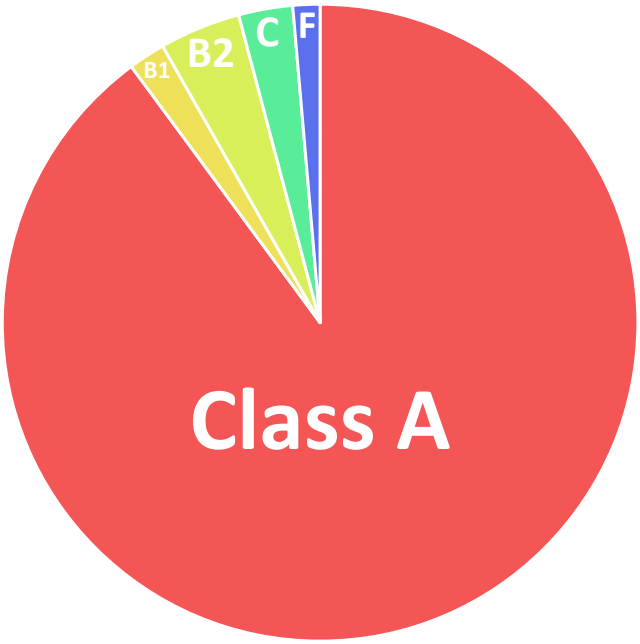

Figure 1-1. Pie chart for ratio of GPCRs of each class: Among 800 GPCRs (from ref.[3]), 719 receptors (89.875%) belong to class A (Rhodopsin family), 15 receptors (1.875%) belong to class B1 (Secretin family), 33 receptors (4.125%) belong to class B2 (Adhesion family), 22 receptors (2.75%) belong to class C (Glutamate family) and 11 receptors (1.375%) belong to class F (Frizzled family).

The canonical mechanism of signal transduction of GPCRs occurs as follows: 1) Extracellular ligand, (e.g. hormone, neurotransmitter, photon) is bound to GPCR and the conformation is altered for binding to G-proteins. 2) Intracellular hetero-trimeric G-protein complex which consists of GDP-bound Gα and Gβγ, obligated complex of Gβ and Gγ, is bound to the intracellular binding site of GPCR. Upon binding, the GDP bound to Gα is released and exchanged to GTP and the trimeric complex is dissociated to GTP-bound Gα and Gβγ. 3) GTP-bound Gα and Gβγ stimulate different effector proteins independently (and GTP-bound Gα hydrolyzes the bound GTP into GDP and the GDP-bound Gα associates with Gβγ and hetero-trimeric G-protein complex is formed again) (**Figure 1-2**) [2].

The downstream effector proteins stimulated by GTP-bound Gα and Gβγ cause different cellular responses. Gα are classified into four families: $G\alpha_s$, $G\alpha_i$, $G\alpha_q$ and $G\alpha_{12}$. $G\alpha_s$ activates adenylyl cyclase, an effector protein which converts ATP into cyclic AMP (cAMP), and increases cellular concentration of cAMP that triggers activation of cAMP-regulated proteins: protein kinase A for example. On the other hand, $G\alpha_i$ deactivates adenylyl cyclase, resulting in decrease of cellular cAMP-level. $G\alpha_q$ activates phospholipase C which converts phosphatidylinositol 4,5-biphosphate into 1,4,5-triphosphate (IP_3) and membrane-bound diacylglycerol (DAG). IP_3 alters IP_3 receptor, a calcium channel on the endoplasmic reticulum (ER) membrane, to the open state and triggers Ca^{2+} release from ER. In parallel, DAG activates protein kinase C (PKC). $G\alpha_{12}$ family activates p115 RhoGEF and related RhoGEF proteins.[6] There are five Gβ genes and twelve Gγ genes, however, independent to the combinations, Gβγ expresses similar activity. The effector proteins activated by Gβγ are diverse: the complex regulates ion channels, adenylyl cyclase, phospholipase C, etc.[7]

The activated biochemical cascades cause various physiological functions, depending on the cells that the biochemical cascades are activated. GPCRs are related with virtually every aspects of human physiological functions and due to the number of related diseases and the easy accessibility of drugs to

the extracellular ligand binding site, they are one of the major drug targets [5]. In fact, approximately one-third of FDA-approved small molecule drugs target GPCRs [8].

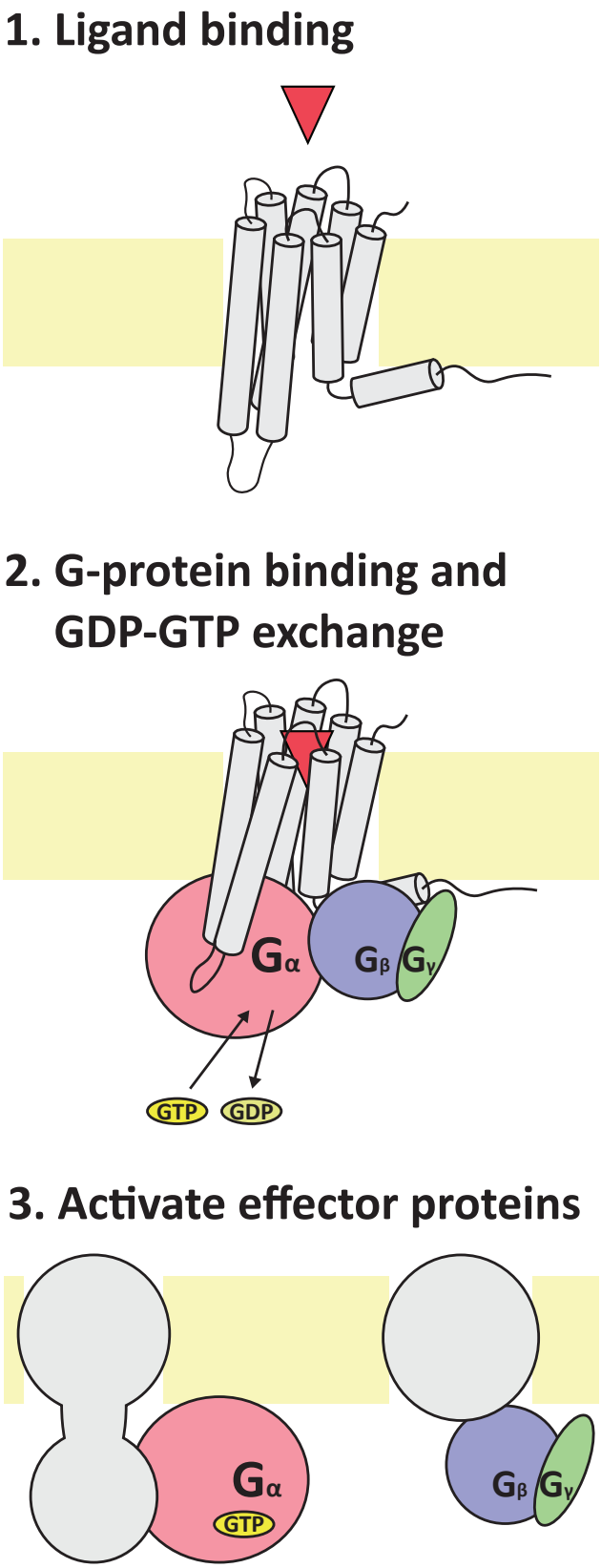

Figure 1-2. Canonical flow of signal transduction via G-protein coupled receptors.

Stabilization of GPCRs

GPCRs are of major scientific interest and among important drug targets, however, because GPCRs are innately unstable, the instability and low homogeneity has caused problems in sample preparation, functional assay and structure determination. To solve these problems, stabilization of GPCRs by protein engineering has been employed [9]. The stabilization strategies used for GPCRs are roughly classified into two types: sidechain-level and mainchain-level engineering.

In the sidechain-level engineering, thermostabilizing mutations has been identified through systematic mutagenesis such as alanine/leucine scan and directed evolution [10, 11]. Thermostabilizing mutations found by these methods do not always stabilize other GPCRs when they are transferred at corresponding residue positions of other GPCRs [9]. Therefore, systematic mutations are still used to generate stabilized GPCRs [12-16]. To rationally stabilize GPCRs, some computational methods to predict thermostabilizing mutations have been developed [17-19].

In the mainchain-level engineering, the N- and C-terminal regions are usually truncated since they are thought to be flexible and thus cause low expression and solubilization of GPCRs [20]. Moreover, GPCRs have another region which is thought to be the major origin of the instability: the intracellular loop 3 (ICL3), the loop connecting transmembrane helix 5 and 6. This ICL3 has variety of length depending on the receptor type and thought as the source of conformational flexibility. Fusing soluble protein domains, so called fusion partner proteins, removing most part of ICL3 is one of the major stabilization methods [21]. The first human GPCR structure, β_2AR was solved using this method with T4-lysozyme [22, 23] as its fusion partner protein and so many GPCR structures have been solved with fusion partner proteins such as apocytochrome b_{562}RIL (BRIL) [21, 24], rubredoxin [25] and glycogen synthetase [26]. The effects of chimerization with fusion partner proteins have not been computationally predicted, therefore the proteins used as fusion partner and the positions to fuse have been identified

through experimental trial and error. In this thesis, I attempted to develop methods to rationally stabilize GPCRs by mainchain-level engineering.

State-selective stabilization of GPCRs

GPCRs alter their conformations upon agonist binding and G-proteins [2] or other intracellular transducer protein coupling (such as arrestin and GPCR kinase [27]). Stabilizing a particular conformational state of GPCRs has been undergone toward solving experimental structures in a particular state [11, 28]. In addition, because GPCRs activate downstream signaling cascades depending on the transducer proteins, drugs that stabilize a particular state of GPCRs to trigger specific signaling cascade over others (biased agonists) are developed [5]. GPCR samples stabilized in a particular state is expected to be useful for screening ligands or antibodies.

Methods to stabilize GPCRs in a particular state have been developed. In both the sidechain-level engineering and mainchain-level engineering, using mutations and fusion partner strategy, respectively, engineered GPCRs stabilized in a particular state were identified by evaluating the stability of the GPCRs in presence of ligands which stabilize a specific state [10, 11, 21]. This identification process is based on experimental trial and error.

Rational methods to stabilize GPCRs in a particular state have been developed. For sidechain-level engineering, methods using computational prediction [19] and knowledge-based prediction [29] have been developed. However, for mainchain-level engineering, except one example combined with mutations [30], there are no report of methods to rationally stabilize GPCRs in a particular state. In this thesis, I attempted to stabilize GPCRs in a particular state, only by rationally engineering the mainchain. In the next section, based on activation mechanisms of GPCRs, I hypothesized on which part of mainchain to engineer.

Mechanism of conformational change of class A GPCRs

For rationally stabilizing GPCR structures in desired states, understanding the mechanism of conformational change is important. The outstanding difference between inactive and active state structures is the transmembrane helix 6 (TM6). In the inactive state structures, TM6 contacts with other transmembrane helices, but in the active state structure, the TM6 is largely moved outward for G-protein coupling (**Figure 1-3**). By comparing the residue-residue contacts of inactive state and active state structures of class A GPCRs, common activation mechanism of class A GPCR is proposed [29]. For labeling the consensus residue positions among class A GPCRs, GPCRdb numbering is employed. In this numbering scheme, the first integer denotes transmembrane helix (TM1-7) and the second integer denotes relative residue number to the most conserved residue position when the most conserved residue position number among belonging transmembrane helix is set to 50. For example, 6×48 indicates a residue position in TM6 (from the first integer, 6) which two residues before the most conserved residue among the helix (from the second integer, 48; 50-2). Receptor activation is thought to occur by the following four layers (**Fig 1-4**). (**Layer 1: Signal initiation**) Ligand binding alters the intrahelical contacts around two residues, $W^{6\times48}$ and $F^{6\times44}$, called toggle switch residue and transmission switch residue, at the bottom of ligand included in the two conserved motifs: CWxP motif and PIF motif, respectively. This rearrangement initiates rotation of cytoplasmic end of TM6 and another conserved motif, called Na^+ pocket, the sodium ion binding site is collapsed and the four residues including $D^{2\times50}$, consisting the motif make denser packing and Na^+ is released from $D^{2\times50}$. (**Layer 2: Signal propagation**) Contacts of the three hydrophobic residues in the middle of transmembrane helices, so called hydrophobic lock is collapsed and packing between TM3 and TM6 is loosened. (**Layer 3: Microswitch rewiring**) The two residues $Y^{7\times53}$ in NPxxY motif and hydrophobic residue at 6×37, called microswitch residues are rewired. $Y^{7\times53}$ packed with two residues on TM1 and the eighth helix at C-terminal loses the contact and forms the three residues including $R^{3\times50}$ on TM3 which are originally packed against TM6.

Via this repacking process, the interaction between TM3 and TM7 is enhanced while that between TM3 and TM6 is weakened. (**Layer 4: G-protein coupling**) The electrostatic interaction called ionic lock, composed of the three residues D/E$^{6\times30}$, D/E$^{3\times49}$ and R$^{3\times50}$, where the latter two residues in conserved DRY motif is canceled so that outward movement of TM6 is enabled to couple with G-proteins. Finally, intracellular G-protein coupling cause large outward movement of TM6.

Starting from agonist binding, the effects propagate and result in large outward movement of TM6. Along with TM6, transmembrane helix 5 (TM5) also shows conformational differences in the inactive and active state structures. TM5 and TM6 are connected by ICL3, the flexible loop which is thought to be the cause of intrinsic instability of GPCRs and the angle of TM5 and TM6 makes differences of conformational states of GPCRs. I hypothesized that by redesigning TM5-ICL3-TM6 region to be stabilized in the angle of desired conformational state, rational state-selective stabilization might be achieved. To achieve this, I utilized computational de novo protein design technology.

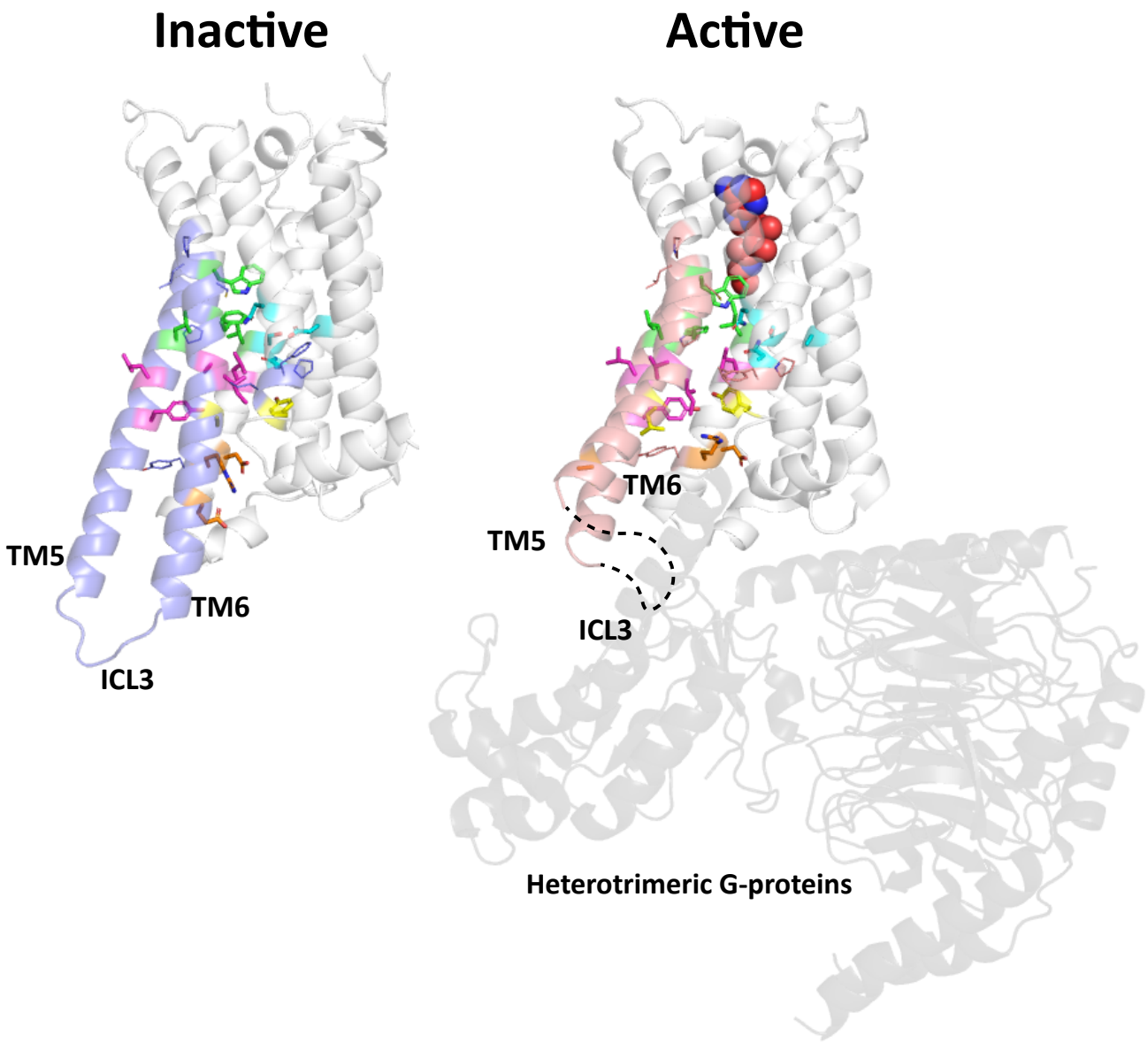

Figure 1-3. Comparison of class A GPCR structures of inactive state and active state. (Left) $A_{2A}R$ structure of the inactive state (PDB: 3PWH). (Right) $A_{2A}R$ structure of the active state (PDB: 6GDG). The missing model of ICL3 is represented by dotted line. Amino acid residues whose sidechains are shown as stick are included in residue clusters that mainly move upon the activation. Residue clusters colored in green, sky blue, magenta, yellow and orange are residues around toggle switch and transmission switch, Na^+ pocket, hydrophobic lock and ionic lock, respectively. Sidechains shown in line are amino acid residues contained in conserved motifs. These residues and TM5 and TM6 region are colored in blue in the inactive structure and pink in the active structure. The ligand shown as sphere is agonist NECA. Heterotrimeric G-protein is transparently shown in black.

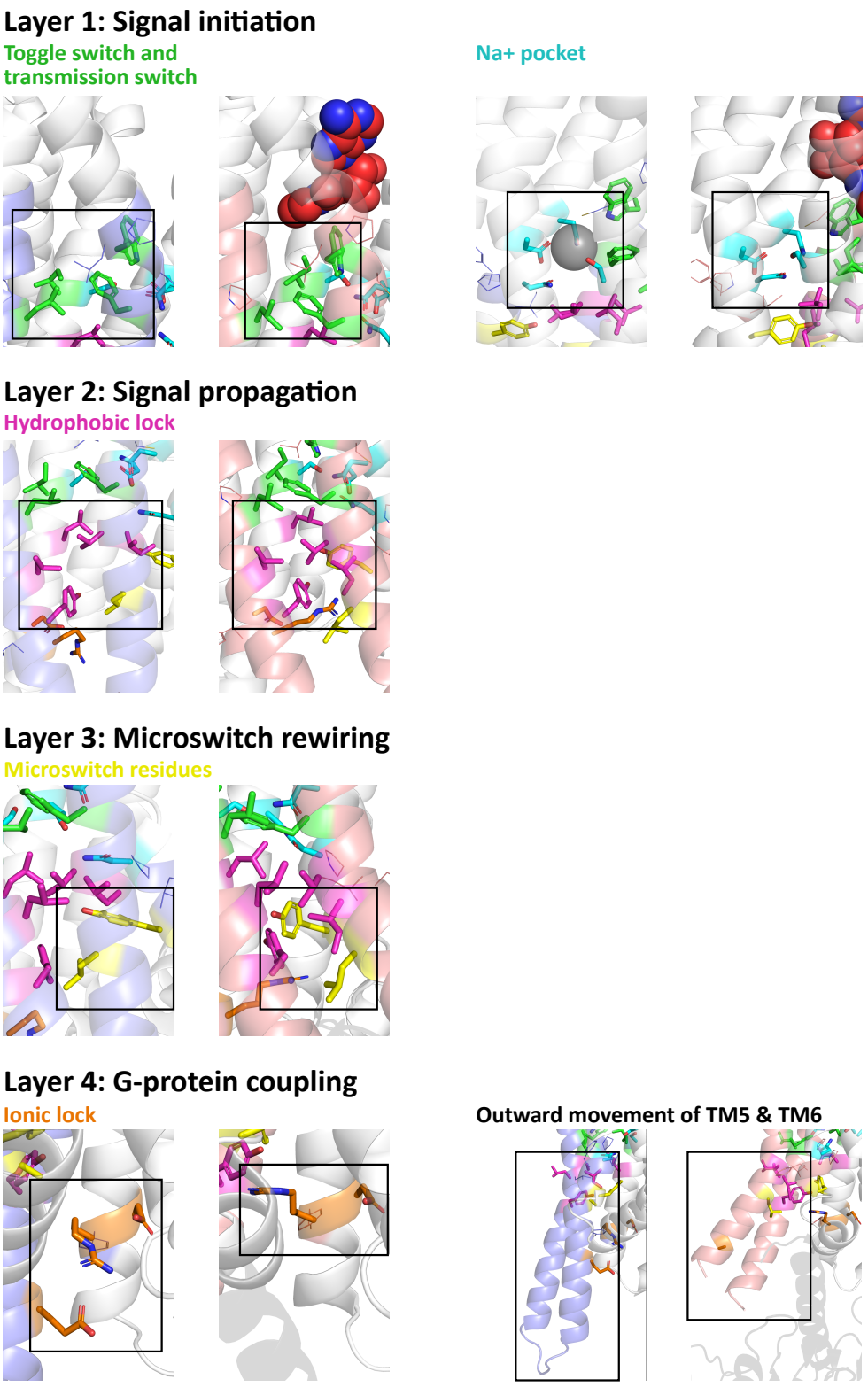

Figure 1-4. Comparison of residue clusters or region that move upon activation of class A GPCRs. Images shown in left are from inactive-state structures and these shown in right are from active-state structures.

