Doctoral Dissertation

博士論文

Ligand recognition and activation mechanism of class A GPCRs

（Class A GPCR におけるリガンド認識および活性化機構）

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2020

令和 2 年 12 月　博士（理学）申請

Department of Biological Sciences, Graduate School of Science,
The University of Tokyo

東京大学大学院理学系研究科生物科学専攻

Tamaki Izume

井爪　珠希

Abstract

Class A GPCRs are the most diverse subfamily of GPCRs in humans and are well-studied as drug targets. The endothelin receptor (ETR) is a class A GPCR that uses the peptide hormone endothelin as a ligand and is involved in a variety of cellular responses such as blood pressure maintenance in vascular endothelial cells. There are three types of endothelin, ET-1, ET-2, and ET-3, and two types of ETR, type A and type B. ET_A receptor (ETAR) induces vasoconstriction, whereas ET_B receptor (ETBR) induces vasorelaxation. Thus, B-type selective agonists are of interest for the treatment of hypertension, heart disease, Alzheimer's disease, and for improving drug delivery to cancer cells.

The ligand-free and ET-1-bound states of ETBR have been used to elucidate the binding mode and conformational changes of ETBR and endothelin. However, since ET-1 is an agonist that binds to both types A and B to the same level, the binding mode and receptor activation mechanism of the type B-selective agonist have been unclear. In this study, I determined the steric structure of the complexes of ETBR with three type B-selective agonists, ET-3 and IRL1620, and sarafotoxin S6b (S6b), which has a weak affinity for ETAR. I discussed (1) the molecular basis of ETBR agonist binding and B-type selectivity, (2) the details of ETBR activation mechanism, and (3) the mechanism of ETBR partial activation of IRL1620.

Lysophosphatidylserine (LysoPS), which is produced when phosphatidylserine, PS, undergoes deacylation by phospholipase A_2 or PS-selective phospholipase A_1. LysoPS is a lipid mediator and induces inflammatory responses. LPS1, 2, and 3 are class A GPCRs that accept LysoPS and thought to belong to the P2Y family. LPS1/GPR34 is the first LysoPS receptor discovered and is known to induce cellular responses by accepting G_i. The contribution of LPS1 *in vivo* has long been unknown, however, its involvement in neuropathic pain was reported in 2019. LPS1 antagonists are expected to be applied to drugs for neuropathic pain. The purpose of this study was to obtain ligand-binding information for the discovery of LPS1 antagonists by structural analysis of LPS1 and LysoPS analogs in G_i complexes using cryo-electron microscopy, to understand how lipid ligands enter the cell membrane from inside or outside the cell, and to elucidate the mechanism of Gi activation by LysoPS analogs. As a result, I successfully purified the complex of LPS1-G_i-scFv16, and obtained the 3.6 Å complex structure by using cryo-electron microscopy (cryo-EM).

In this study, focusing on the ligand binding mode of class A GPCR, I performed X-ray crystallography and cryo-electron microscopy of peptide receptor ETBR, and lipid-sensing receptor LPS1, respectively. From the crystal structures of ETBR and the agonist complex, I discussed the B-type selectivity from the ligand binding mode. The crystal structures of ETBR and the agonist complex were discussed in terms of ligand binding mode and B-type selectivity. Cryo-EM structure of LPS1, its agonist, and Gi complex were discussed in terms of agonist binding sites based on the structures of other lipid-sensing class A GPCRs.

Contents

Table of abbreviations

Abbreviations	Full name
CHS	Cholesteryl hemisuccinate
DDM	n-dodecyl-β-D-maltopyranoside
EC_{50}	half maximal (50%) effective concentration
EGTA	(ethyleneglycol)tetraacetic acid
E_{max}	Maximum responce
GDN	Glyco-diosgenin
GFP	Green fluorescent protein
$GnTI^-$	N-acetylglucosaminyl-transferase I-negative
GPCR	G protein-coupled receptor
LCP	Lipidic Cubic Phase
LMNG	Lauryl Maltose Neopentyl Glycol
MES	2-(N-morpholino)ethanesulfonic acid
MD	molecular dynamics
PDB	Protein Data Bank
PEG	Polyethylene glycol
P2Y	Purinergic 2Y receptor
POPC	1-palmitoyl-2-oleoyl-phosphatidylcholine
RMSD	root mean square deviation
Sf9	*Spodoptera frugiperda* 9
TALON	TALON® metal affinity resin
TCEP	Tris(2-carboxyethyl)phosphine
TEV	tobacco etch virus
TM	transmembrane
Tris	tris(hydroxymethyl)aminomethane