Computational protein design

Computational de novo protein design is the technology to generate proteins with intended structures as designed *in silico*. The first example of computational protein design was reported by Dahiyat and Mayo in 1997 [31]. Using an algorithm to place appropriate rotamers of amino acids based on scoring functions composed of physics-based terms, they designed amino acid sequence on a 28 residue-long backbone of ββα motif from a naturally occurring protein that fold into the shape of the motif with only ≈ 20% sequence identity. The first de novo protein design from the mainchain to the sidechain was reported in 1998 by Harbury et al.[32] They de novo designed tetrameric right-handed coiled coil whose fold was not found in nature to the date. In 2003, the first de novo designed globular protein structure was reported by Kuhlman et al.[33] They generated a backbone structure of a novel α+β-fold from scratch *in silico* and designed the sequence on the backbone by Rosetta, a protein structure prediction software they developed. The next de novo designed globular proteins was reported in 2012 by Koga et al.[34] They found the relationships between loop length and the strongly favored direction to the adjacent secondary structure element, independent to the amino acid sequence. They utilized the found principle for de novo design of αβ-fold protein structures by implicitly disfavoring possible other conformations than the design structure. Since the success by Koga et al., in this decade, computational de novo protein design has greatly progressed. Computationally de novo designed protein structures to date are diverse: TIM-barrel, diverse helix-loop-helix repeat proteins, bulge-containing α+β-fold protein, macrocycles, β-barrel, jerry-roll fold, etc.[35-41]

These designed protein structures are generally thermally stable [35]. The computational algorisms and scoring functions developed for protein design has been utilized for protein stabilization. For sidechain-level engineering, the algorisms have been utilized to predict thermostabilizing mutations [42, 43]. For mainchain-level engineering, methods are limited in idealizing backbones around parts of statically functioning active sites [44-46]. In this thesis, I attempted to apply protein design to regulate

dynamics of proteins by mainchain-level engineering: de novo design of protein structure superseded to unstable region for state-selective stabilization.

Brief description about this thesis

In this thesis, I aimed to rationally stabilize GPCR structures in a desired state, by mainchain-level protein engineering. In class A GPCRs, TM5 and TM6 show conformational differences between the inactive and active states. I hypothesized that by stabilizing TM5 and TM6 region to stabilized in a specific conformation, GPCRs might be stabilized in a desired state. I targeted on adenosine A_{2A} receptor ($A_{2A}R$), a prototypical class A GPCR for the state-selective stabilization in the inactive or active states. The developed two methods in this thesis are following. Chapter 2: designing stable TM5-ICL3-TM6 region by redesigning ICL3 to be short and typical structure, and Chapter 3 and 4: de novo design of super-stable all-α helical fusion partner protein customized to be fused into $A_{2A}R$ through TM5 and TM6 in the desired conformations.

Adenosine A_{2A} receptor

The target GPCR in this thesis is adenosine A_{2A} receptor ($A_{2A}R$). $A_{2A}R$ is one of adenosine receptors which are mediated by adenosine, a ubiquitous autacoid in human body. Binding with adenosine, the receptor alters the conformation to couple with heterotrimeric G-proteins, composed of $G\alpha_s$ and $G\beta\gamma$. The activated $G\alpha_s$ stimulate adenylyl cyclase to produce cAMPs that trigger various physiological functions depending on the location the cAMP level is increased. $A_{2A}R$ is related to physiological functions such as modulation of motor, vascular control and immunosuppression, therefore many agonists or antagonists for $A_{2A}R$ are being developed as drugs targeting diseases including Parkinson's disease, heart failure and cancers.[47]

Human $A_{2A}R$ is a well-studied receptor whose crystal structure was solved in 2008, following bovine Rhodopsin structure in 2000 [48], human β_2AR structure in 2007 [22, 23, 49] and turkey β_1AR structure in 2008 [50]. The structure was first solved using fusion partner strategy using T4-lysozyme in the inactive state with synthetic inverse agonist ZM241385 [51]. Later, in 2011, active-like intermediate state structures with agonist NECA with T4-lysozyme as fusion partner [52] and without fusion partner with thermostabilizing mutations [53]. The active state structures was solved with NECA and engineered $G\alpha_s$ in 2016 [54] and with heterotrimeric G-proteins in 2018 [55]. The structure with engineered $G\alpha_s$ is the second structure of a GPCR solved with G-proteins (the first structure is β_2AR in 2011 [56]). Because the number of the deposited structure in PDB is most abundant among human GPCRs (https://gpcrdb.org [57]) and the solved states include inactive, active-like intermediate and active state (agonist-bound G-protein coupling state), $A_{2A}R$ can be considered as the best receptor to test my proposed design methodologies.

Chapter 2: Redesign of intracellular loop 3

Introduction

In this chapter, I attempted to stabilize $A_{2A}R$ selectively either in an inactive or active state. In contrast to unstable GPCRs, most of computationally de novo designed proteins exhibit extremely high thermal stability [58]. The extreme high stability is thought to be due to the use of short and typical loop structures in computationally designed proteins [59]. Therefore, I tried to redesign TM5-ICL3-TM6 region from its mainchain to have a short and typical loop structure for state-selective stabilization.

Results

Backbone design: Statistics of ABEGO torsion patterns for loops connecting α-helices

For redesigning ICL3 of $A_{2A}R$ to be a typical and short loop structure that favors a particular conformational state, I first attempted to figure out typical ABEGO (**Figure 2-1**) torsion patterns for loops connecting α-helix and α-helix (αα-unit) as Lin and Koga et al. did[60]. Using a non-redundant set of naturally occurring protein structures, I investigated statistics of the ABEGO torsion pattern of αα-unit. As results, ABEGO patterns frequently observed in nature for each loop length were identified (**Figure 2-2** and **Table 2-1**).

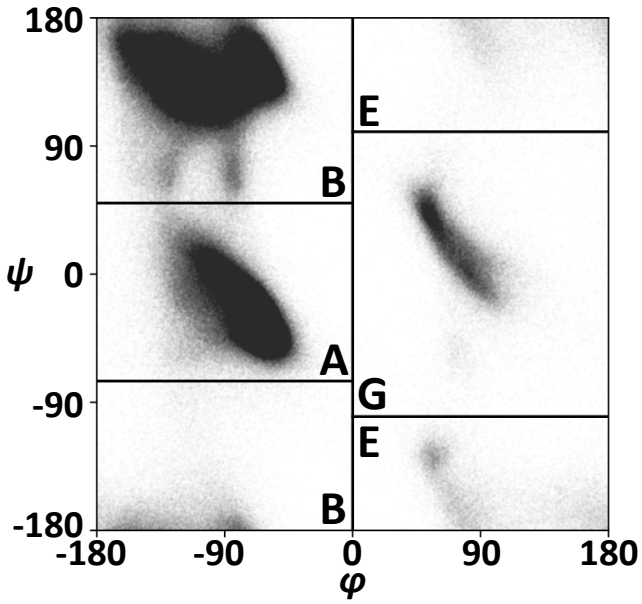

Figure 2-1. The scatter plot of dihedral angles φ and ψ (called Ramachandran plot) from non-redundant set of naturally occurring protein structures. The area A is for left-handed α-helical region, B is for left-handed β-strand region, G is for right-handed helical region and E is for right-handed β-strand region. (O is for cis-peptide.)

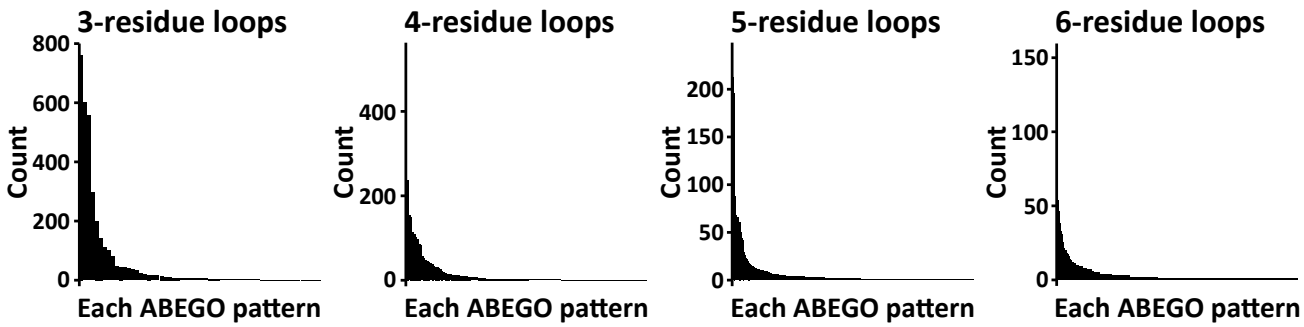

Figure 2-2. The histograms of count of each ABEGO pattern in αα-unit for 3, 4, 5 and 6 residues-long loops.

ABEGO pattern (3-residues)	Count	ABEGO pattern (4-residues)	Count	ABEGO pattern (5-residues)	Count	ABEGO pattern (6-residues)	Count
GBB	762	AGBB	536	BAABB	237	AGBBBB	152
BAB	600	BBBB	238	AGABB	212	ABAABB	71
AGB	556	BAAB	154	GBBBB	196	BBAABB	54
ABB	296	ABAB	151	BABBB	88	AAAABB	50
BBB	200	GABB	115	ABBBB	74	BABBBB	46

Table 2-1. Counts of each ABEGO pattern in αα-unit for 3, 4, 5 and 6 residues-long loops. The patterns within the fifth highest count are shown as typical ones.

Backbone design: Rebuilding ICL3 by identified typical loops

I computationally remodeled the ICL3 structures of $A_{2A}R$ using the inactive state (PDB: 3PWH) and active state (PDB: 5G53) structures. I performed loop remodeling between cytosolic area of TM5 and TM6 using RosettaRemodel [61]. For every combinations of a starting residue in cytosolic area of TM5 and a terminal residue in cytosolic area of TM6, the loop remodeling simulations were performed for 1,000 times independently using the loop lengths from three to six. After the loop remodeling, I investigated the combination of loop length and the residue positions that were able to rebuild ICL3 in high frequency, by counting the numbers of trials that were able to rebuilt ICL3 and subsequently analyzing the ABEGO patterns of the rebuilt ICL3 structures.

Among the structures generated by combinations that closed loops using typical ABEGO patterns within top five frequency, I selected the structures which remain the longest helical residues after loop rebuilding as template structures to carry out sequence design (**Table 2-2** for the inactive structure and **Table 2-3** for the active structure). The selected templates for the inactive structures with ABEGO patterns of BBBB and BABBB were named as Inact-BBBB and Inact-BABBB, respectively. The selected template for the active structure with ABEGO types of AGBB was named as Act-AGBB.

For state-selective stabilization, it is desirable for the designed loop structures to stabilize the target state while destabilizing the other states. I investigated the loop closing ability by comparing the number of computational loop rebuilding trials which closed loops for inactive-state structure and active-state structure with the same combination of the starting residue, terminal residue and loop length to the selected templates (**Figure 2-3**). In the case for corresponding combinations to Inact-BBBB and Act-AGBB, only one trial and no trials, respectively closed loop using $A_{2A}R$ of the other state. The corresponding trials to Inact-BABBB closed loops less frequently in the active-state structure than in the inactive-state structure. From these results, the selected templates are expected to form short and typical loop structure in the targeting state while it is harder to form short loop structure in the other state.

Accept	Starting residue	Terminal residue	Loop length	Number of generated models	Most frequently observed ABEGO pattern	Number of generated models with most frequently observed ABEGO pattern
	204	229	3	957	BBB	953
	207	222	4	881	BBBB	768
✓	211	222	4	734	BBBB	586
✓	211	222	5	759	BABBB	511
	207	226	4	453	BBBA	320

Table 2-2. Simulation results for loop remodeling for inactive $A_{2A}R$ structure. The combinations of starting residue, terminal residue and loop length of the remodeled $A_{2A}R$ structures and the number of generated models, the most frequently generated ABEGO pattern of the remodeled ICL3 and the number of generated models with most frequently observed ABEGO pattern are listed. List is sorted by the number of generated models with most frequently observed ABEGO patter (the combinations within the fifth highest count are shown). The combinations used for final template models for sequence design are marked in the "Accept" column.

Accept	Starting residue	Terminal residue	Loop length	Number of generated models	Most frequently observed ABEGO pattern	Number of generated models with most frequently observed ABEGO pattern
✓	211	224	4	702	AGBB	575
	208	224	5	747	AAABB	473
	209	228	4	413	GBBB	378
	205	228	3	562	EAB	347
	208	224	4	581	BABB	303

Table 2-3. Simulation results for loop remodeling for active $A_{2A}R$ structure. The combinations of starting residue, terminal residue and loop length of the remodeled $A_{2A}R$ structures and the number of generated models, the most frequently generated ABEGO pattern of the remodeled ICL3 and the number of generated models with most frequently observed ABEGO pattern are listed. List is sorted by the number of generated models with most frequently observed ABEGO pattern (the combinations within the fifth highest count are shown). The combination used for final template model for sequence design is marked in the "Accept" column.

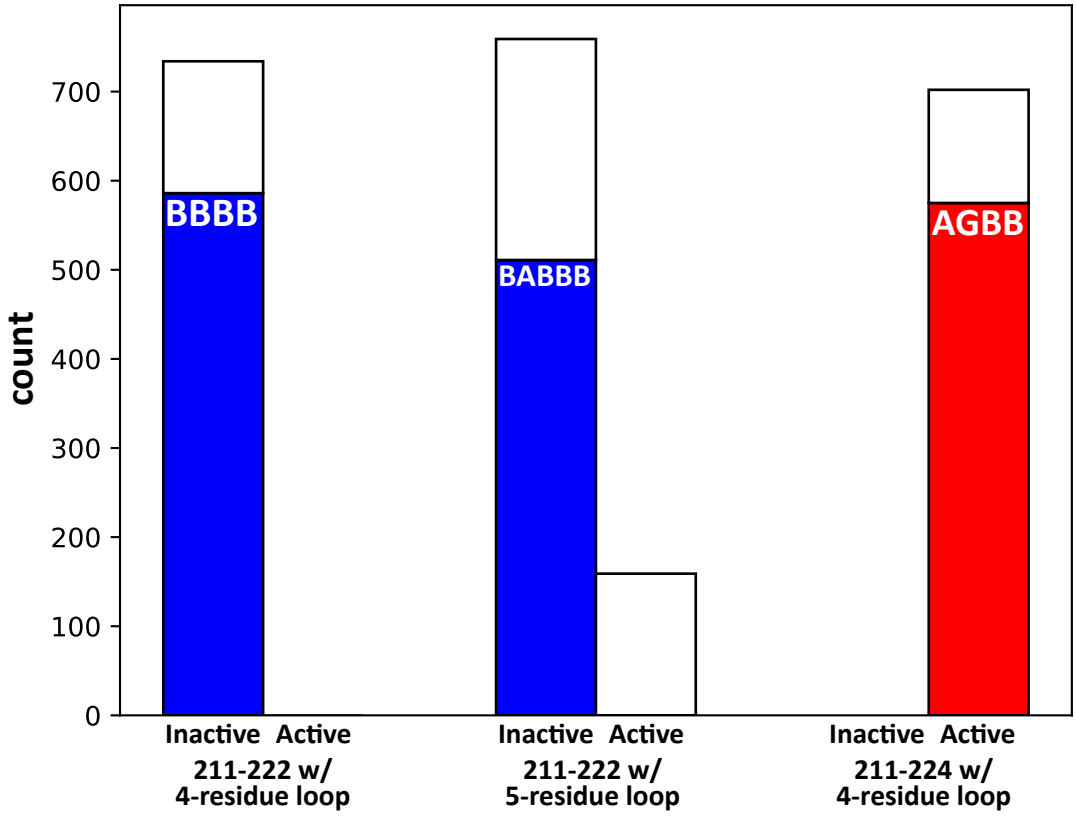

Figure 2-3. The numbers of the generated models using the inactive $A_{2A}R$ structure and the active $A_{2A}R$ structure with the same combination of starting residue, terminal residue and length of the loop to the template structures: Inact-BBBB, Inact-BABBB and Act-AGBB. The shares of the ABEGO patterns of these selected templates are colored in blue for these for inactive state and red for active state.

Sequence design

Sequence design was performed on the cytosolic area of TM5-ICL3-TM6 of Inact-BBBB, Inact-BABBB, and Act-AGBB, using Rosetta [33]. The design trials were performed for 1,000 times with restrictions of amino acids used for design depending on the environment of residue positions (see **Materials and Methods**). Structures whose Rosetta score is best among the individual run were selected for experimental validation (**Figure 2-4**). The selected redesigned $A_{2A}R$ structures on Inact-BBBB and Inact-BABBB were named Inact-SSSS and Inact-SSSSS, respectively. Three designs were selected for the designs on Act-AGBB and they were named Act-SSIS, Act-SGIS and Act-SGIP, respectively. The meaning of alphabets I, S, P and G at the end of the designs are for **I**nsoluble (hydrophobic), **S**oluble (hydrophilic), **P**roline and **G**lycine, respectively. Based on the buriedness of the residue positions on the rebuilt ICL3, limitations of amino acids used for design were set for the patterns of amino acids with these characteristics (named ISPG pattern, see **Materials and Methods**).

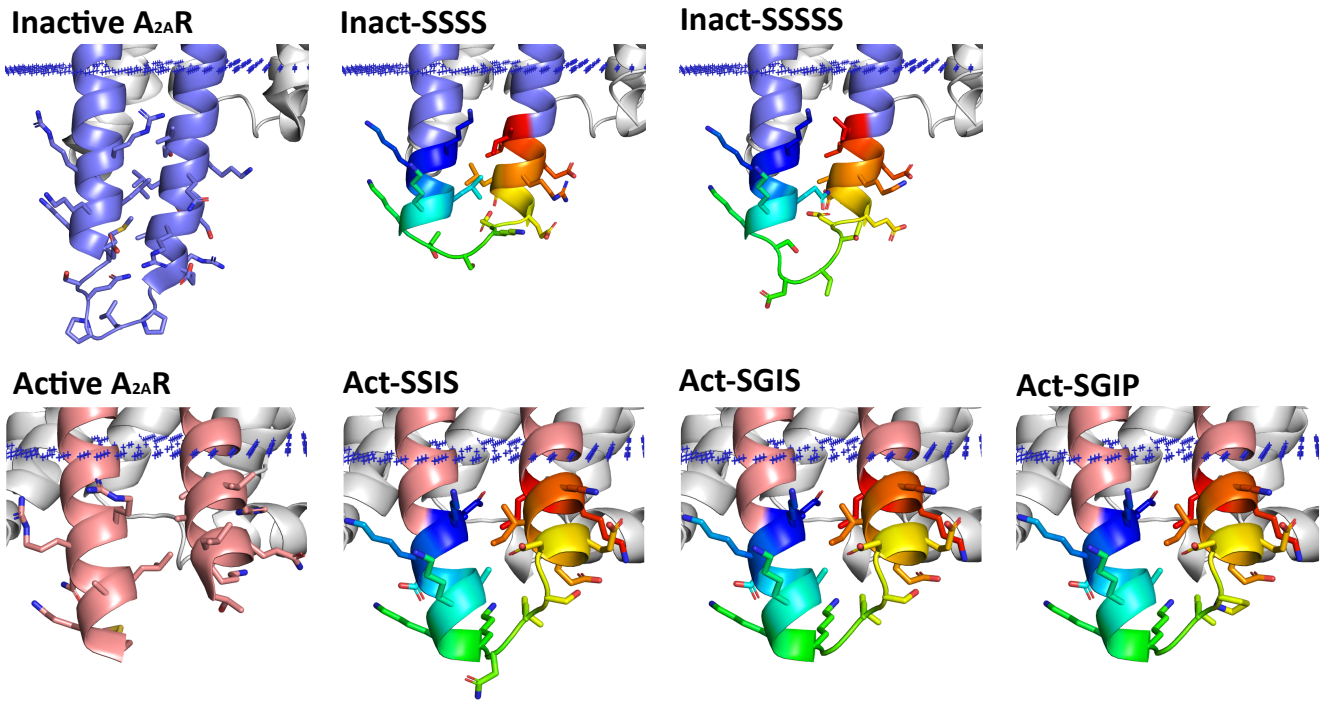

Figure 2-4. Comparison of the original $A_{2A}R$ structures (PDB: 3PWH for structure of the inactive state and PDB: 5G53 for structure of the active state) and the design structure models. For the design structures, sidechains are shown in the area that sequence design were performed and for original $A_{2A}R$ structures, sidechains in the corresponding area are shown.

Solubilization efficiency of redesigned A_{2A}Rs

Redesigned A_{2A}Rs were expressed with tagRFP at the C-terminal. As one of the stability metrices, solubilization efficiency of the redesigned A_{2A}Rs upon detergent extraction (the ratio of the fluorescent intensity after solubilization using n-dodecyl-β-D-maltoside (DDM) over that before solubilization) was evaluated. These results suggest that the redesigned A_{2A}Rs are not stabilized compared to the wild-type A_{2A}R (**Table 2-4**). This experiment was performed by Murata group at Chiba university.

Construct	Solubilization efficiency (%)
A_{2A}R WT	64.7
Inact-SSSS	54.6
Act-SGIS	38.5

Construct	Solubilization efficiency (%)
A_{2A}R WT	56.0
Inact-SSSSS	41.5
Act-SSIS	30.6
Act-SGIP	25.3

Table 2-4. The solubilization efficiencies of A_{2A}R WT and the redesigned A_{2A}Rs. The top data and the bottom data were taken by independent experimental sets.

Discussion

In this chapter, based on computational protein design methodology, I attempted state-selective stabilization of $A_{2A}R$ by redesigning TM5-ICL3-TM6 regions. However, experimental results suggest that the redesigned $A_{2A}Rs$ were not stabilized compared to the wild-type.

For Inact-SSSS and Inact-SSSSS, the redesigned $A_{2A}Rs$ to stabilize the inactive state, possible reasons of the insufficient stability are following.

1) Loss of helical region and hydrophobic interaction between TM5 and TM6. Through redesigning $A_{2A}R$, several residues in the original $A_{2A}R$ structure were lost; two residues in TM5 and four residues in TM6, forming helices in the original structure. In addition, the lost Met 211 and Arg 222 make hydrophobic contacts in the original structure (**Figure 2-5**). Removing these regions may destabilize $A_{2A}R$ structure and the redesigned TM5-ICL3-TM6 regions are thought not to stable enough to overcome the destabilization effect. Designing TM5-ICL3-TM6 region with much more stability or without losing original helices or residues are possible solutions for further stabilization.

2) Designing intended structures of TM5-ICL3-TM6 region was not achieved. To investigate this possibility, the sequence-structure compatibilities of the redesigned TM5-ICL3-TM6 region were evaluated for each designed $A_{2A}R$ models. First, every fragment of nine-residue window of the designed areas were obtained. Second, for each fragment, 200 nine-residue peptide fragments from non-redundant set of naturally occurring proteins were obtained based on the similarity in the secondary structure and sequence. For all nine-residue fragments of the redesigned TM5-ICL3-TM6 region, Cα RMSDs of the obtained fragments are calculated and plotted. Results for Inact-SSSS and Inact-SSSSS indicate that in naturally occurring protein fragments, there are only few fragments whose Cα RMSD are within 2.0 Å, for the area that mainly covered with redesigned loop region; 209-217 (**Figure 2-6A**). In the case that 200 nine-residue peptide fragments were obtained based only on the similarity of the secondary structure without using sequence similarity information, there are more fragments that are similar to the structure

of the redesigned region. This indicates that the backbone structure is typical, however, the designed sequence was not optimal to form the backbone conformation (**Figure 2-6B**). The most frequently found ISPG patterns for the backbones of Inact-BBBB is ISIS and that of Inact-BABBB is SSSIS; not the patterns used in this study (**Table 2-5**, see **Materials and Methods**). Using template backbones that the rebuilt ICL3 in environment whose assigned ISPG patterns are more frequently observed is possible solution for further stabilization.