Table of Amino Acid abbreviations

1 letter	3 letter	Full name	1 letter	3 letter	Full name
A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic acid	P	Pro	Proline
E	Glu	Glutamic acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	leucine	Y	Tyr	Tyrosine

Chapter 1: General introduction

1.1 Class A GPCRs

1.1.1 G-protein coupled receptors

How organisms respond to changes in the environment is important for the survival of individuals and, ultimately, the prosperity of the species. Organisms have responded to changes in the outside world by receiving stimuli, transmitting information within or between cells, and changing their morphology and behavior based on the transmitted information. There are three main types of proteins that accept extracellular stimuli and transmit them into the cell. First, there are iontotropic receptors such as the $GABA_A$ receptor, which accepts GABA and passes Ca^{2+}, and second, there are receptors for growth factors that activate Ras and induce Mitogen-activated protein kinase (MAPK) cascade, and last, G-protein coupled receptors (GPCRs) are the largest superfamily of these receptors. GPCRs are not only the largest superfamily, but also the largest among membrane proteins.

The ligands that GPCRs can accept are diverse, including Ca^{2+} ions, amines, hormones, peptides, and even proteins. G-proteins have three subunits, α, β, and γ, and when they bind to GPCRs, the α subunit is phosphorylated and the βγ subunit is released. The phosphorylated α subunit transmits signals downstream (Figure 1-1).

There are mainly three types of Gα subunit; G_s, G_i, G_q, and $G_{12/13}$[1]. Different GPCRs bind to different α-subunits, and induces different downstream signaling depending on the binding α-subunit. The G_s or G_i family activates or inhibits adenylyl cyclase activity, respectively, and increases intracellular cAMP levels. family activates phospholipase C signaling[2]. $G_{12/13}$ family activates Rho family small GTP binding protein[3].

It has also been reported that not only the G_α subunit but also the $G_{\beta\gamma}$ subunit plays an important signaling role in vivo, such as the activation of GIRK channels. It has also been shown that certain ligands induce the binding of β-arrestin to desensitize the receptor and induce β-arrestin biased agonism, as shown in the 2011 study on β2AR[4].

Thus, GPCRs are involved in diverse physiological phenomena such as embryonic development, cell proliferation, immune response, apoptosis, and neuronal response because they regulate diverse signaling. For this reason, GPCRs are membrane proteins that account for one third of drug targets, and their detailed structural information is important for biological and medical.

1.1.2 Class A GPCRs

GPCRs are broadly classified into six subfamilies from A to F. Especially in vertebrates, they are divided into class A, B, C, and F[5]. Class A GPCRs, also called the rhodopsin family, are a subfamily that accounts for 80% of all GPCRs and is so diverse that it can be further divided into 19 subgroups. Class A GPCRs have a variety of ligands, including amines, peptides, lipids, and nucleic acids. Class B GPCRs are a subfamily that includes The Secretin receptor family or the adhesion family, and mainly accept peptide hormones. Class C GPCRs are a subfamily that includes metabotropic glutamate receptors and $GABA_B$ receptors. Class F GPCRs are a subfamily that includes Smoothened and Frizzled GPCRs, and they accept proteins such as Wnt. In addition, although the three-dimensional structure has not yet been reported, but taste receptors that accept sweet, bitter, and umami tastes are also considered to be GPCRs and are classified as class T.

The vertebrate GPCRs, class A, B, C, F, and T GPCRs, are often classified by GRAFS classification. In this case, class B GPCRs are divided into two, and class F GPCRs and class T GPCRs are lumped together and called the Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2, and Secretin subfamilies. class A GPCRs correspond to "R".

The structural features of the GPCR subfamilies whose structures have been reported in GPCRs are shown in the Figure 1-2. On the other hand, class A GPCRs are characterized by the presence of a ligand pocket in the TM. Class A GPCRs are the most well-studied GPCRs because many of them are potential drug targets, and the structural information of class A GPCRs is particularly important because they are key to the search for receptor agonists and antagonists.

1.1.3 Structural studies of class A GPCRs

The first GPCRs discovered were β-adrenergic receptor and rhodopsin, and analysis using the Hopp and Woods method and the Kyte and Doolittle method suggested that they have a common transmembrane domain region based on amino acid sequence alignments[6].

In 1997, the structure of rhodopsin at 7 Å resolution was reported by two-dimensional crystal structure analysis using electron diffraction, suggesting that rhodopsin is a membrane protein with seven transmembrane helices (TM)[7]. And in 2000, the three-dimensional crystal structure of rhodopsin was reported, which revealed that the

GPCR has seven TMs, extracellular and intracellular loops, and one amphiphilic helix (H8) at the C-terminus[8] (Figure 1-3). In 2011, the complex structure of β2AR, which is a complex structure of ligand-receptor and G_s protein, was reported for the first time in class A GPCRs (Figure 1-4). These papers and many other structural studies of class A GPCRs have revealed that the inward shift of TM6 and TM7 is important for the stable binding of G proteins and the activation of G proteins in general class A GPCRs. In addition, in order to discuss the G-protein activation mechanism in class A GPCRs, a notation called Ballesteros-Weinstein numbering[9] was used. This notation is based on the most highly conserved residue in each TM. Hereafter, the Ballesteros-Weinstein numbering is superscripted to the right of a receptor residue if necessary.

The motifs involved to induce activation are highly conserved among class A GPCRs. The major key motifs are CWxP, PIF, NPxxY, and D/ERY[10]. "x" can apply to any amino acid. More than 70 unique X-ray crystallographic and cryo-EM single-particle structures of class A GPCR show that these key motifs play a central role in the structural change from the inactive to the active form at the TM level.

Especially, $W^{6.48}$, which is a part of the CWxP motif, functions as a Na^+ pocket that coordinates Na^+ ions together with $N^{7.49}$, $D^{2.50}$, and $S^{3.39}$ in the inactive state. when the receptor accepts the ligand, the Na^+ pocket collapses and $W^{6.48}$ rotates to the active form[11]. However, the process from the acceptance of the ligand by the receptor to the collapse of the Na^+ pocket differs among receptors, and not all class A GPCRs have the identical motifs and the same activation mechanism. For example, in Endothelin receptor type B, it is known that the H_2O-cluster contributes to stabilization of the inactive state instead of the Na^+ pocket[12].

1.2 Endothelin receptor type B

1.2.1 Endothelin receptors

Endothelin receptor (ETR) is a G-protein coupled receptor, which is activated by peptide agonist, endothelin (ET). In 1988, endothelin was first isolated as a vasoconstrictor derived from vascular endothelium[13]. Endothelin is the first mammalian peptide with a cyclic structure containing two disulfide bonds (Figure 1-5). It was assumed that the peptides acted on the ion channels in the cell membrane, similarly to neurotoxins such as bee venom and scorpion venom, which have similar structures. This endothelin was identified as ET-1, and three endothelin analogs, ET-1, ET-2, and ET-3,

were isolated. In 1990, the cloning of two endothelin receptors, ETAR and ETBR, was reported[1415]. These endothelin receptors are widely expressed in the vascular endothelium, brain, and other circulatory organs. ET_B is expressed in the vascular endothelium and its stimulation induces nitric oxide-mediated vasorelaxation, while ET_A and ET_B are both expressed in the vascular smooth muscle and their stimulation leads to potent and long-lasting vasoconstriction (Figure 1-6). Therefore, systemic administration of ET-1 causes transient vasodilation (initial endothelial ET_B activation) and hypotension, followed by prolonged vasoconstriction and hypertension (smooth muscle ET_A and ET_B activation).

ET-1 is one of the most potent vasoconstrictors and its toxic is high (LD_{50} for mice is about 0.015 mg/kg body weight). The up-regulation of ET-1-ETAR pathway is significantly related to circulatory-system diseases[4], including pulmonary arterial hypertension (PAH)[16]. Thus, antagonists of endothelin receptor have been developed for the treatment of circulatory system diseases. Recently, the crystal structures of the human ET_B receptor have been determined in complex with agonist ET-1 and an antagonist bosentan, and inverse agonist K-8794.

1.2.2 ETBR ligands

Since the activation of the ET_B receptor has a vasorelaxation effect, unlike the ET_A receptor, ET_B-selective agonists have been studied as vasodilator drugs for the improvement of tumour drug delivery, as well as for the treatment of hypertension.

Only ET-3 is an endothlin type B-selective agonist *in vivo*. Moreover, the ET-3/ET_B signaling pathway has distinct physiological roles, as compared to the ET-1/ET_B pathway. In the brain, ET-3 is responsible for salt homeostasis, by enhancing the sensitivity of the brain sodium-level sensor Na_x channel[17]. Moreover, the ET-3/ET_B signaling pathway is related to the development of neural crest cells, and has an essential role in the formation of the enteric nervous system. Thus, mutations of the ET-3 or ET_B genes cause Hirschsprung's disease, a birth defect in which nerves are missing from parts of the intestine. Generally, the endothelin system participates in a wide range of physiological functions in the human body.