The observed insufficient stability for Act-SSIS, Act-SGIS and Act-SGIP was not explained by the same reasons that were described for Inact-SSSS and Inact-SSSSS. Act-SSIS, Act-SGIS and Act-SGIP do not lose most of helical regions and residue contacts between TM5 and TM6 in the original structure (**Figure 2-7**). Furthermore, the sequence-structure compatibility of Act-SSIS, Act-SGIS and Act-SGIP are good (**Figure 2-8**). The reason of the insufficient stability should be derived from other factors. The original $A_{2A}R$ structure of the active state contains an engineered $G\alpha_s$, however, the redesigned $A_{2A}R$ structures are designed to be stabilized in the conformation without G-protein coupling. The active state of GPCRs are stabilized not only by agonist binding but also by G-protein coupling. Compensating for the absence of the G-protein by further increasing the stability of the TM5-ICL3-TM6 region seems to be a possible strategy to improve the method.

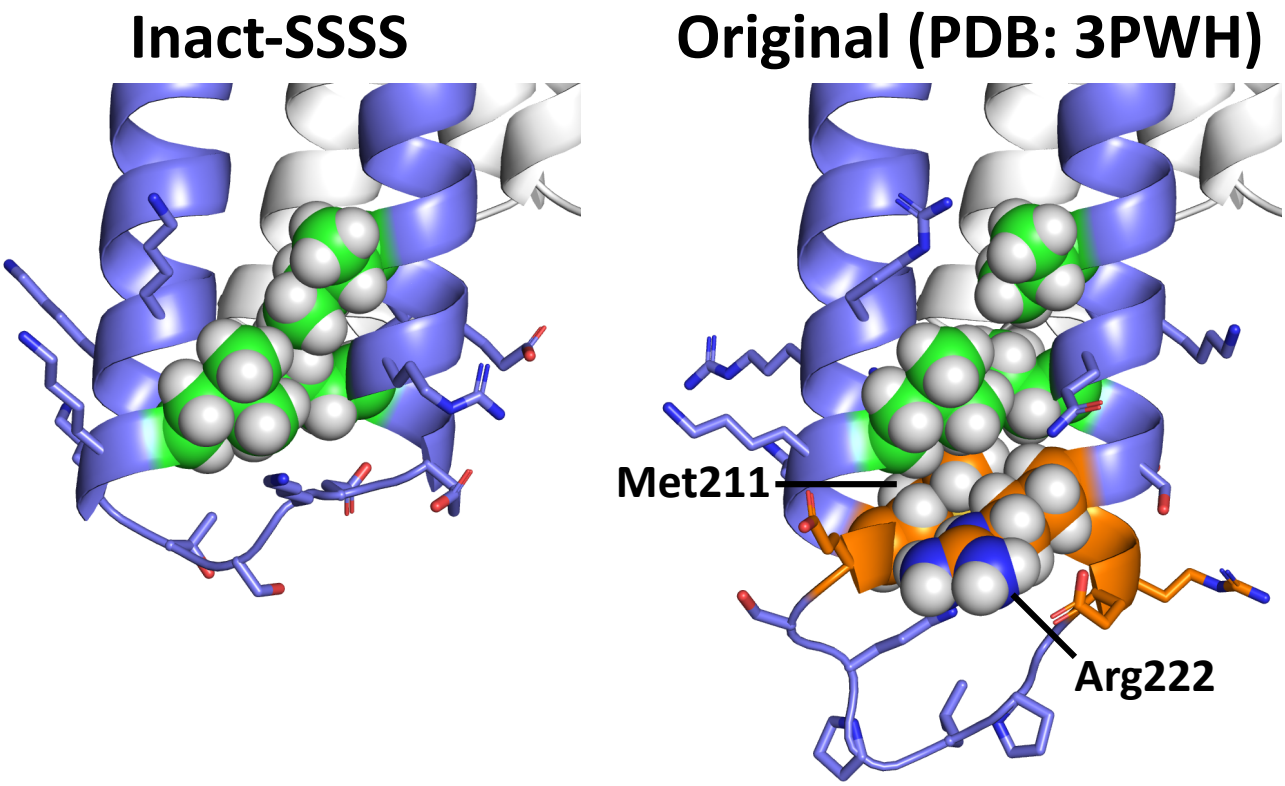

Figure 2-5. (Left) Computational model of Inact-SSSS. The sidechains of amino acids in the designed area are shown. Residues that form hydrophobic core packing are shown in spheres in green and the rest are shown in sticks. (Right) Original structure (PDB: 3PWH). The sidechains of amino acids in the designed area in Inact-SSSS and the corresponding area in the original structure are shown. Residues that form core packing in Inact-SSSS and residues at the corresponding position in the original structure are colored in green. Residues that form extra helices and hydrophobic core packing in TM5 and TM6, compared to Inact-SSSS are colored in orange. Other residue positions belonging to TM5-ICL3-TM6 region are colored in blue.

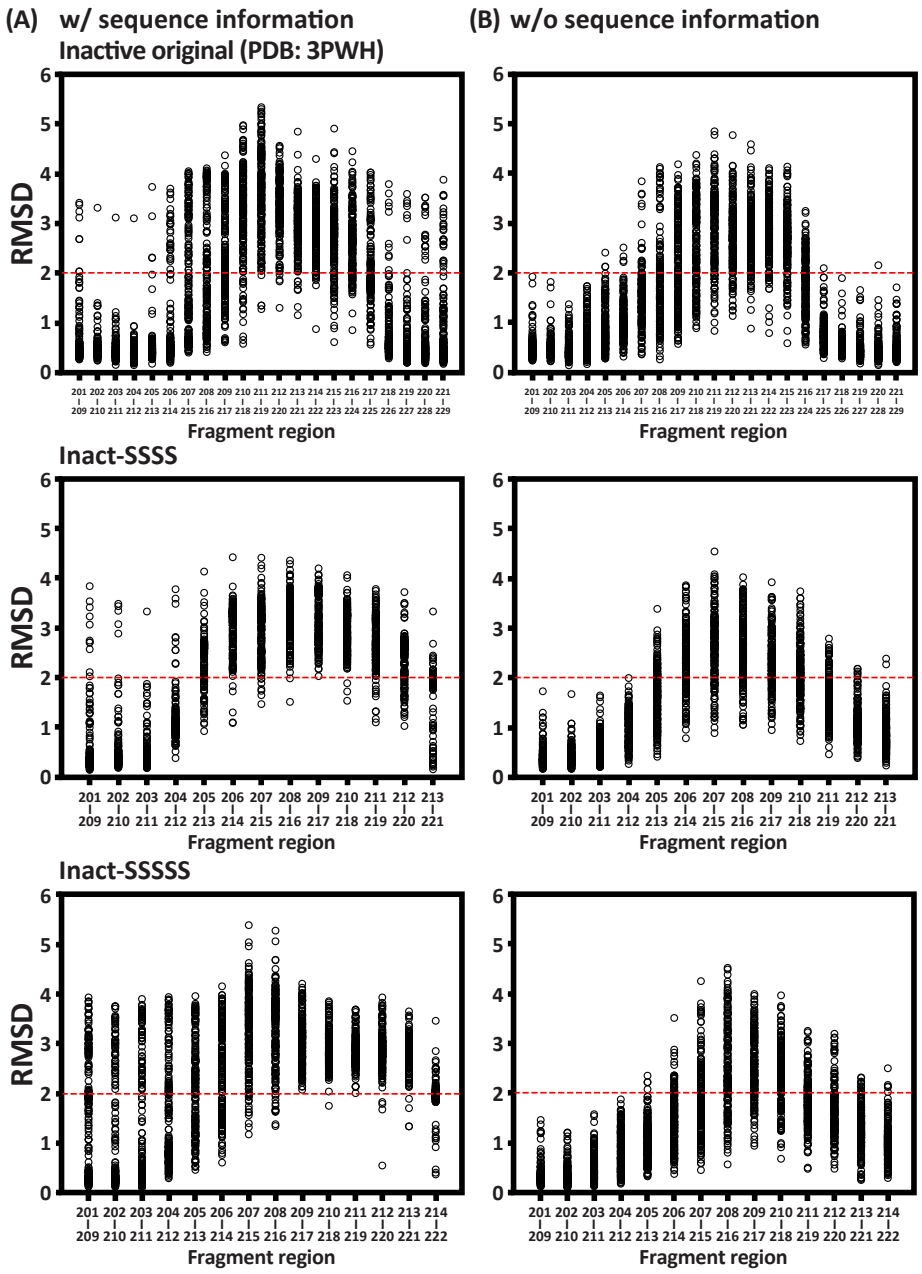

Figure 2-6. Evaluation of sequence-structure compatibility for original and redesigned $A_{2A}R$ structures in the inactive state. (**A**) The plot of RMSD of nine-residue fragments obtained based on the similarity of the sequence and the secondary structure to the corresponding nine-residue structure in the design models. (**B**) The plot of RMSD of nine-residue fragments obtained based only on the similarity of the secondary structure to the corresponding nine-residue structure in the design models. Each circle corresponds to single fragment. The abundance of the circle in low-RMSD area indicates the existence of sufficient number of similar nine-residue structure in naturally occurring proteins. The scarcity of the circles in low-RMSD area indicates there are not many similar nine-residue structures with similar sequences or secondary structures in naturally occurring proteins.

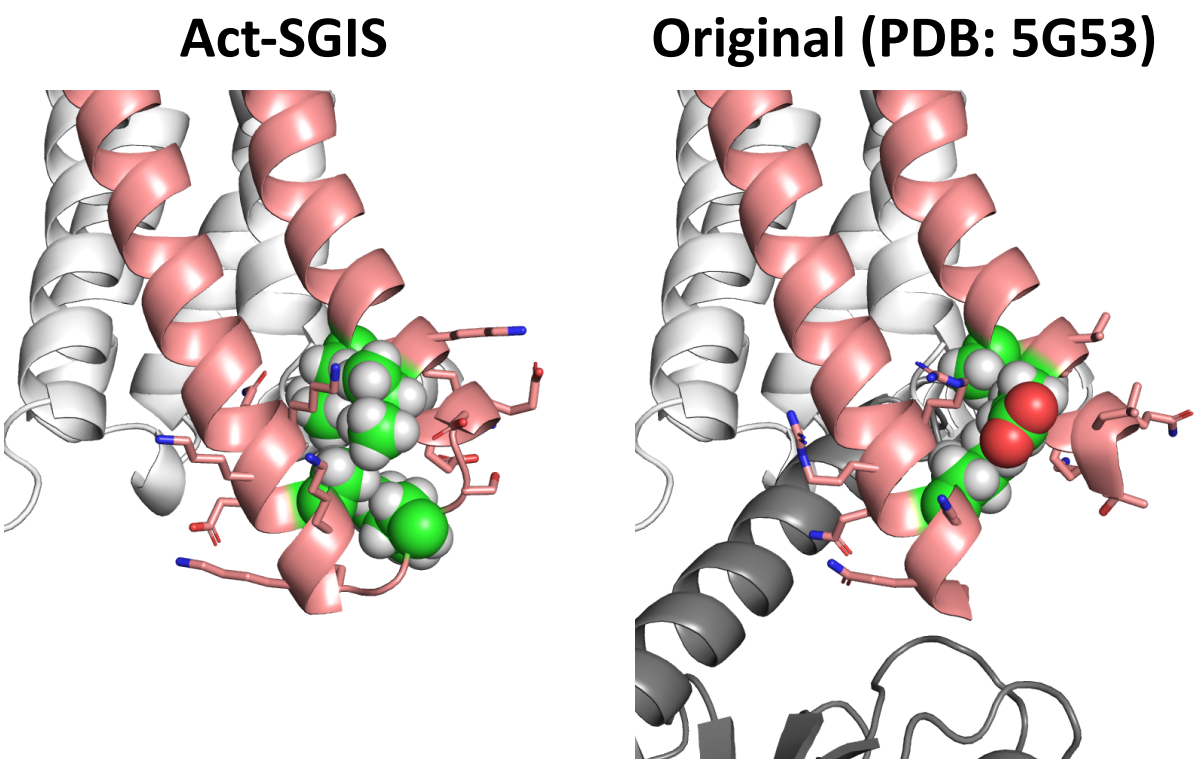

Figure 2-7. (Left) Computational model of Act-SGIS. The sidechains of amino acids in the designed area are shown. Residues that form hydrophobic core packing are shown in spheres in green and the rest are shown in sticks. (Right) Original structure (PDB: 5G53). An engineered $G\alpha_s$ is colored in gray. The sidechains of amino acids in the designed area in Act-SGIS and the corresponding area in the original structure are shown. Missing hydrogens and sidechain model of E228 are modeled by Rosetta. Residues that form core packing in Act-SGIS and residues at the corresponding position in the original structure are colored in green. Other residue positions belonging to TM5-ICL3-TM6 region are colored in pink.

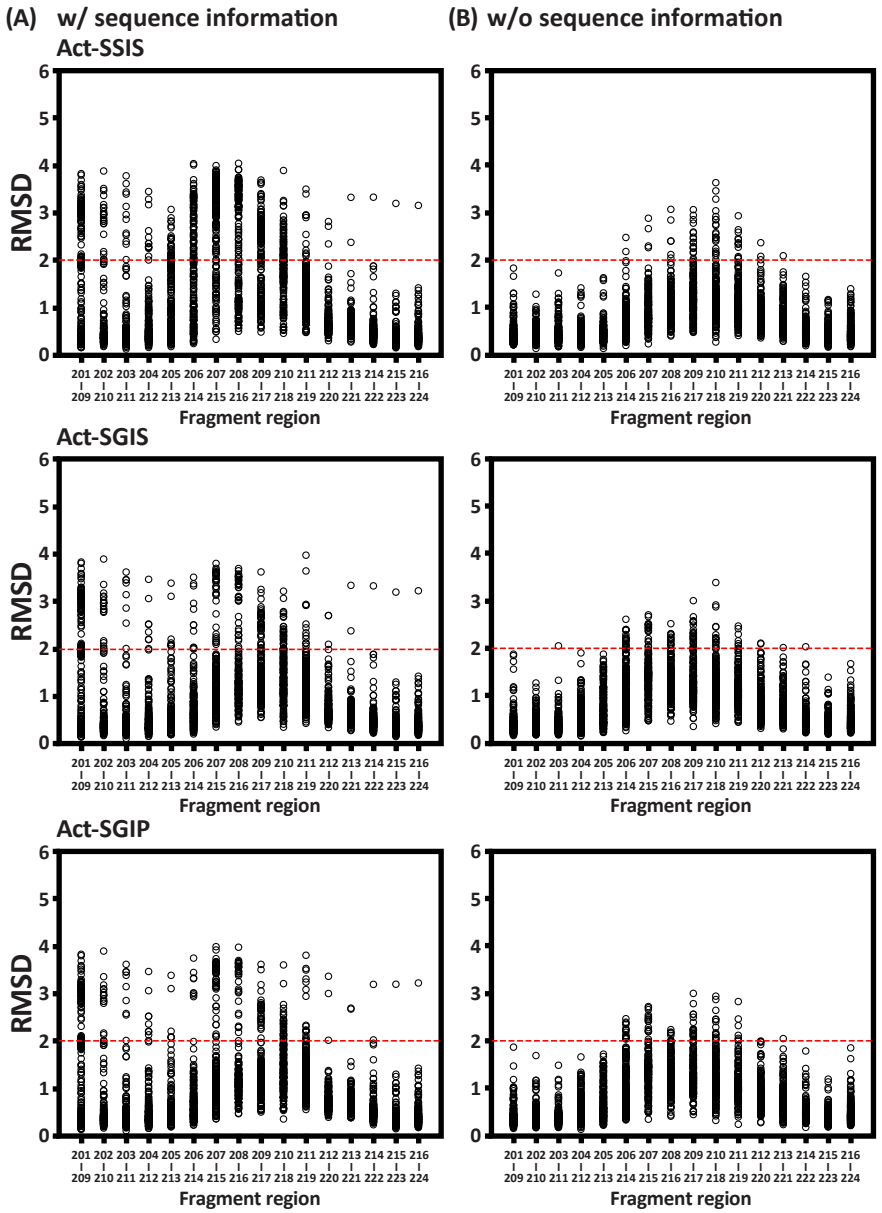

Figure 2-8. Evaluation of sequence-structure compatibility for redesigned $A_{2A}R$ structures in the active state (evaluation was not performed for the original structure, because no models are assigned on the residues of the ICL3 of the original structure (PDB: 5G53)). (**A**) The plot of RMSD of nine-residue fragments obtained based on the similarity of the sequence and the secondary structure to the corresponding nine-residue structure in the design models. (**B**) The plot of RMSD of nine-residue fragments obtained based only on the similarity of the secondary structure to the corresponding nine-residue structure in the design models. Each circle corresponds to single fragment. The abundance of the circles in low-RMSD area indicates there are many similar nine-residue structures with similar sequences or secondary structures in naturally occurring proteins.

Materials and Methods

Computational remodeling of ICL3 to find optimal lengths and positions of de novo ICL3

To investigate the optimal lengths and positions of de novo designed ICL3, computational loop rebuilding was performed between TM5 and TM6 of $A_{2A}R$. The template structures for backbone rebuilding are an inactive state structure of $A_{2A}R$ (PDB: 3PWH) [62] and the active state structure of $A_{2A}R$ (PDB: 5G53) [63]. For every combination of starting residue from TM5 (residue number 198 to 211 for structures of the both states) and ending residue from TM6 (residue number 219 to 234 for inactive state-structure and 224 to 234 for active state-structure) and loop length (three to six residues). Residue models at residue position 222 and 223 are missing in the original $A_{2A}R$ structure of the active state. When loop rebuilding trials were performed using these residues, they were set to be helix. Loop remodeling was performed for 1,000 times using RosettaRemodel protocol[61] in Rosetta software (https://www.rosettacommons.org/software). The residues in cytosolic areas were selected based on positional information of membrane in the target structures of $A_{2A}R$ obtained from Orientation of Protein in Membrane (OPM) database[64]. Note that the structure of engineered G-protein in 5G53 contains was removed.

Backbone selection

After backbone rebuilding trials, based on the number of most frequently observed ABEGO torsion pattern[26] of the generated backbones, the backbone models for sequence design were selected. In detail, from the combinations of loop length and residue positions of rebuilt ICL3 which generated more than 500 structures, the combinations which remained the longest secondary structures were selected. Average structures were generated using generated structures with most frequently observed ABEGO patterns in each run of backbone rebuilding trials.

Limitation of amino acids used for sequence design

For sequence design, limiting the amino acids used in design, based on the environment of the residue position of template backbone models, has been employed by the scoring function or directly restricting the candidate amino acids [31, 33, 34, 60]. There might be typical hydrophobic/hydrophilic patterns among the loop structures to form specific ABEGO patterns that is useful for limiting amino acids for sequence design of loop structures. In addition, because the dihedral angles of Proline and Glycine are distinctive to other 18 kinds of amino acids, the positional information of Proline and Glycine may be additional categories to hydrophobic/hydrophilic patterns, especially in loop structures. To investigate typical patterns of amino acid characteristics, the number of patterns of hydrophobic amino acid, hydrophilic amino acid, Proline and Glycine in naturally occurring αα-units are counted. The four letters corresponding to hydrophobic, hydrophilic, Proline and Glycine were set to be I, S, P and G (ISPG alphabets) from the initials of Insoluble, Soluble, Proline and Glycine, respectively. I named the pattern of the four letters I, S, P and G as ISPG pattern. For hydrophobic amino acids, Ala, Val, Ile, Leu, Phe, Met, Cys and Trp were assigned and for hydrophilic amino acids, Ser, Thr, Tyr, Asn, Gln, Asp, Glu, Lys, Arg and His were assigned. For ABEGO types of BBBB, BABBB and AGBB of αα-units from non-redundant set of protein structures, frequently used ISPG patterns were found. (**Figure 2-9** and **Table 2-5**)

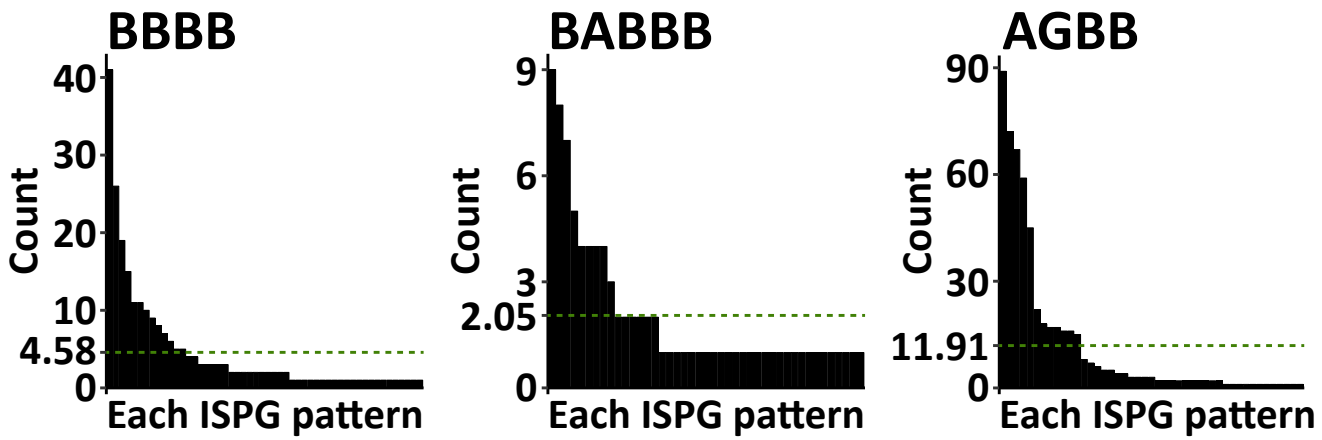

Figure 2-9. The histograms of the counts of each ISPG pattern of the loops of naturally occurring αα-units whose ABEGO patterns are BBBB (left), BABBB (center) and AGBB (right), respectively. Green dotted lines show the average number of the count of each ISPG pattern.

ISPG pattern for BBBB loop	Count	ISPG pattern for BABBB loop	Count	ISPG pattern for AGBB loop	Count
ISIS	41	SSSIS	9	IGIS	89
SSIS	26	ISSIS	8	ISIS	72
IPIS	19	SSSSS	7	SSIS	67
ISSS	15	SSSPS	5	SGIS	59
ISIP	11	SSIIS	4	SGSS	45
SPIS	11	SPSSS	4	IGSS	22
IIIS	10	SPSPS	4	ISIP	18
SSSS	9	ISSSS	4	IGIP	17
SIIS	8	SPIIS	3	SGIP	17
IPSS	7			SSSS	16
GSIS	6			SSIP	16
ISPS	5			ISSS	15
SSIP	5				

Table 2-5. Counts of each ISPG pattern of the ABEGO pattern of BBBB, BABBB and AGBB. Only the patterns more frequently observed than the average count are shown.