IRL1620 (N-Suc-[E9, A11, 15] ET-1_{8-21}[18], a N-terminal truncated peptide analogue of ET-1, is the smallest agonist that can selectively stimulate the ET_B receptor, and currently no non-peptidic ET_B-selective agonists have been found. The affinity of IRL1620 to the ET_B receptor is comparable to that of ET-1, while it essentially does not

activate the ET_A receptor, and thus it shows high ET_B selectivity of over 100,000-fold. Due to its large molecular weight, IRL1620 is not able to use as orally active drug and thus requires intravenous delivery. Despite its pharmacokinetic disadvantages, IRL1620 is an attractive candidate for the treatments of diseases related to the ET_B receptor. Since the ET_B-selective signal induces vasorelaxation and improves blood flow, IRL1620 could be utilized for the improved efficacy of anti-cancer drugs at the same dose by increasing the efficiency of drug delivery, as shown in rat models of prostate and breast cancer[18,19,20]. In addition, this strategy can also be applied to radiotherapy in the treatment of solid tumours, as the radiation-induced reduction in the tumour volume was enhanced by IRL1620[20]. On the top of that, IRL1620 has vasodilation and neuroprotection effects in the brain. IRL1620 reduced neurological damage following permanent middle cerebral artery occlusion of focal ischaemic stroke in a rat model[21]. Moreover, the stimulation of the ET_B receptor by IRL1620 reduces the cognitive impairment induced by β-amyloid (1-40), a pathological hallmark of Alzheimer's disease, in rat experiments[22,23]. These data suggest that ET_B selective agonists might offer new therapeutic method for neuroprotection and Alzheimer's disease. The safety and maximal dose of IRL1620 were investigated in a phase I study. Though a recent phase 2 study of IRL1620 in combination with docetaxel as the second-line drug reported no significant improvement in the treatment of advanced biliary tract cancer (ABTC)[24], further trials for selected patients based on tumour types with various choices for the second-line drugs are still expected. Simultaneously, the pharmacological properties of IRL1620 could also be improved for better clinical applications. However, little is known about the receptor selectivity and activation mechanism of this artificially designed agonist peptide, although the ET_B structures in complex with ET-1 and antagonists have been determined[25,26].

Sarafotoxins (SRTXs) are a group of toxins purified from a venom of *Atractaspididae* family of snakes[27,28,29], and share high structural and functional homology with the endothelins. SRTXs include the four major isopeptides S6a-d, all composed of 21 amino acids. SRTXs stimulate the endothelin receptors and induce vasoconstriction, which is followed by left ventricular dysfunction, bronchoconstriction, and increase of the airway resistance. S6b binds to both ET_A and ET_B receptors with similar affinities to ET-1, and thus are highly lethal by causing cardiac arrest and death in mice within minutes of intravenous administration (LD50 for mice is about 0.015 mg/kg body weight) [29]. S6b is the most potent in sarafotoxins. By contrast, S6c shows 100- to 10,000-fold reduced

affinity for the ET_A receptor and thus works as an ET_B-selective agonist. As a result, S6c shows the modest toxicity as compared with S6b (LD50 for mice is about 0.30 mg/kg body weight). In addition, S6b possesses modest matrix metalloproteinase inhibitory activity, because S6b shares a common fold with the core region of the tissue inhibitors of metalloproteinases (TIMPs)[30]. Several ET_B structures have been determined, however, the binding modes and selectivities of the sarafotoxins remained to be elucidated.

Amino acid sequences and the B-type selection system of ETBR ligands are shown in Table 1-1. There is a high degree of sequence conservation in the C-terminal region, in contrast to the diversity of the N-terminal region.

1.2.3 Structural studies of ETBR

Figure 1-7 lists the ETBR structures that have already been reported. The ligand unbound state and the ET-1 bound structure, a potent agonist, were reported in 2016. These two structures revealed the three-dimensional structure of ETBR and the inward movement of extracellular TMs 6 and 7 is important for receptor activation. Then, the antagonist Bosentan, K8794-bound structure was reported in 2017. This paper revealed that a network of water molecules in the core of the receptor is responsible for stabilizing the inactive state of ETBR. However, there are still some unanswered questions about the structural information of ETBR. First, ET-1, which was reported as an agonist-bound state, is a “nonselective” agonist, that means ET-1 binds to ETAR and ETBR to the same level, and the ETBR-binding mode of B-type selective agonists is unknown. Seconds, although the water molecule network in the receptor core contributes to the stabilization of the inactive state of the receptor, no water molecules were observed in the medium resolution structure of the ET-1 binding form, and the activation mechanism of ETBR focusing on the water molecule network was unknown. Third, IRL1620 had failed in clinical trials despite being similar to endothelin in binding capacity, and the cause was unknown. In addition, sarafotoxin, a snake venom of endothelin-like peptide, is known for its strong vasoconstrictor effect, but its binding mechanism to ETRs was unknown.