In a previous report by Koga et al., the residue positions for hydrophobic core and hydrophilic surface are specified by the solvent accessible surface area (SASA) and the amino acids used for design are limited based on these classifications [34]. In Inact-BBBB, Inact-BABBB and Act-AGBB, the residue position with SASA $\leq$ 35 Å^2 are set to be core position and SASA > 35 Å^2 are set to be surface position. Considering that the I of ISPG pattern corresponds to hydrophobic core and the S corresponds to hydrophilic surface, on each of residue positions of the rebuilt ICL3s, the ISPG alphabet I and S were assigned for residues at core position and surface position defined by the above criteria of SASA, respectively. The ISPG patterns of the rebuilt ICL3s were SSSS for Inact-BBBB, SSSSS for Inact-BABBB and SSIS for Act-AGBB (**Figure 2-9**). Because all the assigned ISPG patterns are appeared more frequently observed than average (**Table 2-5**), the amino acid sequence which satisfy the limitation of the ISPG pattern are designed. If any of ISPG alphabets are replaced to P or G and the variant ISPG pattern appears still more frequently than average, sequence design with limitation of amino acids of the ISPG pattern are also performed.

The restrictions of candidate amino acids are following: 1) For core area defined by SASA, hydrophobic amino acids Ala, Val, Ile, Leu and Phe are used. If I of the ISPG alphabets is set on a loop position, the same set of amino acids are used for design on the position. 2) For surface area defined by SASA, candidate set of amino acids are limited in frequently observed amino acids in three-residue fragments from non-redundant set of naturally occurring proteins with the same ABEGO pattern (for the residue position and one-residue before and after the position) with similar buriedness. For the position one residue before TM6, instead of these candidate amino acids, Asp, Asn, Thr and Ser are used for helix capping, based on a previous report by Koga et al.[34] Considering electrostatic repulsion caused by helix dipole and sidechains, negatively charged amino acids Asp and Asn are removed from the last three residues of TM5 and positively charged residues Lys, Arg and His are removed from the first three residues of TM6. 3) If P of the ISPG alphabets is set on a loop position, Proline is used for design. 4) If

G of the ISPG alphabets is set on a loop position, Glycine is used for design. 5) Cysteine was removed from the set to eliminate disulfide bonds in designed region that is not formed in cytosol and Histidine was removed and multiple protonation states that potentially make unexpected conformational difference depending on pH conditions.

Sequence design and selection

Sequence design was performed on Inact-BBBB, Inact-BABBB and Act-AGBB. The positions designed were set to be between cytosolic areas of TM5 and TM6, based on the positional information of the membrane obtained by OPM database [64]. For the selected averaged backbone structures, sequence design was performed for 1,000 trials with above-mentioned restriction of amino acids using FlxbbDesign protocol in Rosetta (for the scoring function, talaris2014 [65] was used). The designs with the lowest score were selected. For experimental validation, in order of frequency of the ISPG patterns, designs using ISPG patterns of SSSS and SSSSS were selected among designs for Inact-BBBB and Inact-BABBB, respectively (these designs were named Inact-SSSS and Inact-SSSSS). Three designs using ISPG patterns of SSIS, SGIS and SGIP were selected among the designs for Act-AGBB (these designs were named Act-SSIS, Act-SGIS and Act-SGIP, respectively).

Note that the structure of 3PWH is that of a thermally stabilized mutant; however, the wild-type sequence with the mutation N154Q was used for preventing glycosylation in the experiments. For the structure of 5G53, it has a point mutation N154A for preventing glycosylation, but the sequence with N154Q was used in the experiments for the same reason.

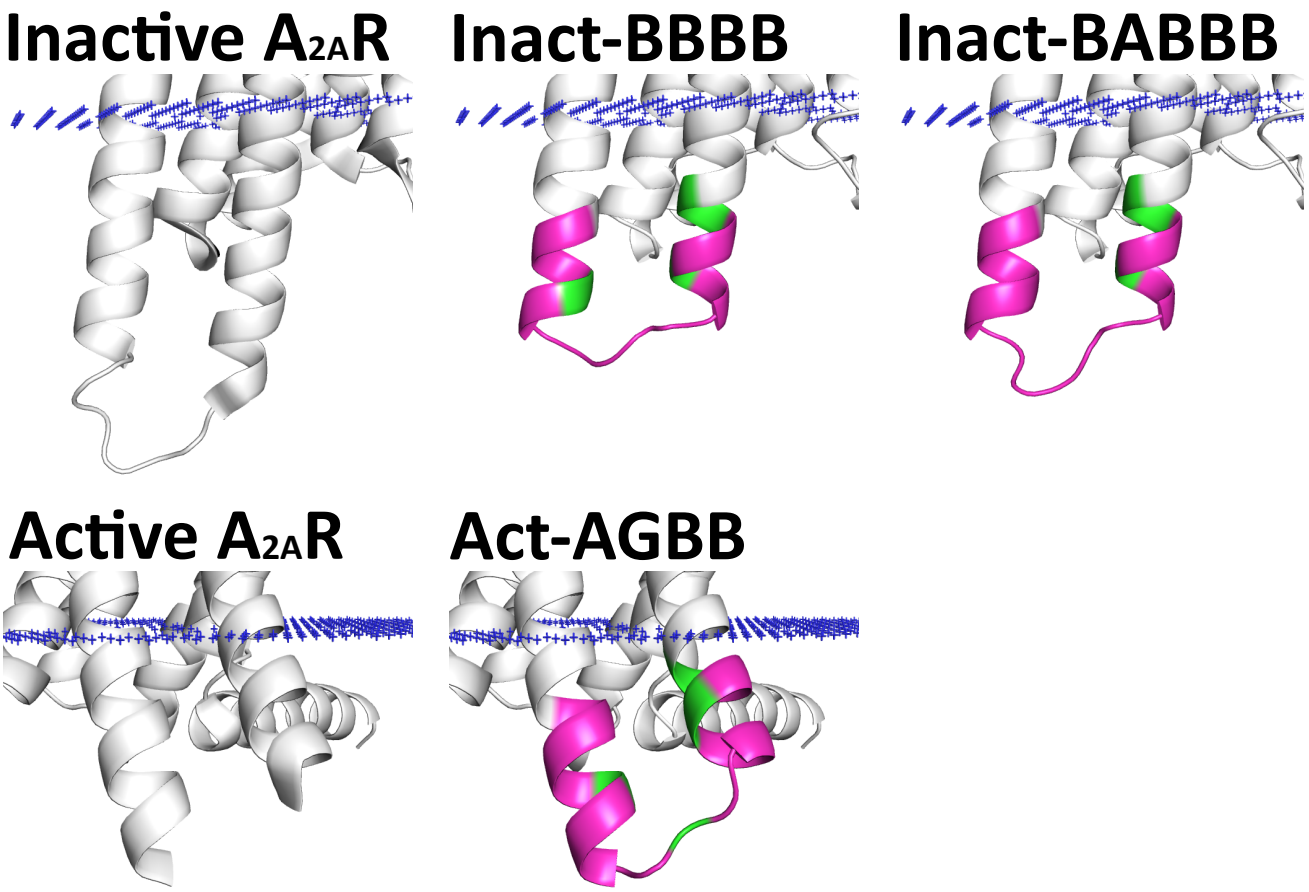

Figure 2-9. Comparison of the original $A_{2A}R$ structures (PDB: 3PWH for $A_{2A}R$ structure in inactive state and PDB: 5G53 for $A_{2A}R$ structure in active state) and the selected template backbone models for design. The residue positions selected for sequence design are colored. Residue positions whose SASA is larger than 35 Å are classified as surface and colored in pink. The other residue positions are classified as core and colored in green. The predicted membrane positions obtained by OPM server[64] are shown with three-dimensional crosses colored in blue.

Experiments of redesigned A_{2A}Rs: DNA construction

Using primers coding redesigned TM5-ICL3-TM6 sequences, inverse PCR was performed on pRS426 vector coding A_{2A}R with C-terminal tagRFP as a template. The PCR product was purified, phosphorylated and ligated. Using the ligation construct, *E. coli* DH5α was transformed and the plasmids coding sequences of redesigned A_{2A}R with C-terminal RFP were extracted. The sequences of these plasmids were confirmed by sequencing.

Experiments of redesigned A_{2A}Rs: Solubilization efficiency

Wild-type A_{2A}R and its variants (Inact-SSSS, Inact-SSSSS, Act-SSIS, Act-SGIS and Act-SGIP) with C-terminal tagRFP were expressed in *Saccharomyces cerevisiae* strain FGY217, and the membranes were prepared as described previously [66]. The membrane samples were resuspended in a membrane buffer (Tris buffer (pH 8.0), 120 mM NaCl, 10% glycerol) and the membrane suspension (4 mg/mL) was solubilized using n-dodecyl-β-D-maltoside (DDM) (final concentration, 1%) with cholesteryl hemisuccinate tris salt (CHS) (final concentration, 0.2%) for an hour at 4 °C. The solubilized samples were centrifuged at 43,000 rpm for 30 min at 4 °C. The red fluorescent protein (RFP) intensity was measured before solubilization and after centrifugation at 595 nm (excitation at 535 nm) using a FilterMax F5 microplate reader (Molecular Devices, Sunnyvale, CA, USA). The solubilization ratio was evaluated as the RFP intensity of the supernatant of solubilized samples after centrifugation divided by the RFP intensity before DDM solubilization.

Supplementary Table

Construct	Sequence
A_{2A}R WT	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAIT ISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWV LSFAIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLG VYLRIFLA**ARRQLKQMESQPLPGERARSTLQKEVHA**AKSLAIIVGLFALCWLPLHIINCFTFFCP DCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAHHHHHHHHH H
Inact-SSSS	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAIT ISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWV LSFAIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLG VYLRIFLAAKKQLKKTSKDEELRELIHAAKSLAIIVGLFALCWLPLHIINCFTFFCPDCSHAPLW LMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAHHHHHHHHH
Inact-SSSSS	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAIT ISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWV LSFAIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLG VYLRIFLAAKKQQKKSDTTDEEAKELLHAAKSLAIIVGLFALCWLPLHIINCFTFFCPDCSHAPL WLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAHHHHHHHHH
Act-SSIS	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAIT ISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWV LSFAIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLG VYLRIFLA**QKKEAKKKNLSDEEIKKL**AKSLAIIVGLFALCWLPLHIINCFTFFCPDCSHAPLWLM YLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAHHHHHHHHH
Act-SGIS	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAIT ISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWV LSFAIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLG VYLRIFLA**QKKEAKKKGLSDEEIKKL**AKSLAIIVGLFALCWLPLHIINCFTFFCPDCSHAPLWLM YLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAHHHHHHHHH
Act-SGIP	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAIT ISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWV LSFAIGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLG VYLRIFLA**QKKEAKKKGLPDEEIKKL**AKSLAIIVGLFALCWLPLHIINCFTFFCPDCSHAPLWLM YLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAHHHHHHHHH

Supplementary Table. The design sequences. Designed area for inactive A_{2A}R and the corresponding area in WT A_{2A}R are shown with underlines. Designed area for active A_{2A}R and the corresponding area in WT A_{2A}R are shown with bold fonts.

Chapter 3: De novo design of fusion partner proteins to stabilize an inactive state

Background

Studies in **Chapter 2** indicate that for stabilization of $A_{2A}R$, the redesigned TM5-ICL3-TM6 region needs to be further more stabilized. A widely used method to stabilize GPCRs by engineering this ICL3 region is fusion partner strategy, the method to replace ICL3 with soluble protein domains (fusion partner proteins) [21-23]. Successes in the previous fusion partner strategy indicate that engineering ICL3 region to be stable as naturally occurring proteins is sufficient to stabilize GPCR structures. Computationally de novo designed protein structures have desired overall structure and are generally much more stable than naturally occurring proteins [58, 59] and these features are thought to be expedient for designing fusion partner proteins tailored to stabilize GPCRs in a specific conformational state. In this chapter, I de novo designed super-stable fusion partner proteins that are customized to stabilize the inactive state of $A_{2A}R$.

This chapter is based on the paper: State-Targeting Stabilization of Adenosine A_{2A} Receptor by Fusing a Custom-Made De Novo Designed α-Helical Protein., Masaya Mitsumoto, Kanna Sugaya, Kazuki Kazama, Ryosuke Nakano, Takahiro Kosugi, Takeshi Murata and Nobuyasu Koga., International Journal of Molecular Sciences. 2021, Volume 22, Issue 23, 12906; https://doi.org/10.3390/ijms222312906. I computationally designed fusion partner proteins and experimentally characterized them. Experiments for chimeric $A_{2A}Rs$ with these fusion partner proteins were performed by Murata group, Chiba university.

Introduction

The fusion partner strategy has been widely used, in which the intracellular loop 3 (ICL3) connecting the transmembrane helices 5 and 6 (TM5 and TM6) is replaced with soluble protein domains, such as T4-lysozyme [22, 23], apocytochrome b_{562}RIL (BRIL) [21, 24], rubredoxin [25], and glycogen synthetase [26]. However, these fusion partner proteins and residue positions for fusion have been identified through experimental trial and error. Moreover, stabilizing GPCRs in a specific state using the fusion approach, which is useful for screening out state-dependent ligands and antibodies [5, 67], has not been achieved.

Recently, principles for designing protein structures from scratch have been developed, which made it possible to create a wide range of new protein structures with high thermal stability [34, 58, 59]. A method to create a diverse set of all-α protein structures ranging from bundle-like topologies with parallel-aligned helices to complicated ones with irregularly arranged helices was previously developed [68]. Using the developed method, I sought to rationally design fusion partner proteins customized for not only thermally stabilizing GPCRs but also stabilizing them in a target state compared to the other states.

In this study, I designed fusion partner proteins customized for stabilizing one of the class A GPCRs, adenosine A_{2A} receptor (A_{2A}R), in an inactive state [51, 62]. A_{2A}R plays important physiological roles, such as the modulation of motor, vascular control, and immunosuppression; therefore, A_{2A}R is a drug target for various diseases, including Parkinson's disease, heart failure, and cancer [47]. Class A GPCRs are the largest subfamily of GPCRs, and the receptors in the class have been suggested to undergo large conformational changes in TM6 associated with TM5 upon the state transitions (**Figure 3-1A**). Therefore, I sought to stabilize A_{2A}R in the inactive state by making a fusion with de novo designed proteins customized to fix the conformation of the two helices in the inactive state (**Figure 3-1B**).

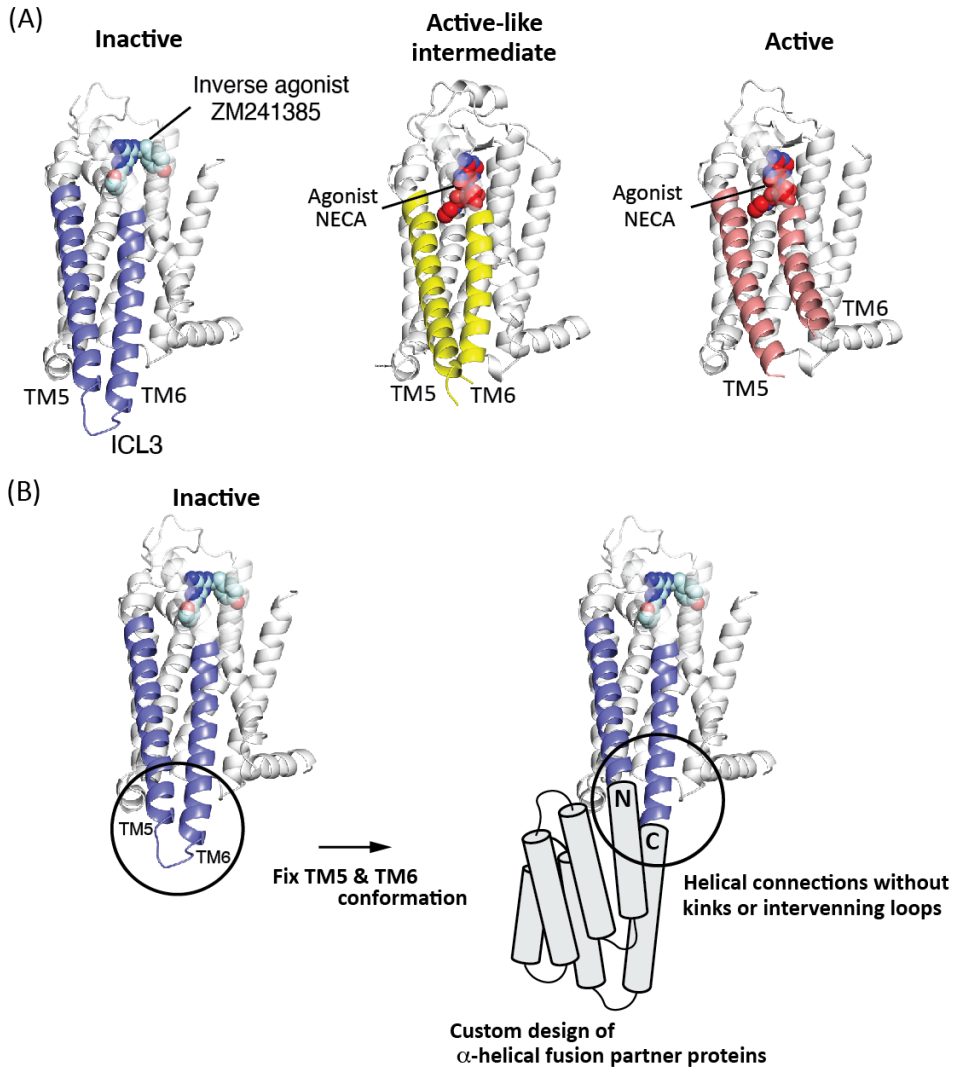

Figure 3-1. Strategy for state-selective stabilization of GPCR, using de novo designed fusion partner proteins. (**A**) Multiple conformational states of $A_{2A}R$. From left, the presented structures, respectively, correspond to the ones in an inactive, active-like intermediate, and activate states (PDB ID: 3PWH, 2YDV, and 5G53, respectively) [63, 69, 70]. The structure in the inactive state binds to the inverse agonist ZM241385, and the structures in the active-like intermediate, and active states bind to the agonist NECA. ZM241385 and NECA are shown in a sphere model. TM5 and TM6 are colored in blue for the inactive state, yellow for the active-like intermediate state, and pink for the active state. The loop connecting TM5 and TM6 is called ICL3. TM6 with TM5 exhibits large conformational changes upon the state transitions. (**B**) My strategy for the state-selective stabilization of $A_{2A}R$. The TM5 and TM6 conformation in a targeted state was tried to fix through the fusion strategy. To this end, I designed α-helical proteins, which can be fused into $A_{2A}R$ in the targeted state through straight helical connections without kinks or intervening loops. In this work, I tested this idea by stabilizing the inactive state.

Results

Computational design of α-helical fusion partner proteins

I assumed that the TM5 and TM6 conformation could be fixed in a specific state through straight helical connections between a fusion partner protein and $A_{2A}R$. Therefore, I sought to design α-helical protein structures de novo, of which the N- and C-terminal helices are, respectively, connected to TM5 and TM6 of an inactive state $A_{2A}R$ structure (PDB ID: 3PWH; this structure is bound to the inverse agonist ZM241385) without any kinks or intervening loops (**Figure 3-2**) (details are described in the **Materials and Methods**).

Using 1688 globular all-α backbone structures with six helices [68] whose N- and C-terminal helices are close to each other, I elongated the N- and C-terminal helices by seven residues, respectively, to fuse with TM5 and TM6 of the $A_{2A}R$ inactive structure. I then selected a set of 389 backbone structures whose terminal helices were elongated without steric clash. From the generated set, I selected backbone structures whose N- and C-terminal helices are well-superimposable to TM5 and TM6 in the inactive-state $A_{2A}R$ structure by calculating root mean square deviation (RMSD) values for the main chain of superimposed residues.

Next, I designed amino-acid sequences that stabilize each of the selected backbone structures by carrying out the cycles of amino acid sequence optimization and optimization of the entire structure [33]. Among the resulting designs with tight core packing [71] and high compatibility between the local sequence and structure [34], the designs whose N- and C-terminal helices were better superimposable to TM5 and TM6 were selected. Note that the designed structures that were inside of the predicted membrane region or had clashes with $A_{2A}R$ after the fusion with $A_{2A}R$ were discarded. (The positional information of the membrane was obtained from the Orientation of Proteins in Membrane (OPM) database (https://opm.phar.umich.edu) [64]).

Next, I selected the designed proteins that exhibited funnel-shaped energy landscapes in Rosetta ab initio folding simulations [34]. Among the selected designs, I further selected those whose N- and C-terminal helices exhibited low fluctuation in molecular dynamics (MD) simulations for experimental characterization. Finally, the designed proteins, FiX1 and FiX2 (**FiX** stands for a **F**usion partner protein customized for **i**nactivation and e**X**tra stabilization), were selected (one of the residues in FiX2 was mutated manually using Foldit [72]. See the **Material and Methods**).

Experimental characterization of FiX1 and FiX2

I experimentally characterized the de novo designed proteins, FiX1 and FiX2, without $A_{2A}R$. These two designs were expressed in *Escherichia coli* and purified using a Ni-NTA column. Both of the designs were found to be well expressed and highly soluble, and are then characterized via circular dichroism (CD) spectroscopy, size-exclusion chromatography combined with multi-angle light scattering (SEC-MALS), and $^{1}H–^{15}N$ heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) spectroscopy (**Figure 3-3**). Both designs showed CD spectra of all-α proteins from 25 to 98 °C, were monomeric in SEC-MALS, and showed well-dispersed sharp NMR peaks. These results indicate that the designs fold into unique α-helical structures as monomers with high thermal stability.

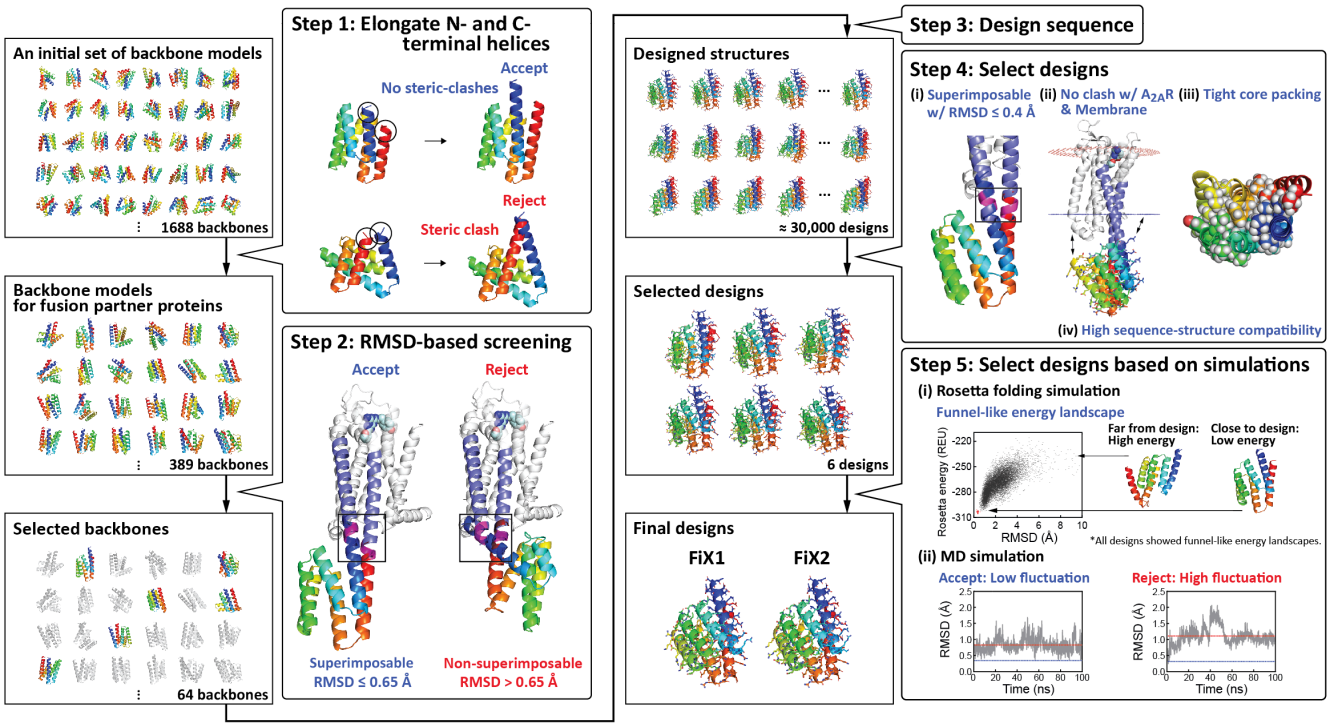

Figure 3-2. Computational protocol for designing fusion partner proteins. As an initial set of backbone structures, 1688 globular all-α backbone structures with six helices, extracted from the previously created all-α backbone structure library [68], were used. (**Step 1**) For the initial set, the N- and C-terminal helices of the structures were elongated to fuse into $A_{2A}R$ through TM5 and TM6. Then, 389 backbone structures whose helices were elongated without steric clash were selected. (**Step 2**) Backbone structures whose elongated N- and C-terminal helices were well-superimposable (the mainchain root mean square deviation (RMSD) for the fused region, equal to or less than 0.65 Å) to TM5 and TM6 of the inactive state $A_{2A}R$ structure were selected. (**Step 3**) For each selected backbone structure, amino acid sequences that stabilize the backbone structure were designed. (**Step 4**) Six designed structures were selected on the basis of the following criteria: the designs are superimposable to $A_{2A}R$ with mainchain RMSD values less than and equal to 0.4 Å for the fused region, have tight core packing [71] and high sequence-structure compatibility [34] and can be fused into $A_{2A}R$ without steric crash with $A_{2A}R$ and the membrane. (**Step 5**) Two designed proteins that exhibited funnel-shaped energy landscapes in Rosetta folding simulations [34] and whose N- and C-terminal helices show low fluctuation in molecular dynamics (MD) simulations were selected. In the MD simulations, blue and red lines show initial and averaged mainchain RMSD values, respectively. Details are described in the **Results** and **Materials and Methods**.

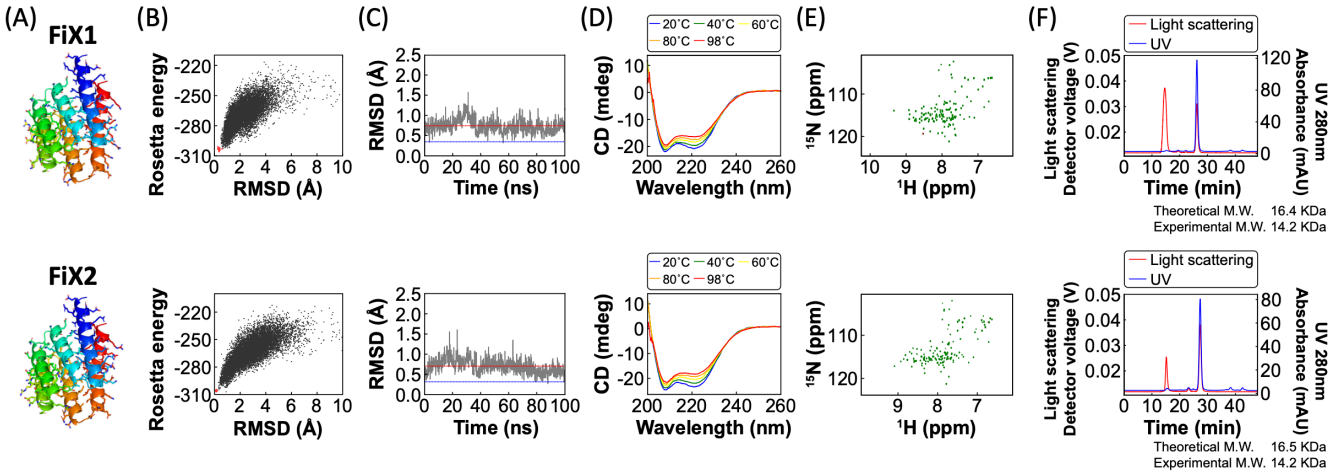

Figure 3-3. Characterization of the de novo designed fusion partner proteins, FiX1 (upper) and FiX2 (lower) without $A_{2A}R$. (**A**) Computational models. (**B**) The energy landscape of each designed protein obtained from Rosetta folding simulations [34]. Each dot represents the lowest energy structure obtained from an independent trajectory starting from an extended chain (black) or the design model (red), the x-axis shows the Cα RMSD from the design model, and the y-axis shows the Rosetta full-atom energy. (**C**) Structural fluctuations of the N- and C-terminal helices in the MD simulation, starting from each design model. The mainchain RMSD of the N- and C-terminal helices of each snapshot structure during an MD trajectory against TM5 and TM6 of $A_{2A}R$ is shown along the time course. (Red and blue lines are the averaged and initial RMSD values, respectively.) (**D**) Far-ultraviolet CD spectra at various temperatures from 20 to 98 °C. (**E**) Two-dimensional ^{1}H–^{15}N HSQC spectra at 25 °C and 600 MHz (in parts per million, p.p.m). (**F**) Size-exclusion chromatograms combined with multi-angle light scattering (SEC-MALS) demonstrate that these designed proteins are monomeric in solution. M.W. stands for molecular weight.

Experimental characterization of $A_{2A}R$ fused with FiX1 and FiX2

The genes encoding the wild-type $A_{2A}R$ ($A_{2A}R$ WT) and chimeras with fusion partner proteins, $A_{2A}R$ fused with FiX1 ($A_{2A}R$–FiX1), FiX2 ($A_{2A}R$–FiX2), and BRIL ($A_{2A}R$–BRIL), were constructed and expressed in yeast as described previously (for each construct, a red fluorescent protein (RFP) was appended at the C-terminus) [66, 73]. It is known that the innate instability of GPCRs gives rise to low yields upon detergent extraction from the membrane. As one of the stability metrics, the solubilization efficiency upon detergent extraction (the ratio of the fluorescent intensity after solubilization using n-decyl-D-maltopyranoside (DM) over that before solubilization) was evaluated.

$A_{2A}R$–FiX1 showed significantly improved solubilization efficiency compared to the wild-type; the efficiency was greater than that of $A_{2A}R$–FiX2 (**Table 3-1**). Therefore, the stability of $A_{2A}R$–FiX1 was further studied by measuring the apparent melting temperatures in the clear-native polyacrylamide gel electrophoresis (CN–PAGE) method [66]. The melting temperature was found to be significantly increased (**Figure 3-4B**), which is consistent with the solubilization efficiency results. For comparison, the solubilization efficiency and melting temperature for $A_{2A}R$–BRIL were also measured; these values were comparable to those of $A_{2A}R$–FiX1 (see **Discussion**).

Construct	Solubilization efficiency (%)
$A_{2A}R$ WT	24 ± 10
$A_{2A}R$–BRIL	57 ± 19
$A_{2A}R$–FiX1	59 ± 21
$A_{2A}R$–FiX2	42 ± 17

Table 3-1. Solubilization efficiencies of $A_{2A}R$ fused with or without fusion partner proteins. Results are reported as mean ± standard deviation for n = 3 independent measurements.

Next, the ligand-binding affinities of $A_{2A}R$ WT, $A_{2A}R$–BRIL, and $A_{2A}R$–FiX1 were investigated using the radioligands of an inverse agonist [^{3}H]-ZM241385 and an agonist [^{3}H]-NECA (see **Materials and Methods**). The binding affinities of $A_{2A}R$ WT and $A_{2A}R$–BRIL against ZM241385 were similar to each other, and those against NECA were also similar to each other (the obtained equilibrium dissociation constants (K_d) of $A_{2A}R$ WT and $A_{2A}R$–BRIL against ZM241385 were 10.5 nM and 15.7 nM, respectively; K_d values of $A_{2A}R$ WT and $A_{2A}R$–BRIL against NECA were 161.5 nM and 193.1 nM, respectively) (**Table 3-2** and **Figure 3-4C**). While the binding affinity of $A_{2A}R$–FiX1 against ZM241385 was similar to those of $A_{2A}R$ WT and $A_{2A}R$–BRIL, but significant binding against NECA was not observed (**Table 3-2** and **Figure 3-4C**). Moreover, the binding affinity of $A_{2A}R$–FiX1 to NECA was investigated using the inhibition assay, which was approximately 100-times lower than that of $A_{2A}R$ WT (**Figure 3-4D**).

For $A_{2A}R$, the correspondence between the ligand-binding states and the conformational states has been well studied: the binding of the inverse agonist ZM241385 shifts the conformational equilibrium to the inactive state, and that of the agonist NECA shifts the equilibrium to the active state [74, 75]. Therefore, these results indicate that the shifting of the conformational equilibrium of $A_{2A}R$ toward the inactive state by fusion with FiX1 was successful. The experimental structure information of $A_{2A}R$–FiX1 is required to further support my conclusion.

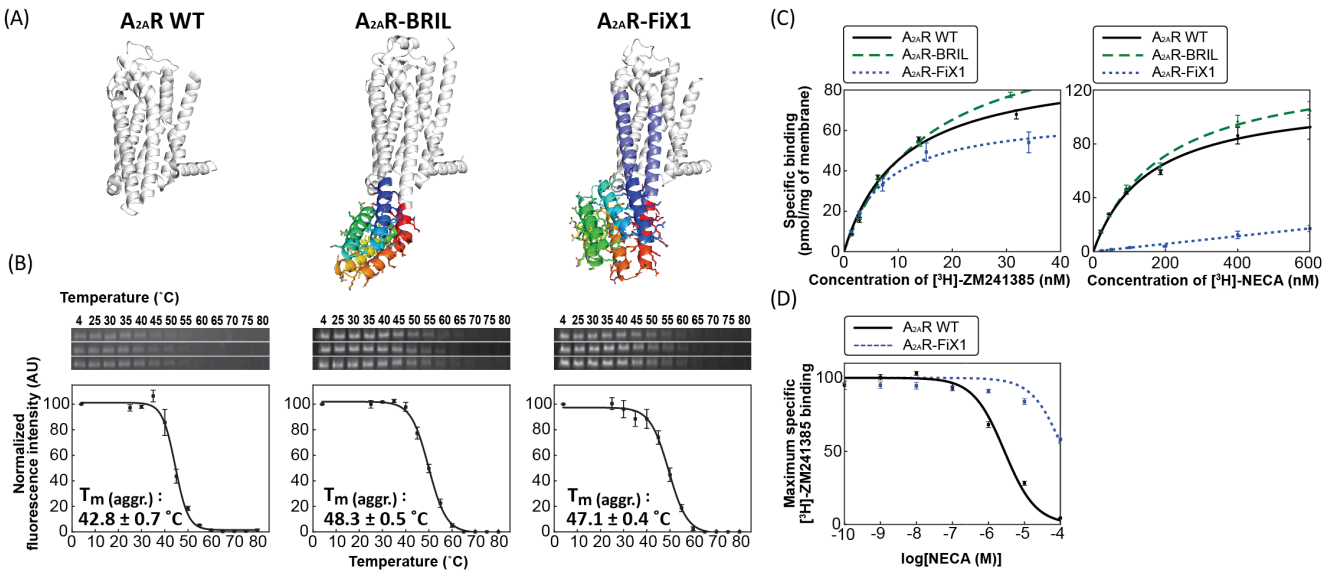

Figure 3-4. Experimental characteristics of $A_{2A}R$ fused with or without fusion partner proteins. (**A**) Crystal structures of $A_{2A}R$ WT (PDB ID: 3VG9) [76], $A_{2A}R$–BRIL (PDB ID: 4EIY) [24], and the computational model of $A_{2A}R$–FiX1. (**B**) (top) Monomer bands in the clear-native PAGE for each $A_{2A}R$ sample heated at various temperatures. (bottom) The fluorescence intensities of the gel images for each $A_{2A}R$ sample with temperature. The thermal transition from soluble to aggregated states was fitted (solid line) to obtain the midpoint temperature, Tm(aggr.) (**C**) Saturation binding curves of [^{3}H]-ZM241385 (left) and [^{3}H]-NECA (right) to $A_{2A}R$ WT (solid line), $A_{2A}R$–BRIL (dashed line), and $A_{2A}R$–FiX1 (dotted line). (**D**) Inhibition of [^{3}H]-ZM241385 binding to $A_{2A}R$ WT (solid line) and $A_{2A}R$–FiX1 (dotted line) by NECA. The binding of [^{3}H]-ZM241385 in the absence of NECA was set to 100%. All measurements were carried out three times independently; dots show the average and whiskers show the s.e.m. for n = 3.

Construct	K_d (nM)	
	ZM241385	**NECA**
$A_{2A}R$ WT	10.5 ± 0.3	162 ± 44
$A_{2A}R$–BRIL	15.7 ± 0.6	191 ± 22
$A_{2A}R$–FiX1	7.3 ± 0.5	N. D.

Table 3-2. Dissociation constants (K_d) of $A_{2A}R$ fused with or without fusion partner proteins by saturation binding assay. N.D. stands for not detected. Results are reported as the mean ± standard deviation for an n = 3 independent assay.

Discussion

I succeeded in rationally designing a fusion partner protein, which thermally stabilized one of the class A GPCRs, $A_{2A}R$, and stabilized it in an inactive state. I carried out the custom-made de novo design of fusion partner proteins, of which the N- and C-terminal helices are, respectably, connectable to the TM5 and TM6 in the inactive state without kinks or intervening loops. The de novo designed fusion partner proteins FiX1 and FiX2 were found to fold as monomers with high thermal stability. The fusion of $A_{2A}R$ with FiX1 was found to be not only thermally stabilized but also stabilized in the inactive state, as I designed.

I expected that $A_{2A}R$–FiX1 would be more stable than $A_{2A}R$–BRIL, since the melting temperature of FiX1 is over 98 °C in circular dichroism, which is far more than that of BRIL, around 65 °C [77]. However, the apparent melting temperatures of $A_{2A}R$–FiX1 and $A_{2A}R$–BRIL were almost the same. This suggests that the overall stability of the $A_{2A}R$ chimera is determined by the transmembrane helices or the other loops rather than the ICL3 region; therefore, the stabilization by making fusion proteins may be saturated.

Previous fusion partner strategies have used naturally occurring proteins or their mutants as fusion partner proteins. However, the number of naturally occurring proteins is limited, and their structures have been optimized during evolution to express their functions. Among naturally occurring protein structures, the ones in which the terminal helices are in close distances would be readily found. However, for fusion partner proteins to be connected to TM5 and TM6 in a specific state using straight helices, the terminal helices must have a specific distance, angle, and helical cycle. It would be difficult to find naturally occurring proteins whose terminal helix geometries exactly satisfy all the three conditions. Moreover, most naturally occurring proteins are not stable; therefore, their folding ability can be impaired by only a few mutations. In contrast, my de novo protein design approach allows us to create stable proteins with specific helix geometries and high stability, without experimental trial and error.

The next question is whether the developed method can stabilize $A_{2A}R$ in other states (i.e., active-like intermediate or active states). GPCRs in the basal condition favor the inactive state over the other states [2]. Therefore, stabilization of $A_{2A}R$ in the other states may be more difficult than stabilization in the inactive state. Nevertheless, the success in the stabilization in the inactive state indicates that my developed method has the potential for stabilization of $A_{2A}R$ in the other states and further for the state-selective stabilization of other GPCRs.

Materials and Methods

Selection of backbone structure models for fusion partners

1688 backbone structures from the previously created all-α backbone structure library with six helices [68] was extracted with the following restrictions: (1) the maximum consecutive buried residues (a residue with accessible surface area < 5.0 $Å^2$ calculated using FreeSASA (https://freesasa.github.io)[78] with a probe of radius 3.0 Å is regarded as buried) in a structure is less than 4, (2) the distance between the N- and C-terminal Cα atoms is less than 12.0 Å corresponding to a rough distance between TM5 and TM6, and (3) a low radius of gyration. This selection was done by Koya Sakuma, a former Ph. D. course student in SOKENDAI.

Next, the N- and C-terminal helices of the extracted backbones were extended by appending seven helical residues using the RosettaRemodel protocol [61] in Rosetta software (https://www.rosettacommons.org/software). The calculations were attempted 100 times, and if 100 backbone structures were successfully generated, their averaged structure was used as a backbone structure in the following calculation; ultimately, 389 averaged backbone structures were obtained. Next, among these structures, I selected those whose terminal helices were fusible to TM5 and TM6 of the inactive state $A_{2A}R$ structure (PDB: 3PWH, obtained from the PDB OPM database [64]). Note that the structure of 3PWH is that of a thermally stabilized mutant; however, the wild-type sequence with the mutation N154Q was used for preventing glycosylation in the experiments. To this end, mainchain RMSD values were calculated by superimposing all pairs of three consecutive residues in the N- and C-terminal helices of the backbone structure (the residues are selected from those of the residue number from 2 to 11, and those from 113 to 123, respectively) against all pairs of the three consecutive residues in TM5 and TM6 at the cytoplasmic side (the residues were selected from those of the residue number from 204 to 211, and those from 219 to 229, respectively) (see **Table S3-4**). I selected the backbone

structures that were superimposable with a mainchain RMSD value less than or equal to 0.65 Å as fusible ones; 64 backbone structures were obtained.

Sequence design for further backbone selection

I further screened 64 backbone structures via the sequence design of each backbone structure, followed by entire structure optimization, using the FlxbbDesign protocol in Rosetta (for the score function, talaris2014 [65] was used). In the sequence design, amino acid residue types used for each residue position were restricted based on the buriedness: hydrophobic residues were used in the protein core, hydrophilic residues on the surface, and both hydrophobic and hydrophilic residues at the boundary. Cysteine was not used to prevent unintentional disulfide bond formation; histidine was not used because of its several protonation states; glycine was used for the first and last helix residues. After the sequence design, the designs whose backbone ABEGO torsion patterns ("A" indicates the alpha region of the Ramachandran plot; "B" the beta region; "G" and "E", the positive phi region; and "O", the cis peptide conformation [60]) were different from those of original backbone structures were discarded. I performed the design calculations 50 times independently, and then selected backbone structures from which almost all designs were successfully generated without a change in the backbone ABEGO torsion pattern. Finally, I obtained three backbone structures.

Sequence design

For each of the selected three backbone structures, sequence designs were performed 10,000 times, using the design protocol described in above section, with additional restrictions for used amino acid residue types. (1) When the backbone dihedral angle was classified as G based on the ABEGO classification [60], the amino acid type of the residue was fixed to glycine; (2) serine and threonine on α-helices were not used, except for the first and last helix residues, because these residues have a tendency to bend α-helices [79]; (3) positively charged residues, lysine and arginine, were not used in the first three helix residues, based on a previous report [80].

Selection criteria after sequence design

After the sequence design, the designs were selected by the following criteria. (1) the mainchain RMSD value between the N- and C-terminal helices and TM5 and TM6 of $A_{2A}R$ is less than or equal to 0.4 Å and (2) designs with tight-core packing calculated by Rosetta Holes [71] (more than 0 and less than 2.0) and Packstat (more than 0.6) in Rosetta software (https://www.rosettacommons.org/software). Then, designed structures that had clashes with $A_{2A}R$ in a fused structure were discarded (a clash was identified by the distances between the Cα atoms of a design and $A_{2A}R$ being less than 5.5 Å). In addition, designed structures that were to be inside of the membrane were also discarded (it is not allowed that even one of the atoms of a designed protein in a fused structure is in the membrane region; the membrane region was obtained from the Orientation of Proteins in Membrane (OPM) database [64]). Moreover, designs with high compatibility between the local sequence and structure were selected in the following manner. For each nine-residue frame of a designed protein, 200 nine-residue fragments were collected from a non-redundant set of X-ray structures based on the sequence similarity and secondary structure prediction. Then, for each frame, the RMSD of the local structure against each of the 200 fragments was calculated. Designs were ranked according to the summation of the log-ratio of the fragments, for which the RMSD was less than 1.5 Å, across all nine-residue frames, and six design sequence with high values were selected.

Rosetta folding simulation

Energy landscapes of the designed sequences were obtained from Rosetta folding simulations [34]. For each amino acid sequence of designed proteins, 10,000 predicted structure models were generated starting from a completely extended structure. Furthermore, 200 energy-minimized structure models were generated starting from each of the designed protein structures. The energy landscape of each designed structure was evaluated by the shape of the scatter plot of the Rosetta score of the generated models versus the corresponding RMSD values to the designed structure. I confirmed that the predicted energy landscapes for all the designs were funnel-like.

Molecular dynamics (MD) simulation

MD simulations were performed to select designed structures whose N- and C-terminal helices did not fluctuate significantly in the simulations. Mainchain RMSD values were calculated by superimposing three consecutive residues in the N- and C- terminal helices of each snapshot structure generated during an MD trajectory against three consecutive residues in TM5 and TM6. The positions for the three consecutive residues were those used in the RMSD-based screening calculation (**Table S4-4**). The designs with average RMSD values of more than 0.75 Å or unexpected hydrophilic interactions in the MD simulation trajectories were discarded.

The AMBER16 software suite [81] was used to perform the MD simulations. Hydrogen atoms were added using the LEaP module in AMBER16, after removing those from the design models. A box with a 12 Å buffer of water models around the protein model in each direction was created. TIP3P [82] and AMBER ff99SB force fields [83] were used as the water model and protein force field, respectively. Periodic boundary conditions were set at a cut-off distance of 10 Å. Long-range electrostatic interactions were treated using the particle mesh Ewald method.

At the beginning of the simulations, energy minimization of the solvent was performed with harmonic restriction for protein atoms, and subsequently energy minimization without restriction was performed. Next, the temperature of the system was gradually increased from 0 to 300 K in 100 ps in an NVT ensemble with a Langevin thermostat and harmonic positional restriction for the protein atoms. After the heating step, a 100 ns MD simulation was performed at 1 atm at 300 K in an NPT ensemble with isotropic position scaling, setting one step as 0.002 ps.

A manual mutation using Foldit

The Tyr residue at the position 67 in FiX2 was manually mutated to Leu using Foldit [72] to optimize the core packing.

Experiments of de novo designed fusion partner proteins: protein expression and purification

Plasmids with FiX1 or FiX2 DNA sequences between the NdeI and XhoI restriction sites in pET21b vectors were purchased from FASMAC (Kanagawa, Japan). *E. coli* BL21 Star (DE3) competent cells were transformed with the plasmids and cultured in MJ9 minimal media containing ^{15}N-labeled ammonium sulfate as a nitrogen source, and ^{15}N-labeled FiX1 and FiX2 were expressed. After the cells were spun down, they were suspended in phosphate-buffered saline (PBS) buffer, 137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, and 1.8 mM KH2PO4 at pH 7.4 with BugBuster (EMD Millipore Corp., Billerica, MA, USA), protease inhibitor, lysozyme, and deoxyribonuclease.

From the cell lysates, FiX1 and FiX2 samples with a His-tag at C-terminus were purified using a Ni-NTA column. The purified samples were dialyzed against PBS buffer at pH 7.4. The purity of the FiX1 and FiX2 samples was confirmed via SDS-PAGE (**Figure S3-1**) and mass spectrometry.

Experiments of De Novo Designed Fusion Partner Proteins: Circular Dichroism (CD)

CD spectra were measured using J-1500 KS (JASCO Corp., Tokyo, Japan). By heating the samples from 20 to 98 °C at a rate of 1 °C per min, far-UV CD spectra were measured from 260 to 200 nm at various temperatures of 20, 40, 60, 80, and 98 °C using 10 μM FiX1 and FiX2 samples in PBS buffer (pH 7.4) in a 1-mm path length cuvette.

Experiments of de novo designed fusion partner proteins: size exclusion chromatography combined with multi-angle light scattering (SEC-MALS)

SEC-MALS measurements were performed using a miniDAWN TREOS static light scattering detector (Wyatt Technology Corp., Santa Barbara, California, USA) and a high-performance liquid chromatography (HPLC) system (1260 Infinity LC, Agilent Technologies, Santa Clara, CA, USA). Approximately 180 μM FiX1 and FiX2 samples in PBS buffer (pH 7.4) were injected into a Superdex 75 increase 10/300 GL column (GE Healthcare) equilibrated with PBS at a flow rate of 0.5 mL/min. Sample concentrations were evaluated based on the absorbance at 280 nm detected by using HPLC system. Static light scattering data at the angles of 43.6°, 90.0°, and 136.4° were obtained using a 659 nm laser. The data were analyzed using ASTRA software (https://store.wyatt.com/shop/viscostar/viscostar-iii/astra-software/) (version 6.1.2, Wyatt Technology Corp., Santa Barbara, California, USA) with a dn/dc value of 0.185 mL/g.

Experiments of de novo designed fusion partner proteins: 2D ^{1}H-^{15}N HSQC measurement

For ^{15}N-labeled FiX1 and FiX2 samples of 400 to 600 μM in 90% H_2O/10% D_2O PBS buffer (pH 7.4), 2D HSQC NMR spectrum measurements were performed using a JNM-ECA 600 MHz spectrometer (JEOL, Tokyo, Japan). The obtained NMR spectra were analyzed using the Delta NMR software (https://nmrsupport.jeol.com/Software) (version 5.2.1, JEOL, Tokyo, Japan).

Experiments of $A_{2A}R$-designed fusion partner proteins: DNA construction

The coding sequence of the human adenosine A_{2A} receptor ($A_{2A}R$) from residues 1−316 was amplified by using the polymerase chain reaction (PCR) method, in which N154 was replaced by Q to eliminate N-linked glycosylation [76]. The DNA fragment was inserted into the plasmid pDDGFP-2 [84], including TagRFP-His8 at the C-terminus [66]. The intracellular loop 3 (ICL3) of the $A_{2A}R$ (denote $A_{2A}R$ WT) was replaced with FiX1 or FiX2. The residue numbers of the de novo designed fusion partner proteins refer to the original amino acid sequences on the pET21b vectors (**Table S3-1**).

Experiments of $A_{2A}R$-designed fusion partner proteins: Solubilization efficiency

Wild-type $A_{2A}R$ and its variants ($A_{2A}R$–BRIL, $A_{2A}R$–FiX1, and $A_{2A}R$–FiX2) were expressed in *Saccharomyces cerevisiae* strain FGY217, and the membranes were prepared as described previously[66]. Briefly, membranes were resuspended in a solubilization buffer (50 mM Tris, 120 mM NaCl, 20% glycerol, and 1 μg/mL 6-amidinonaphthalen-2-yl 4-guanidinobenzoate bis (methanesulfonate) (Alfresa Pharma Corp., Osaka, Japan); pH 8.0). The membrane suspension (5 mg/mL) was solubilized using n-decyl-β-D-maltopyranoside (DM) (final concentration, 1%) (Anatrace, Maumee, OH, USA) for 30 min at 4 °C. The red fluorescent protein (RFP) intensity was measured before and after solubilization at 595 nm (excitation at 535 nm) using a FilterMax F5 microplate reader (Molecular Devices, Sunnyvale, CA, USA). The solubilization ratio was evaluated as the RFP intensity of the unpurified $A_{2A}R$–RFP divided by that of the whole membrane protein mixture soon after DM solubilization.

Experiments of $A_{2A}R$-designed fusion partner proteins: Clear-native PAGE

To evaluate the apparent melting temperatures of the wild-type $A_{2A}R$ and its variants ($A_{2A}R$–BRIL, $A_{2A}R$–FiX1) solubilized in 1% n-dodecyl β-D-maltopyranoside (DDM; Anatrace, Maumee, OH, USA) containing 0.2% cholesterol hemisuccinate (CHS; Sigma-Aldrich, Saint Louis, MO, USA), using clear native polyacrylamide gel electrophoresis (CN-PAGE) with modified Coomassie Brilliant Blue G-250 (mCBB) stain[66]. The samples were fused with RFP at the C-terminus, exhibiting fluorescence at 595 nm.

The samples were heated at each prescribed temperature for 5 min (the temperatures were prescribed in the range of 25–80 °C), and then immediately cooled on ice. CN-PAGE was performed using 10% Tris-glycine separation gel applied to the treated samples with CN-PAGE buffer (200 mM Tris-HCl, 20% glycerol, 1.0% mCBB, and 1.0% DDM; pH 8.6) at a ratio of 1:1. The samples on the gel were visualized (i.e., gel imaging was performed) using FUSION SOLO 7S (Vilber–Lourmat, Marne-la-Vallée, France) after a 5-s exposure to green light at 530 nm with a 655 nm cutoff filter.

The melting temperatures of the samples were determined from the fluorescence intensities of the monomeric bands on the CN-PAGE gel. The normalized fluorescence intensity was calculated by dividing the fluorescence intensity of the monomeric bands after heating by that before heating and is represented as a percentage. The obtained intensities of the monomeric bands were quantified using the ImageJ software (https://imagej.nih.gov/ij). The melting temperatures were calculated using GraphPad Prism 4.0 (GraphPad Software, San Diego, CA, USA) as previously described [66].

Experiments of $A_{2A}R$-designed fusion partner proteins: Radioligand binding assay

Radioligand binding assays were performed using yeast cell membranes expressing the wild-type $A_{2A}R$ and its variants ($A_{2A}R$–BRIL and $A_{2A}R$–FiX1). The protein concentrations of the membranes were determined by using the bicinchoninic acid method (Thermo Fisher Scientific, Waltham, MA, USA) with bovine serum albumin as a standard. All experiments were performed in triplicate (independent expressions). For the saturation-binding assay, 10 μg of membranes were incubated (3 h on ice) with the inverse agonist [^{3}H]-ZM241385 (American Radiolabeled Chemicals, Saint Louis, MO, USA) at concentrations ranging from 5 to 80 nM and the agonist [^{3}H]-NECA (PerkinElmer, Waltham, MA, USA) at concentrations ranging from 25 to 600 nM. Non-specific binding was determined in the presence of 10 μM ZM241385 (Sigma-Aldrich, St. Louis, MO, USA) and 100 μM NECA (Sigma-Aldrich, St. Louis, MO, USA), respectively.

For the competition-binding assay, 10 μg of membranes were incubated with 20 nM of [^{3}H]-ZM241385 and unlabeled NECA at concentrations ranging from 1 nM to 100 μM of [^{3}H]-NECA for 3 h on ice. The unbound ligand was removed by rapid vacuum filtration over GF/F filters (GE Healthcare, Chicago, IL, USA). Filtration was performed using a MINI-VAC (Yamato Scientific, Tokyo, Japan) at room temperature. The filters were washed twice with solubilization buffer. After adding 5 mL of Filter-Count (PerkinElmer, Waltham, MA, USA), radioactivity was measured using an LSC-6100 liquid scintillation counter (Hitachi ALOKA Medical, Tokyo, Japan). The collected data were analyzed by using a nonlinear regression-fitting program in GraphPad Prism 8 (GraphPad Software, San Diego, CA, USA).

Supplementary figures and tables

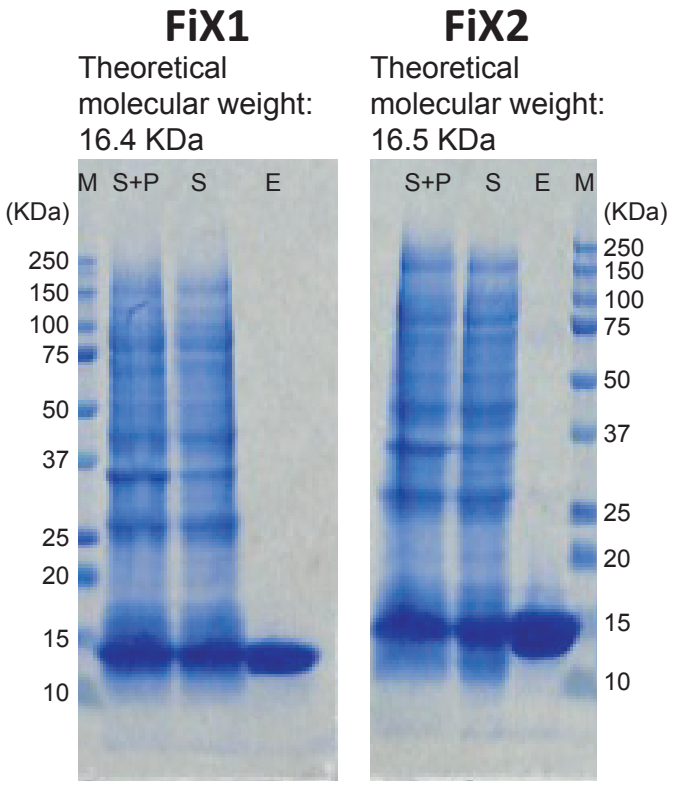

Figure S3-1 | SDS-PAGE results for FiX1 and FiX2

M, S+P, S and E indicate marker, supernatant+pellet, supernatant and elution, respectively.

Construct	Truncated regions in $A_{2A}R$	Inserted regions in fusion partner proteins
$A_{2A}R$–BRIL	K209 - G218	A1 - L106
$A_{2A}R$–FiX1	K209 - R220	R10 - E119
$A_{2A}R$–FiX2	K209 - E219	R10 - E118

Table S3-1 | Sequence regions for $A_{2A}R$ and fusion partner proteins in $A_{2A}R$ chimeras

Construct	Sequence
FiX1	mGEEEERRRRLLELLKRIAELLERGDLEEALKLVKKLAKEQGRQEIIDYIEEVLRLYQEGNREEARKLLE ELLRRLEKEGDTEFRELIRIILEFLELEERGDLEEAKKLARELKKQVDEQEKRLGlehhhhhh
FiX2	mGEEEERRRRLLELLERIARLLKRGDLEEALKLVKKLAKEQGEQEIIDFIEEVLRLYQEGNREQARELLE RLLRNLEKRGNQDFRNLIEIILRILELEQRGNQEEIKKLAEELRREVEERKRKLGlehhhhhh

Table S3-2 | Amino acid sequences of FiX1 and FiX2

Computationally designed amino acid sequences are described in uppercase and residues added to allow expression, purification, and the spacers between a designed sequence and the C-terminal His-tag are described in lowercase.

Construct	Sequence
A_{2A}R–FiX1	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAITISTG FCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFAIGLT PMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLGVYLRIFLAARRQ LRLLELLKRIAELLERGDLEEALKLVKKLAKEQGRQEIIDYIEEVLRLYQEGNREEARKLLEELLRRLE KEGDTEFRELIRIILEFLELEERGDLEEAKKLARELKKQVDEARSTLQKEVHAAKSLAIIVGLFALCWL PLHIINCFTFFCPDCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKA genlyfqgvskgeelikenmhmklymegtvnnhhfkctsegegkpyegtqtmrikvveggplpfafdil atsfmygsrtfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvnfp sngpvmqkktlgweantemlypadgglegrsdmalklvggghlicnfkttyrskkpaknlkmpgvyyvd hrlerikeadketyveqhevavarycdlpsklghklnhhhhhhhh
A_{2A}R–FiX2	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAITISTG FCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFAIGLT PMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLGVYLRIFLAARRQ LRLLELLERIARLLKRGDLEEALKLVKKLAKEQGEQEIIDFIEEVLRLYQEGNREQARELLERLLRNLE KRGNQDFRNLIEIILRILELEQRGNQEEIKKLAEELRREVERARSTLQKEVHAAKSLAIIVGLFALCWL PLHIINCFTFFCPDCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKA genlyfqgvskgeelikenmhmklymegtvnnhhfkctsegegkpyegtqtmrikvveggplpfafdil atsfmygsrtfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvnfp sngpvmqkktlgweantemlypadgglegrsdmalklvggghlicnfkttyrskkpaknlkmpgvyyvd hrlerikeadketyveqhevavarycdlpsklghklnhhhhhhhh

Table S3-3 | Amino acid sequences of A_{2A}R fused with FiX1 and FiX2

Amino acid sequences of fusion partner proteins, FiX1 and FiX2, are highlighted by underlines. Glycine spacer + TEV protease recognition site + TagRFP + 8xHis-tag (lowercase) is added at the C-terminal of each designed protein.

	TM5 and TM6 in $A_{2A}R$	N- and C-terminal regions in FiX1 and FiX2
$A_{2A}R$ vs FiX1	R206-L208 and A221-S223	R7-R9 and Q120-K122
$A_{2A}R$ vs FiX2	R206-L208 and R220-R222	R7-R9 and E119-K121

Table S3-4 | Residues used for superposition between $A_{2A}R$ and de novo designed fusion partner proteins

Chapter 4: De novo design of fusion partner proteins to stabilize an active state

Introduction

In **Chapter 3**, I succeeded in stabilizing $A_{2A}R$ in an inactive state by custom-made de novo design of a fusion partner protein. To investigate the potential of the developed strategy to stabilize GPCRs in desired states other than the inactive state, I attempted to stabilize $A_{2A}R$ in an active state using the developed strategy: I designed fusion partner proteins tailored to be fused to TM5 and TM6 of the active-state conformation without kinks or intervening loops, and experimentally characterized them. Experiments for chimeric $A_{2A}R$s with these fusion partner proteins were performed by Murata group, Chiba university.

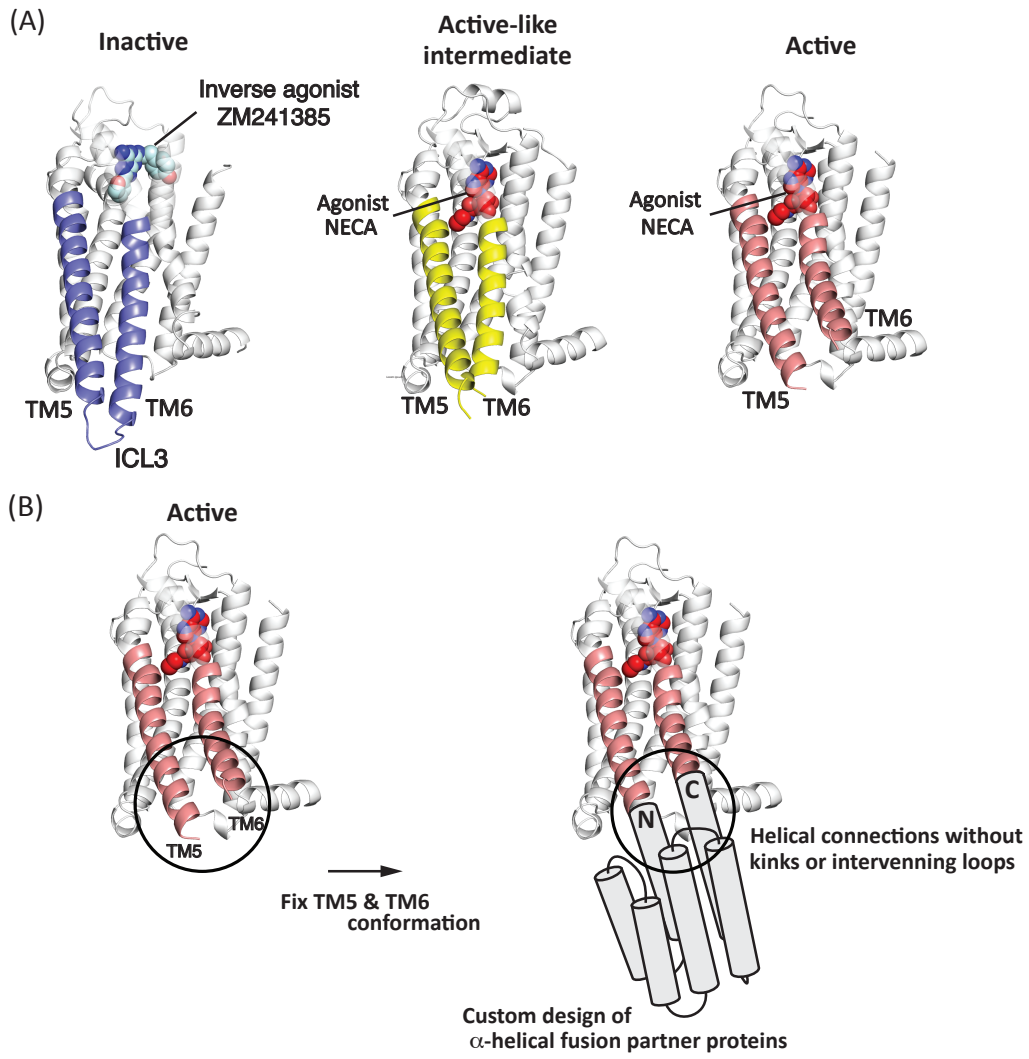

Figure 4-1. Strategy for state-selective stabilization of GPCR, using de novo designed fusion partner proteins. (**A**) Multiple conformational states of $A_{2A}R$. From left, the presented structures respectively correspond to the ones in an inactive, active-like intermediate, and activate states (PDB ID: 3PWH, 2YDV and 5G53, respectively).[54, 69, 85] The structure in the inactive state binds to the inverse agonist ZM241385, and the structures in the active-like intermediate and active states bind to the agonist NECA. ZM241385 and NECA are shown in a sphere model. TM5 and TM6 are colored in blue for the inactive state, yellow for the active-like intermediate state, and pink for the active state. The loop connecting TM5 and TM6 is called ICL3. TM6 with TM5 exhibits large conformational changes upon the state transitions. (**B**) My strategy for the state-selective stabilization of $A_{2A}R$. The TM5 and TM6 conformation in a targeted state was tried to fix through the fusion strategy. To this end, I designed α-helical proteins, which can be fused into $A_{2A}R$ in the targeted state through straight helical connections without kinks or intervening loops. In this work, I tested this idea by stabilizing the active state.

Results

Computational design of α-helical fusion partner proteins

Using the same protocol described in **"Computational design of α-helical fusion partner proteins"**, in **Chapter 3**, I aimed to design fusion partner proteins customized to stabilize the active state of $A_{2A}R$. From the 389 all-α helical backbone models with elongated terminal helices, 36 backbones that are superimposable to TM5 and TM6 of $A_{2A}R$ active state structure (PDB: 5G53), with mainchain RMSD of 0.65 Å or lower, were obtained. Next, sequence design calculations followed by whole structure optimization were performed for the selected backbones. In addition to the set of these designed structures, designs created by the **Step 3** in **Figure 3-2** were added to the set for further obtaining potential fusion partner proteins. In the design set, designed proteins with funnel-like energy landscapes in Rosetta ab initio folding simulations [34] and low fluctuations of the N- and C-terminal helices in MD simulations were further selected for experimental validation. Finally, five designs, FaX1, FaX2, FaX3, FaX4 and FaX5 were created (**FaX** stands for a **F**usion partner protein customized for **a**ctivation and e**X**tra stabilization) (several N-terminal residues of FaX2, FaX3 and FaX4 were truncated for adjusting the length of terminal helices and some residues in FaX1, FaX2, FaX3 and FaX4 were manually mutated using Foldit [72]. See the **Material and Methods**).

Experimental characterization of FaX1, FaX2, FaX3, FaX4 and FaX5

I experimentally characterized the de novo designed proteins, FaX1, FaX2, FaX3, FaX4 and FaX5 without $A_{2A}R$. These designs were expressed in *Escherichia coli* and purified using a Ni-NTA column. All the designs were found to be well expressed and highly soluble and showed CD spectra of all-α proteins from 25 to 98℃, were monomeric in SEC-MALS, and showed well-dispersed sharp NMR peaks. These results indicate that the designs fold into unique α-helical structures as monomers with high thermal stability. For experimental characterizations in $A_{2A}R$-fused forms, FaX3 was selected from the four designs FaX1, FaX2, FaX3 and FaX4 whose sequences were designed from the same backbone structure, because FaX3 showed the least aggregation tendency through the above experimental processes. FaX5, the design using another backbone, was also selected for experimental validation as $A_{2A}R$-fused form.

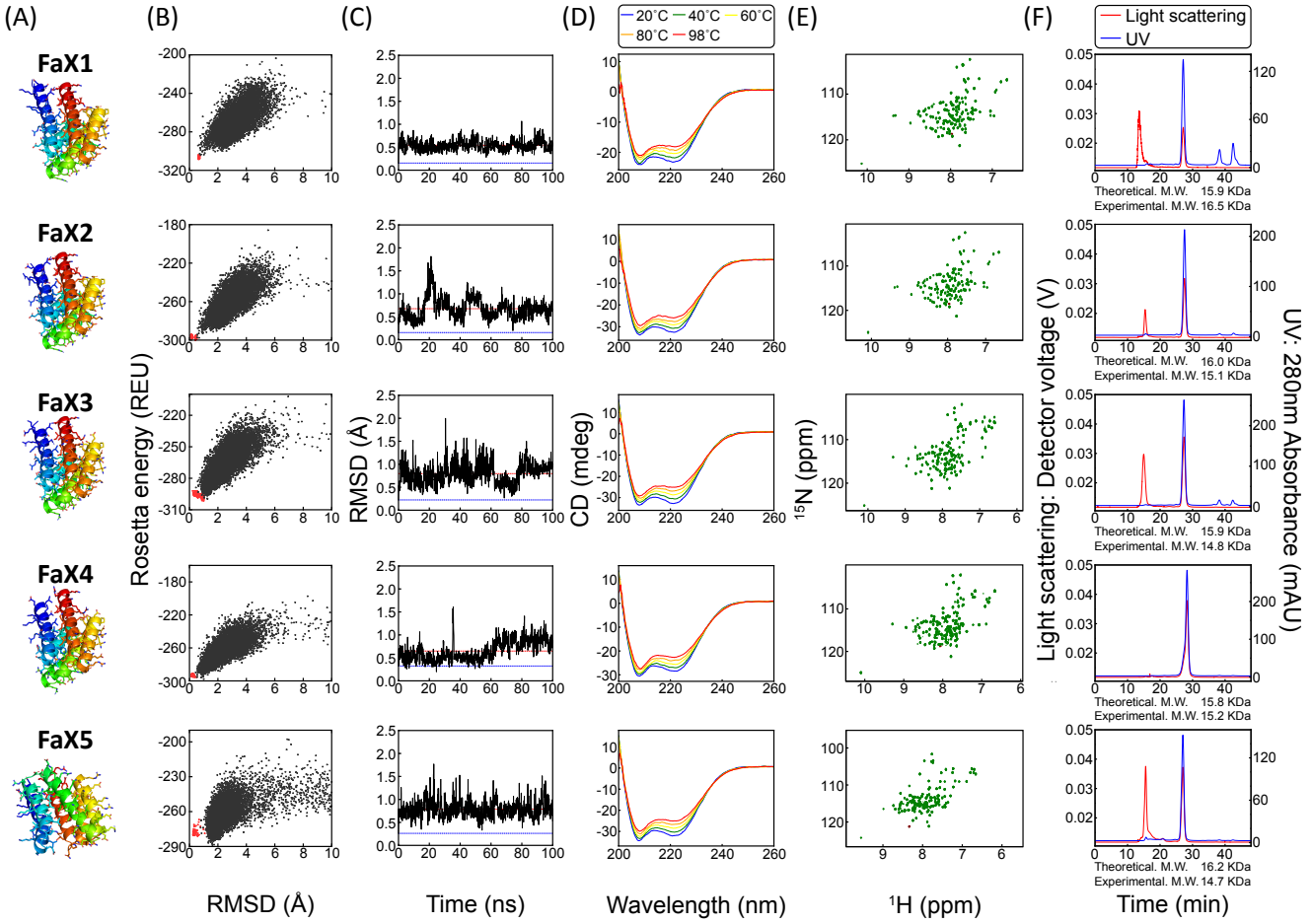

Figure 4-2. Characterization of de novo designed fusion partner proteins without $A_{2A}R$. De novo designed fusion partner proteins, FaX1, FaX2, FaX3, FaX4, and FaX5. (**A**) Computational models. (**B**) The energy landscape of each designed protein obtained from Rosetta folding simulations [86]. Each dot represents the lowest energy structure obtained from an independent trajectory starting from an extended chain (black) or the design model (red), the x-axis shows the Cα RMSD from the design model, and the y-axis shows the Rosetta full-atom energy. (**C**) Structural fluctuations of the N- and C-terminal helices in the MD simulation, starting from each design model. The mainchain RMSD of the N- and C- terminal helices of each snapshot structure during a MD trajectory against TM5 and TM6 of $A_{2A}R$ is shown along the time course. (Red and blue lines are the averaged and initial RMSD values, respectively.) (**D**) Far-ultraviolet CD spectra at various temperatures from 20 to 98°C. (**E**) Two-dimensional ^{1}H-^{15}N HSQC spectra at 25°C and 600 MHz (in parts per million, p.p.m). (**F**) Size-exclusion chromatograms combined with multi-angle light scattering (SEC-MALS) demonstrate that these designed proteins are monomeric in solution.

Experimental characterization of $A_{2A}R$ fused with FaX3 and FaX5

The genes encoding the chimeras with fusion partner proteins, $A_{2A}R$ fused with FaX3 ($A_{2A}R$–FaX3) and FaX5 ($A_{2A}R$–FaX5) were constructed and expressed in yeast as described previously (for each construct, a red fluorescent protein (RFP) was appended at the C-terminus) [66, 73]. As one of the stability metrics, the solubilization efficiency upon detergent extraction was measured. $A_{2A}R$–FaX3 showed significantly improved solubilization efficiency compared to the wild-type; the efficiency was greater than that of $A_{2A}R$–FaX5 (**Table 4-1**). Therefore, the stability of $A_{2A}R$–FaX3 was further studied by measuring the apparent melting temperatures in the clear-native polyacrylamide gel electrophoresis (CN–PAGE) method [66]. The melting temperature was found to be significantly increased, which is same degree as $A_{2A}R$-FiX1: 47.1˚C (**Figure 3-4B** and **Figure 4-3B**) and consistent with the solubilization efficiency results. The solubilization efficiency and melting temperature for $A_{2A}R$–BRIL were comparable to those of $A_{2A}R$–FaX3.

Construct	Solubilization efficiency (%)
$A_{2A}R$ WT	24 ± 10
$A_{2A}R$–BRIL	57 ± 19
$A_{2A}R$–FaX3	53 ± 16
$A_{2A}R$–FaX5	19 ± 10

Table 4-1. Solubilization efficiencies of $A_{2A}R$ fused with or without fusion partner proteins. Results are reported as mean ± standard deviation for n = 3 independent measurements. The shown data for $A_{2A}R$ WT and $A_{2A}R$–BRIL is from the results in **Table 3-1** in **Chapter 3**.

The ligand-binding affinities of $A_{2A}R$–FaX3 using the radioligands of an inverse agonist [^{3}H]-ZM241385 and an agonist [^{3}H]-NECA (see **Materials and Methods**) were also measured. The binding affinity of $A_{2A}R$–FaX3 against ZM241385 and NECA were both not observed. (**Table 4-2** and **Figure 4-3C**). These results indicate that the conformational equilibrium was shifted from the inactive state to a different state by the fusion with FaX3. However, shifting to the active state was not achieved.

Construct	K_d (nM)	
	ZM241385	**NECA**
$A_{2A}R$ WT	10.5 ± 0.3	162 ± 44
$A_{2A}R$–BRIL	15.7 ± 0.6	191 ± 22
$A_{2A}R$–FaX3	N. D.	N. D.

Table 4-2. Dissociation constants (K_d) of $A_{2A}R$ fused with or without fusion partner proteins by saturation binding assay. N. D. stands for not detected. Results are reported as mean ± standard deviation for n = 3 independent assay. The shown data for $A_{2A}R$ WT and $A_{2A}R$–BRIL is from the results in **Table 3-2** in **Chapter 3**.

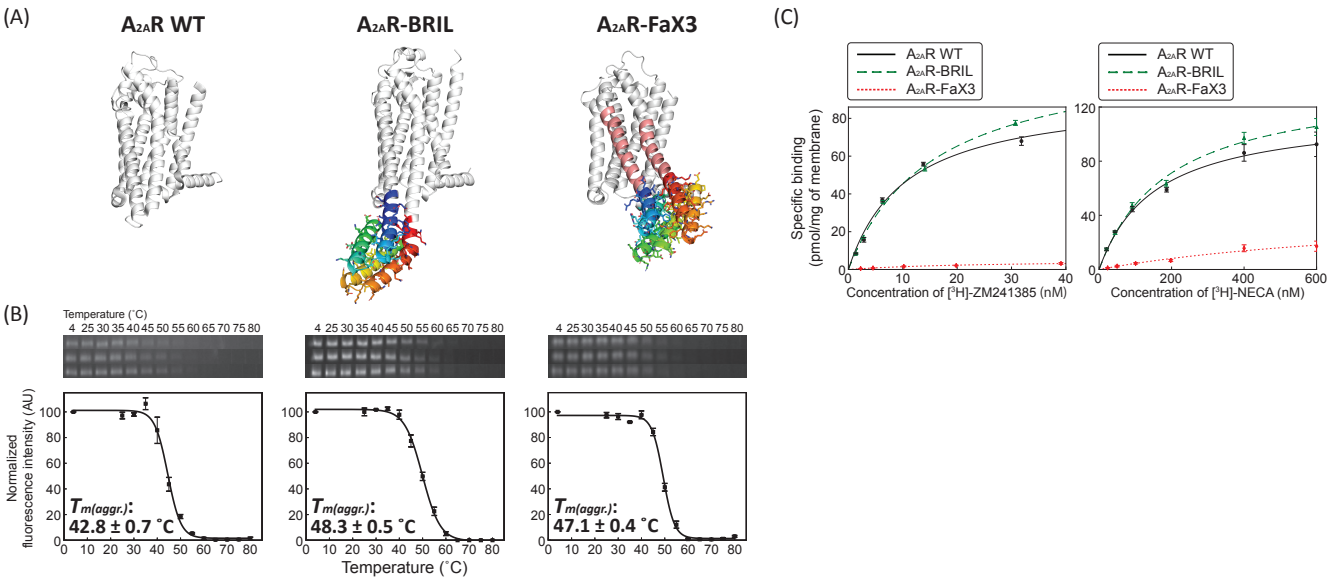

Figure 4-3. Experimental characteristics of $A_{2A}R$ fused with or without fusion partner proteins. (**A**) Crystal structures of $A_{2A}R$ WT (PDB ID: 3VG9) [69], $A_{2A}R$–BRIL (PDB ID: 4EIY) [87], and the computational model of $A_{2A}R$–FaX3. (**B**) (top) Monomer bands in the clear-native PAGE for each $A_{2A}R$ sample heated at various temperatures. (bottom) Fluorescence intensities of the gel images for each $A_{2A}R$ sample with temperature. The thermal transition from soluble to aggregated states was fitted (solid line) to obtain the midpoint temperature, Tm(aggr.). (**C**) Saturation binding curves of [^{3}H]-ZM241385 (left) and [^{3}H]-NECA (right) to $A_{2A}R$ WT (solid line), $A_{2A}R$–BRIL (dashed line), and $A_{2A}R$–FaX3 (dotted line). All measurements were carried out three times independently; dots show the average and whiskers show s.e.m. for n = 3. The shown data for $A_{2A}R$ WT and $A_{2A}R$–BRIL is from **Figure 3-4** in **Chapter 3**.

Discussion

By custom-made design of fusion partner proteins targeting active-state structure of $A_{2A}R$, I succeeded in thermally stabilizing $A_{2A}R$. However, the engineered $A_{2A}R$, $A_{2A}R$–FaX3, was found to be stabilized in neither inactive nor active states. The possible reasons are following:

1) The targeted active-state conformation was not stabilized as the computational model. The targeted $A_{2A}R$ structure (PDB: 5G53) originally contains an engineered G-protein that stabilizes the active state. Only by fusion of FaX3, stabilization of an active state that is originally achieved by G-protein coupling was not realized. Further stabilization of active state is required, possibly achieved not only by stabilization of TM5 and TM6 through straight helical connection but also non-covalent interaction with other transmembrane helices like G-proteins. Considering the success of stabilizing another class A GPCR in active state by designing short ICL3 with mutations[30], additional mutations may also be a strategy to stabilize $A_{2A}R$–FaX3 in an active state. TM5 and TM6 are positioned at different angles between the active state (agonist-bound, G-protein coupling state) and the active-like intermediate state (agonist-bound state with thermostabilizing mutations) (**Figure 4-1**). Therefore, if the purpose of the design is limited to agonist screening, another strategy could be targeting the TM5 and TM6 in the active-like intermediate conformation.

2) FaX3 forms different structure from the designed structure. Very recently, the state-of-the-art structure prediction software AlphaFold2 was developed [88]. For evaluating the agreement of the computational models generated by Rosetta and that by AlphaFold2, I performed structure prediction for FiX1 and FaX3 using AlphaFold2. In the case of FiX1, the mainchain RMSD of N- and C-terminal helices of Rosetta-generated structure and AlphaFold2-predicted structure was 0.365 Å. The Rosetta-generated model showed good agreement with an AlphaFold2-predicted model (**Figure 4-4A**). On the other hand, in the case of FaX3, the mainchain RMSD of N- and C-terminal helices of the Rosetta-generated structure and the AlphaFold2-predicted structure was 1.811 Å; the terminal helices showed

different angles (**Figure 4-4B**). This inconsistency of the model structures suggests that FaX3 structure possibly have different N- and C-terminal helical angle. From the result, FaX3 possibly stabilized TM5 and TM6 in different angle than that of inactive, active-like intermediate or active state of $A_{2A}R$ structure thorough straight helical linkage.

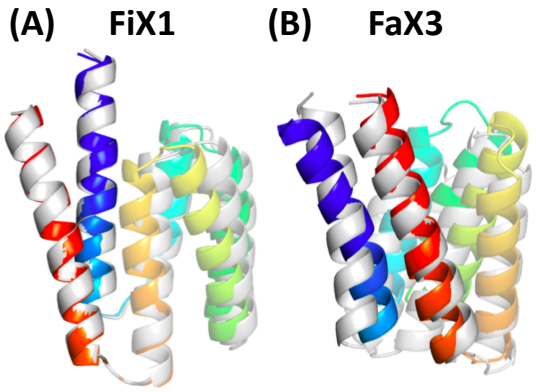

Figure 4-4. Superimposed computational models of (**A**) FiX1 and (**B**) FaX3. Computational models generated by Rosetta are colored in rainbow and AlphaFold2-predicted structures are colored in white. Superimposition was performed for the N- and C-terminal helices.

Materials and Methods

Selection of backbone structure models for fusion partners

The extracted initial set of 1688 backbones and the set of 389 backbones with elongated N- and C-terminal helices were the same as described in "**Selection of backbone structure models for fusion partners**" in **Chapter 3**. Next, among these structures, I selected those whose terminal helices were fusible to TM5 and TM6 of the active state $A_{2A}R$ structure (PDB: 5G53, obtained from the PDB OPM database [64]). To this end, mainchain RMSD values were calculated by superimposing all pairs of three consecutive residues in the N- and C- terminal helices of the backbone structure (the residues are selected from those of the residue number from 2 to 11, and those from 113 to 123, respectively) against all pairs of the three consecutive residues in TM5 and TM6 at the cytoplasmic side (the residues were selected from those of the residue number from 204 to 209, and those from 224 to 229, respectively) (See **Table S4-4**). I selected the backbone structures that were superimposable with a mainchain RMSD value less than or equal to 0.65 Å as fusible ones; 36 backbone structures were obtained.

Sequence design for further backbone selection

I further screened 36 backbone structures via the sequence design of each backbone structure, followed by entire structure optimization, using the FlxbbDesign protocol in Rosetta. The method was same as described in **"Sequence design for further backbone selection"** in **Chapter 3**. Finally, I obtained two backbones.

Sequence design and selection

Sequence design was performed as described in **"Sequence design"** in **Chapter 3**. After the sequence design, the designs were selected by the same criteria as **"Selection criteria after sequence design"** in **Chapter 3**, setting the target $A_{2A}R$ structure as the active-state structure (PDB: 5G53).

In addition to the designed structures from the 36 backbones, the selection criteria were also tested for the protein structures designed in **Step 3** in **Figure 3-2**, originally made for $A_{2A}R$ structure of the inactive state. From these designed proteins, candidate designed proteins were also investigated with the same criteria.

Rosetta folding simulation and molecular dynamics (MD) simulation

Rosetta folding simulation and molecular dynamics simulation were performed as described in **"Rosetta folding simulation"** and **"Molecular dynamics (MD) simulation"** in **Chapter 3**.

Manual truncation of N-terminal residues and mutations using Foldit

For FaX1, to avoid unexpected polar contacts observed in MD simulation, Thr at the position 23 was mutated to Glu (T23E). Ile residue at residue position 68 was mutated to Arg to eliminate exposed hydrophobic residue (I68R).

To eliminate possibly disordered region at N-terminal helix by the lack of sufficient contact against the C-terminal helix, several residues at the N-terminal helix were manually truncated. Residue number 1 to 4, 1 to 3 and 1 to 5 were removed from FaX2, FaX3 and FaX4, respectively. This deletion of the residues was done by text editing for these PDB files.

FaX2 is mutation variant of FaX1. In addition to T23E and I68R mutations corresponding to T19E and I64R in FaX2, two Arg residues at the position 2 and 4 on the new last three residues in N-terminal helix by the terminal truncation were mutated to Glu, to avoid electrostatic repulsion between the charged

sidechain and helix dipole. To make salt bridges to interact with these two Glu residues, Lys residues at position 5 and 8 were mutated to Arg.

For FaX3, to avoid electrostatic repulsion between the sidechain and helix dipole, Arg at residue position 3 was mutated to Glu. To make a salt bridge with the Glu residue, mutation was introduced to alter Lys at residue position 7 to Arg. To eliminate unexpected electrostatic interaction observed in MD simulation, Arg at residue position 5 was mutated to Lys. For appropriate hydrophobic packing between helices, Glu at residue postion 8 was mutated to Ile and the residues that originally made salt bridges with the Glu, Arg at residue position 12 and 118 were both altered to Gln.

For FaX4, considering the repulsion between sidechains and helix dipole, residues at position 3 and 4 were mutated from Arg to Glu and Gln, respectively. Lys at residue position 5 was mutated to Glu to avoid unexpected formation of salt bridge observed in MD simulation and Glu at residue position 6 was mutated to Ile for hydrophobic packing between helices.

Experiments of de novo designed fusion partner proteins

FaX1, FaX2, FaX3, FaX4 and FaX5 were expressed, purified and tested as described in the sections start with **"Experiments of de novo designed fusion partner proteins:"** in **Chapter 3**.

For ^{15}N-labeled FaX1, FaX2, FaX3, FaX4 and FaX5 samples of 400 to 650 μM in 90% H_2O/10% D_2O PBS buffer (pH 7.4), 2D HSQC NMR spectrum measurements were performed using a JNM-ECA 600 MHz spectrometer (JEOL, Tokyo, Japan). The obtained NMR spectra were analyzed using the Delta NMR software (https://nmrsupport.jeol.com/Software) (version 5.2.1, JEOL, Tokyo, Japan).

Experiments of A_{2A}R-designed fusion partner proteins

A_{2A}R–FaX3 and A_{2A}R–FaX5 were expressed and characterized as described in the sections start with **"Experiments of A_{2A}R-designed fusion partner proteins:"** in **Chapter 3**.

Supplementary figures and tables

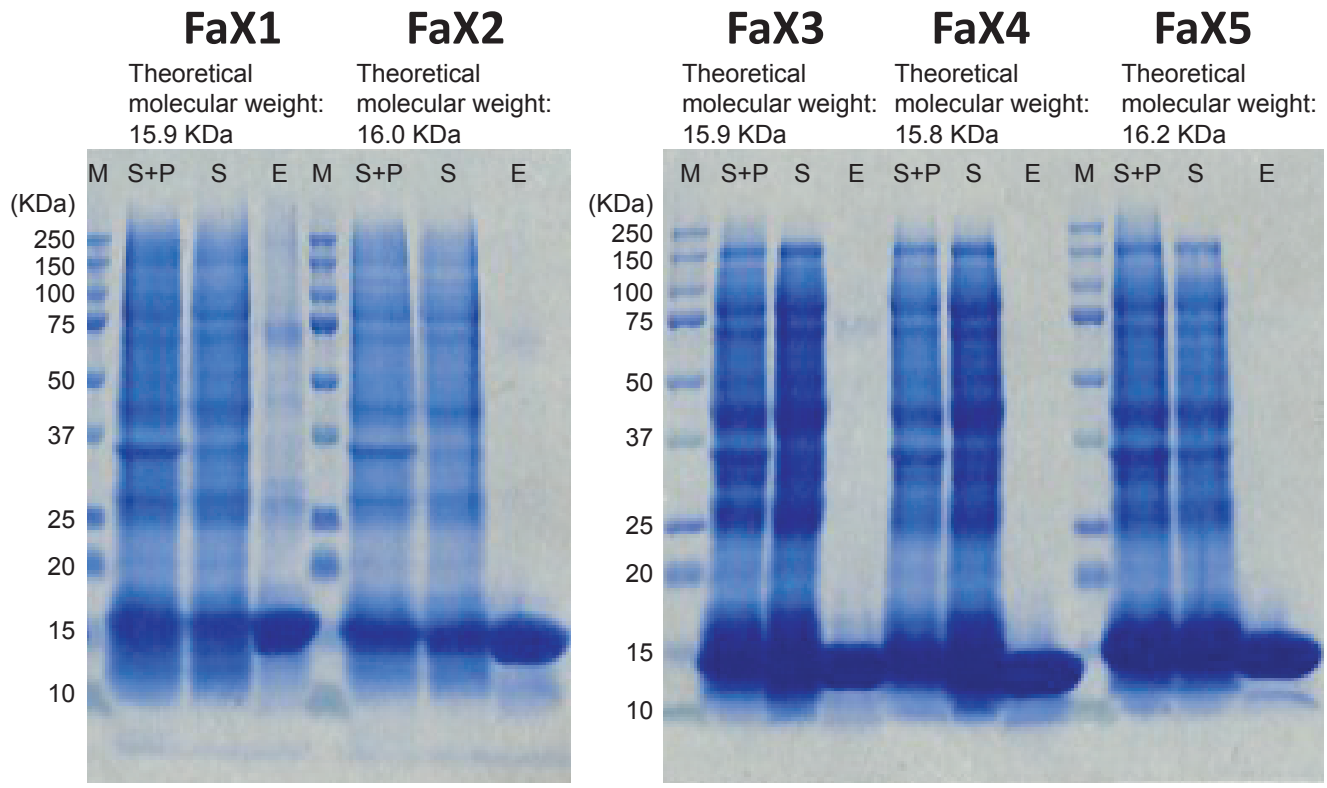

Figure S4-1 | SDS-PAGE results for FaX1, FaX2, FaX3, FaX4 and FaX5

M, S+P, S and E indicate marker, supernatant+pellet, supernatant and elution, respectively.

Construct	Truncated regions in $A_{2A}R$	Inserted regions in fusion partner proteins
$A_{2A}R$–BRIL	K209 - G218	A1 - L106
$A_{2A}R$–FaX3	Q207 - Q226	E11 - K113
$A_{2A}R$–FaX5	Q210 - L225	K8 - I120

Table S4-1 | Sequence regions for $A_{2A}R$ and fusion partner proteins in $A_{2A}R$ chimeras

Construct	Sequence
FaX1	mGEDEKRLRKEEKRLRRILEAFREGDLQEAARELADLAQEKNPNNKNVEKLVREVRELIRRGDRENAFRK ALKALIEVAKDEGDPELEELLKRILKLFEEGDRENAIKLLKKFLEEEIKKLRKKGSWSlehhhhhh
FaX2	mGELEREERRLRRILEAFREGDLQEAARELADLAQEKNPNNKNVEKLVREVRELIRRGDRENAFRKALKA LIEVAKDEGDPELEELLKRILKLFEEGDRENAIKLLKKFLEEEIKKLRKKGSWSlehhhhhh
FaX3	mGEELKRRIKELQRILEAFKTGDLQEAARLLADLAQKKNPNNKNVNELVRQVRELIKRGDRENAFIKALQ ALIEVAKDEGDPELEELLKRILKLFQEGDRDNAIELLRKFLEKEIKKLQKNGSWSlehhhhhh
FaX4	mGLEQQIKELRRILEAFRTGDLQEAARLLAELARRKNPNNKNVEELVRRVEELIRRGDRENAFIEALRAL IEVAKDEGDPDLEELLKRILELFQRGNREDAIKLLKEFLEREIKKLRKNGSWSlehhhhhh
FaX5	mGEDDLKKEAKERVREALELLKKEGNLDLLELLKRLLELLQRGDIEEFRKLLQKLLDELLKQAKKEGERE IYEYIKRVKELLDRGDTEEAEKRAEELQRKRPNNEIVQVLKALVDLFKIIERLRGlehhhhhh

Table S4-2 | Amino acid sequences of FaX1, FaX2, FaX3, FaX4 and FaX5

Computationally designed amino acid sequences are described in uppercase and residues added to allow expression, purification, and the spacers between a designed sequence and the C-terminal His-tag are described in lowercase.

Construct	Sequence
A_{2A}R–FaX3	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAITIS TGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFA IGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLGVYLRIF LAARRELQRILEAFKTGDLQEAARLLADLAQKKNPNNKNVNELVRQVRELIKRGDRENAFIKALQAL IEVAKDEGDPELEELLKRILKLFQEGDRDNAIELLRKFLEKKEVHAAKSLAIIVGLFALCWLPLHII NCFTFFCPDCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRSHVLRQQEPFKAgen lyfqgvskgeelikenmhmklymegtvnnhhfkctsegegkpyegtqtmrikvveggplpfafdila tsfmygsrtfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqdgcliynvkirgvnf psngpvmqkktlgweantemlypadgglegrsdmalklvggghlicnfkttyrskkpaknlkmpgvy yvdhrlerikeadketyveqhevavarycdlpsklghklnhhhhhhhh
A_{2A}R–FaX5	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAIPFAITIS TGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTRAKGIIAICWVLSFA IGLTPMLGWNNCGQPKEGKQHSQGCGEGQVACLFEDVVPMNYMVYFNFFACVLVPLLLMLGVYLRIF LAARRQLKKEAKERVREALELLKKEGNLDLLELLKRLLELLQRGDIEEFRKLLQKLLDELLKQAKKE GEREIYEYIKRVKELLDRGDTEEAEKRAEELQRKRPNNEIVQVLKALVDLFKIIQKEVHAAKSLAII VGLFALCWLPLHIINCFTFFCPDCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFRKIIRS HVLRQQEPFKAgenlyfqgvskgeelikenmhmklymegtvnnhhfkctsegegkpyegtqtmrikv veggplpfafdilatsfmygsrtfinhtqgipdffkqsfpegftwervttyedggvltatqdtslqd gcliynvkirgvnfpsngpvmqkktlgweantemlypadgglegrsdmalklvggghlicnfkttyr skkpaknlkmpgvyyvdhrlerikeadketyveqhevavarycdlpsklghklnhhhhhhhh

Table S4-3 | Amino acid sequences of A_{2A}R fused with FaX3 and FaX5

Amino acid sequences of fusion partner proteins, FaX3 and FaX5, are highlighted by underlines. Glycine spacer + TEV protease recognition site + TagRFP + 8xHis-tag (lowercase) is added at the C-terminal of each designed protein.

	TM5 and TM6 in $A_{2A}R$	N- and C-terminal regions in FaX3 and FaX5
$A_{2A}R$ vs FaX3	A204-R206 and K227-V229	R8-K10 and E114-K116
$A_{2A}R$ vs FaX5	Q207-K209 and Q226-E228	D5-K7 and E121-L123

Table S4-4 | Residues used for superposition between $A_{2A}R$ and de novo designed fusion partner proteins

Chapter 5: Conclusion

I attempted to develop methodologies for state-selective stabilization of GPCRs, based on computational de novo protein design technology.

In **Chapter 2**, I aimed to state-selectively stabilize $A_{2A}R$ by redesigning TM5-ICL3-TM6, using short and typical loop structures. However, the proposed method was not effective enough to stabilize $A_{2A}Rs$. For the redesigned $A_{2A}Rs$ to be stabilized in an inactive state, possible reasons of the insufficient stability are loss of helices or hydrophobic interaction formed in the original structure or low sequence-structure compatibility of the designed ICL3 region. For the $A_{2A}Rs$ redesigned to be stabilized in an active state, one of the reasons is that the redesigned TM5-ICL3-TM6 region did not fully compensate for the decreased stability caused by the absence of G-protein. Both cases suggest that more stability of the redesigned region is required to stabilize $A_{2A}R$.

To further stabilize the region between TM5 and TM6, in **Chapter 3**, I computationally designed super-stable fusion partner proteins customized to stabilize $A_{2A}R$ in an inactive state. The designed fusion partner proteins experimentally validated to form helical structures even under 98˚C. Moreover, $A_{2A}R$ fused with one of the designed fusion partner proteins, named FiX1, was successfully stabilized in an inactive state.

In **Chapter 4**, to investigate the potential of the methodology developed in **Chapter 3**, I computationally designed super-stable fusion partner proteins customized to stabilize $A_{2A}R$ in an active state. The designed fusion partner proteins exhibited super-stabilities. In addition, by fusion of a designed protein named FaX3, $A_{2A}R$ was thermally stabilized. However, the stabilized state was neither inactive and active state. Design of fusion partner proteins interacting with other region of GPCR to fix $A_{2A}R$ in an active state would be required to improve the method.

I achieved the state-selective stabilization of GPCRs by custom-made design of super-stable fusion partner proteins (**Chapter 3**). The developed method still requires improvements for stabilizing an active

state (**Chapter 4**). However, the developed method is expected to dedicate on rationally designing stabilized GPCRs for structure determination in a specific state or state-dependent ligand/antibody screening. Moreover, to the best of my knowledge, this is the first example of a state-selective stabilization of proteins including GPCRs only by rational mainchain-level engineering. The developed method may be applied for other proteins undergoing conformational changes.

References

[1] Lagerstrom MC, Schioth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. Nature Reviews Drug Discovery. 2008;7:339-57.
[2] Weis WI, Kobilka BK. The Molecular Basis of G Protein-Coupled Receptor Activation. In: Kornberg RD, editor. Annual Review of Biochemistry, Vol 872018. p. 897-919.
[3] Alexander SPH, Christopoulos A, Davenport AP, Kelly E, Mathie A, Peters JA, et al. THE CONCISE GUIDE TO PHARMACOLOGY 2019/20: G protein-coupled receptors. British Journal of Pharmacology. 2019;176:S21-S141.
[4] Munk C, Isberg V, Mordalski S, Harpsoe K, Rataj K, Hauser AS, et al. GPCRdb: the G protein-coupled receptor database - an introduction. British Journal of Pharmacology. 2016;173:2195-207.
[5] Yang DH, Zhou QT, Labroska V, Qin SS, Darbalaei S, Wu YR, et al. G protein-coupled receptors: structure- and function-based drug discovery. Signal Transduction and Targeted Therapy. 2021;6.
[6] Syrovatkina V, Alegre KO, Dey R, Huang XY. Regulation, Signaling, and Physiological Functions of G-Proteins. Journal of Molecular Biology. 2016;428:3850-68.
[7] Lin Y, Smrcka AV. Understanding Molecular Recognition by G protein beta gamma Subunits on the Path to Pharmacological Targeting. Molecular Pharmacology. 2011;80:551-7.
[8] Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nature Reviews Drug Discovery. 2017;16:19-34.
[9] Heydenreich FM, Vuckovic Z, Matkovic M, Veprintsev DB. Stabilization of G protein-coupled receptors by point mutations. Frontiers in Pharmacology. 2015;6.
[10] Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. Conformational thermostabilization of the beta 1-adrenergic receptor in a detergent-resistant form. Proceedings of the National Academy of Sciences of the United States of America. 2008;105:877-82.
[11] Magnani F, Shibata Y, Serrano-Vega MJ, Tate CG. Co-evolving stability and conformational homogeneity of the human adenosine A(2a) receptor. Proceedings of the National Academy of Sciences of the United States of America. 2008;105:10744-9.
[12] Hollenstein K, Kean J, Bortolato A, Cheng RKY, Dore AS, Jazayeri A, et al. Structure of class B GPCR corticotropin-releasing factor receptor 1. Nature. 2013;499:438-43.
[13] Jazayeri A, Dore AS, Lamb D, Krishnamurthy H, Southall SM, Baig AH, et al. Extra-helical binding site of a glucagon receptor antagonist. Nature. 2016;533:274-7.
[14] Oswald C, Rappas M, Kean J, Dore AS, Errey JC, Bennett K, et al. Intracellular allosteric antagonism of the CCR9 receptor. Nature. 2016;540:462-5.
[15] Cheng RKY, Fiez-Vandal C, Schlenker O, Edman K, Aggeler B, Brown DG, et al. Structural insight into allosteric modulation of protease-activated receptor 2. Nature. 2017;545:112-5.
[16] Robertson N, Rappas M, Dore AS, Brown J, Bottegoni G, Koglin M, et al. Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. Nature. 2018;553:111-4.
[17] Popov P, Peng Y, Shen L, Stevens RC, Cherezov V, Liu ZJ, et al. Computational design of thermostabilizing point mutations for G protein-coupled receptors. Elife. 2018;7:e34729.
[18] Yasuda S, Kajiwara Y, Toyoda Y, Morimoto K, Suno R, Iwata S, et al. Hot-Spot Residues to be Mutated Common in G Protein-Coupled Receptors of Class A: Identification of Thermostabilizing Mutations Followed by Determination of Three-Dimensional Structures for Two Example Receptors. Journal of Physical Chemistry B. 2017;121:6341-50.

[19] Chen KYM, Zhou FG, Fryszczyn BG, Barth P. Naturally evolved G protein-coupled receptors adopt metastable conformations. Proceedings of the National Academy of Sciences of the United States of America. 2012;109:13284-9.
[20] Warne T, Chirnside J, Schertler GFX. Expression and purification of truncated, non-glycosylated turkey beta-adrenergic receptors for crystallization. Biochimica Et Biophysica Acta-Biomembranes. 2003;1610:133-40.
[21] Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, et al. Fusion Partner Toolchest for the Stabilization and Crystallization of G Protein-Coupled Receptors. Structure. 2012;20:967-76.
[22] Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, et al. High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor. Science. 2007;318:1258-65.
[23] Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, et al. GPCR engineering yields high-resolution structural insights into beta(2)-adrenergic receptor function. Science. 2007;318:1266-73.
[24] 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.
[25] Tan QX, Zhu Y, Li J, Chen ZX, Han GW, Kufareva I, et al. Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. Science. 2013;341:1387-90.
[26] Yin J, Mobarec JC, Kolb P, Rosenbaum DM. Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. Nature. 2015;519:247-50.
[27] Gurevich VV, Gurevich EV. GPCR Signaling Regulation: The Role of GRKs and Arrestins. Frontiers in Pharmacology. 2019;10.
[28] Kang YY, Zhou XE, Gao X, He YZ, Liu W, Ishchenko A, et al. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. Nature. 2015;523:561-+.
[29] Zhou QT, Yang DH, Wu M, Guo Y, Guo WJ, Zhong L, et al. Common activation mechanism of class A GPCRs. Elife. 2019;8.
[30] Yin J, Chen KYM, Clark MJ, Hijazi M, Kumari P, Bai XC, et al. Structure of a D2 dopamine receptor-G-protein complex in a lipid membrane. Nature. 2020;584:125-9.
[31] Dahiyat BI, Mayo SL. De novo protein design: Fully automated sequence selection. Science. 1997;278:82-7.
[32] Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science. 1998;282:1462-7.
[33] Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science. 2003;302:1364-8.
[34] Koga N, Tatsumi-Koga R, Liu GH, Xiao R, Acton TB, Montelione GT, et al. Principles for designing ideal protein structures. Nature. 2012;491:222-7.
[35] Huang PS, Feldmeier K, Parmeggiani F, Velasco DAF, Hocker B, Baker D. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nature Chemical Biology. 2016;12:29-34.
[36] Brunette TJ, Parmeggiani F, Huang PS, Bhabha G, Ekiert DC, Tsutakawa SE, et al. Exploring the repeat protein universe through computational protein design. Nature. 2015;528:580-4.
[37] Dou JY, Vorobieva AA, Sheffler W, Doyle LA, Park H, Bick MJ, et al. De novo design of a fluorescence-activating beta-barrel. Nature. 2018;561:485-91.
[38] Doyle L, Hallinan J, Bolduc J, Parmeggiani F, Baker D, Stoddard BL, et al. Rational design of alpha-helical tandem repeat proteins with closed architectures. Nature. 2015;528:585-8.

[39] Marcos E, Basanta B, Chidyausiku TM, Tang YF, Oberdorfer G, Liu GH, et al. Principles for designing proteins with cavities formed by curved beta sheets. Science. 2017;355:201-6.
[40] Marcos E, Chidyausiku TM, McShan AC, Evangelidis T, Nerli S, Carter L, et al. De novo design of a non-local beta-sheet protein with high stability and accuracy. Nature Structural & Molecular Biology. 2018;25:1028-34.
[41] Hosseinzadeh P, Bhardwaj G, Mulligan VK, Shortridge MD, Craven TW, Pardo-Avila F, et al. Comprehensive computational design of ordered peptide macrocycles. Science. 2017;358:1461-6.
[42] Malakauskas SM, Mayo SL. Design, structure and stability of a hyperthermophilic protein variant. Nature Structural Biology. 1998;5:470-5.
[43] Korkegian A, Black ME, Baker D, Stoddard BL. Computational thermostabilization of an enzyme. Science. 2005;308:857-60.
[44] Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, et al. Proof of principle for epitope-focused vaccine design. Nature. 2014;507:201-6.
[45] Silva DA, Yu S, Ulge UY, Spangler JB, Jude KM, Labao-Almeida C, et al. De novo design of potent and selective mimics of IL-2 and IL-15. Nature. 2019;565:186-91.
[46] Linsky TW, Vergara R, Codina N, Nelson JW, Walker MJ, Su W, et al. De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. Science. 2020;370:1208-14.
[47] Borea PA, Gessi S, Merighi S, Vincenzi F, Varani K. Pharmacology of Adenosine Receptors: The State of the Art. Physiological Reviews. 2018;98:1591-625.
[48] Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, et al. Crystal structure of rhodopsin: A G protein-coupled receptor. Science. 2000;289:739-45.
[49] Rasmussen SGF, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, et al. Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor. Nature. 2007;450:383-7.
[50] Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, et al. Structure of a beta(1)-adrenergic G-protein-coupled receptor. Nature. 2008;454:486-91.
[51] Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EYT, 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.
[52] Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, et al. Structure of an Agonist-Bound Human A(2A) Adenosine Receptor. Science. 2011;332:322-7.
[53] Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, et al. Agonist-bound adenosine A(2A) receptor structures reveal common features of GPCR activation. Nature. 2011;474:521-5.
[54] Carpenter B, Nehmé R, Warne T, Leslie AGW, Tate CG. Structure of the adenosine A2A receptor bound to an engineered G protein. Nature. 2016;536:104-7.
[55] Garcia-Nafria J, Lee Y, Bai XC, Carpenter B, Tate CG. Cryo-EM structure of the adenosine A(2A) receptor coupled to an engineered heterotrimeric G protein. Elife. 2018;7.
[56] Rasmussen SGF, DeVree BT, Zou YZ, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the beta(2) adrenergic receptor-Gs protein complex. Nature. 2011;477:549-55.
[57] Kooistra AJ, Mordalski S, Pandy-Szekeres G, Esguerra M, Mamyrbekov A, Munk C, et al. GPCRdb in 2021: integrating GPCR sequence, structure and function. Nucleic Acids Research. 2021;49:D335-D43.
[58] Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537:320-7.

[59] Koga R, Yamamoto M, Kosugi T, Kobayashi N, Sugiki T, Fujiwara T, et al. Robust folding of a de novo designed ideal protein even with most of the core mutated to valine. Proceedings of the National Academy of Sciences of the United States of America. 2020;117:31149-56.
[60] Lin YR, Koga N, Tatsumi-Koga R, Liu G, Clouser AF, Montelione GT, et al. Control over overall shape and size in de novo designed proteins. Proceedings of the National Academy of Sciences of the United States of America. 2015;112:E5478-E85.
[61] Huang PS, Ban YEA, Richter F, Andre I, Vernon R, Schief WR, et al. RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. Plos One. 2011;6:e24109.
[62] Dore 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.
[63] Carpenter B, Nehme R, Warne T, Leslie AGW, Tate CG. Structure of the adenosine A(2A) receptor bound to an engineered G protein. Nature. 2016;536:104-7.
[64] Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Research. 2012;40:D370-D6.
[65] O'Meara MJ, Leaver-Fay A, Tyka MD, Stein A, Houlihan K, DiMaio F, et al. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. Journal of Chemical Theory and Computation. 2015;11:609-22.
[66] Suzuki N, Takamuku Y, Asakawa T, Inai M, Hino T, Iwata S, et al. An efficient screening method for purifying and crystallizing membrane proteins using modified clear -native PAGE. Analytical Biochemistry. 2018;548:7-14.
[67] Jo M, Jung ST. Engineering therapeutic antibodies targeting G-protein-coupled receptors. Experimental and Molecular Medicine. 2016;48.
[68] Sakuma K, Kobayashi N, Sugiki T, Nagashima T, Fujiwara T, Suzuki K, et al. Design of complicated all-α protein structures. bioRxiv. 2021:2021.07.14.449347.
[69] Doré Andrew S, Robertson N, Errey James C, Ng I, Hollenstein K, Tehan B, et al. Structure of the Adenosine A2A Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine. Structure. 2011;19:1283-93.
[70] Lebon G, Edwards PC, Leslie AGW, Tate CG. Molecular Determinants of CGS21680 Binding to the Human Adenosine A(2A) Receptor. Molecular Pharmacology. 2015;87:907-15.
[71] Sheffler W, Baker D. RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein Science. 2009;18:229-39.
[72] Kleffner R, Flatten J, Leaver-Fay A, Baker D, Siegel JB, Khatib F, et al. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. Bioinformatics. 2017;33:2765-7.
[73] Shiroishi M, Tsujimoto H, Makyio H, Asada H, Yurugi-Kobayashi T, Shimamura T, et al. Platform for the rapid construction and evaluation of GPCRs for crystallography in Saccharomyces cerevisiae. Microbial Cell Factories. 2012;11.
[74] Ye LB, Van Eps N, Zimmer M, Ernst OP, Prosser RS. Activation of the A(2A) adenosine G-protein-coupled receptor by conformational selection. Nature. 2016;533:265-8.
[75] Huang SK, Pandey A, Tran DP, Villanueva NL, Kitao A, Sunahara RK, et al. Delineating the conformational landscape of the adenosine A(2A) receptor during G protein coupling. Cell. 2021;184:1884-94.
[76] 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.

[77] Chu R, Takei J, Knowlton JR, Andrykovitch M, Pei WH, Kajava AV, et al. Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. Journal of Molecular Biology. 2002;323:253-62.
[78] Mitternacht S. FreeSASA: An open source C library for solvent accessible surface area calculations. F1000Research. 2016;5.
[79] Ballesteros JA, Deupi X, Olivella M, Haaksma EEJ, Pardo L. Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. Biophysical Journal. 2000;79:2754-60.
[80] Baker EG, Bartlett GJ, Crump MP, Sessions RB, Linden N, Faul CFJ, et al. Local and macroscopic electrostatic interactions in single alpha-helices. Nature Chemical Biology. 2015;11:221-8.
[81] Case DA, Betz RM, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, et al. AMBER 2016. University of California, San Francisco. 2016.
[82] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics. 1983;79:926-35.
[83] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Structure Function and Bioinformatics. 2006;65:712-25.
[84] Newstead S, Kim H, von Heijne G, Iwata S, Drew D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:13936-41.
[85] Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, et al. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. Nature. 2011;474:521-5.
[86] Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, et al. Principles for designing ideal protein structures. Nature. 2012;491:222.
[87] 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 (New York, NY). 2012;337:232-6.
[88] 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.

Acknowledgement

First of all, I deeply appreciate for all the mentoring by Associate Prof. Nobuyasu Koga, Assistant Prof. Takahiro Kosugi and Dr. Rie Koga, not only for scientific and technical coaching but also for life guidance and mental care. This thesis is definitely made of all their supports. I am so grateful for Prof. Takeshi Murata for collaboration, experimental validation of designed GPCRs and letting me start my career as a protein design scientist for applicational use in pharmacology-related sciences. All the characteristics of designed A_{2A}Rs were validated by Ryosuke Nakano, Nanao Suzuki, Kanna Sugaya and Kazuki Kazama in Murata group at Chiba University. I appreciate all their works. I appreciate Dr. Koya Sakuma for the initial set of all-α helical backbone structures. His behavior to pursue his ideal scientist looked so harsh (honestly…), but it is definitely inspiring for me. I appreciate Dr. Shintaro Minami for sharing useful computational tools for protein design. I learned a lot from his way of doing research with being so kind to other people that finally achieved impressive results. I thank the Functional Genomics Facility, NIBB Core Research Facilities, especially Yumiko Makino, for mass spectrometry analysis, and the Instrument Center, Okazaki, for HSQC NMR measurements. From Oct 2019 to Sep 2020, I learned a lot through one-year internship in Neoleukin Therapeutics, Inc. I deeply appreciate for the people I involved with there, starting from Dr. Daniel Adriano Silva Manzano and Dr. Thomas Linsky for giving me such wonderful opportunity to experience cutting-edge protein design for pharmacological use. I am so grateful all the acceptance and kindness from all the people I involved in Neoleukin. For the internship, I was financially supported by SOKENDAI Student Dispatch Program. During my study as a Ph. D. course student, I was financially supported by RA and SRA hiring system in IMS and scholarship by Daiko foundation. I really appreciate these supports for being able to research with physical and mental health. At the end of this thesis and my long-term school life, I deeply appreciate for my parents for letting me free with so much supports and patience, from the bottom of my heart.