1.3 LysoPS receptor 1 (GPR34)

1.3.1 Metabolic pathways

Lipid is found in various locations in vivo, including cell membranes and fat droplets. Especially, lipids that are involved in cellular responses, such as being accepted as ligands

for GPCRs, are called lipid mediators. Lipid-receiving class A GPCRs have been well studied and more than 10 structures of lipid-receiving class A GPCRs have been reported.

Also, it is known that lipid mediators are likely to be involved in immune response and inflammation. The most representative, oldest- and best-known lipid mediators are prostaglandins, which are pro-inflammatory lipid mediators. When membrane phospholipids are cleaved at the *sn-2* position by phospholipases, arachidonic acid or EPA and DHA are released. The arachidonic acid is metabolized by the COX pathway into inflammatory mediators such as prostaglandins and leukotrienes. Loxoprofen, which is often used as a headache medicine, is a well-known COX pathway inhibitor and suppresses the production of prostaglandins.

EPA and DHA are other unsaturated fatty acids that are coordinated to the *sn-2* position. It is often said that eating fish is "healthier" than eating meat, because EPA and DHA are more abundant in fish than in meat. Metabolites of EPA and DHA, such as RvD1 and RvE1, have been shown to act as pro-resolving mediators.[31]

Thus, the metabolic pathways and functions of fatty acids released from the *sn-2* position of phospholipids have been well studied, but research on lysophospholipids, "the rest" of phospholipids, has been lagging behind due to the lack of progress in the identification of lysophospholipid receptors[32]. In 1996, the successful cloning of LPA1 led to the classification of multiple orphan (but that is now known as a lipid-sensing) GPCRs. At present, more than 10 class A GPCRs are known to receive lysophospholipids[5] .

1.3.2 Structural studies of lipid-sensing GPCRs

Typical lipid receptors whose structures have been solved are shown in the Figure 1-8. The lipid receptors, except for FFAR1, LPA6, and BLT1, exhibit a tightly folded extracellular domain called the lid structure that restricts ligand access to the receptor core from the outside of the cell. This lid structure is formed by the N-terminal region of the receptor or the ECL2 region, or both. This lid structure is widely conserved even among receptors with low sequence conservation. The lid structure and the amphiphilicity of lipid mediators suggest that lipid mediators may enter lipid receptors by lateral access from inside the plasma membrane. The gaps were located on the extracellular side of TM1-7 or on the TM4-5 of the receptor. EP3 and BLT1 have no gap in TMs. As for the gap at TM4-5, the crystal structure of LPA6 suggests that it may be a lateral access region

or, moreover, the binding pocket. However, no agonist density has been found in structure of lipid-receiving GPCRs, even in the presence of agonist and its active state. Cryo-EM structure of the PGE_2-EP4-G_s complex was reported in December 2020[33], but the ligand binding site has not been definitive. Therefore, ligand recognition of lipid receptors is still controversial.

1.3.3 LysoPS receptor LPS1 (GPR34)

LysoPS, like LPA and S1P, is a type of lysophospholipid that consists of LPA with serine attached to it (Figure 1-9). LysoPS, like LPG, LPE, and LPI, exists in low concentrations in vivo, and its role has long been unknown[33]. Orphan receptor GPR34 was found to receive LysoPS in 2006[33]. GPR34 did not respond to other lysophospholipids, including LPA, S1P, LPC, and LPE; only D-LysoPS was found to be accepted by GPR34 as a weak agonist[33]. This suggests that LPS1 strictly recognizes the three-dimensional structure of the phosphate group of LysoPS. After that, LPS2 (P2Y10) and LPS3 (GPR174) were discovered as LysoPS receptor.

GPR34 was reported that it was involved in neuropathic pain in vivo in 2019. Neuropathic pain can be caused by external injuries such as nerve damage or compression, such as herpes zoster, or by abnormalities in the brain or spinal cord, such as phantom limb pain, etc. LPS1 was noticed to be upregulated in spinal cord dorsal horn microglia within a week after nerve injury. Mutant analysis in wild-type mice and LPS1KO mice showed that LPS1 induces inflammation in the dorsal horn of the spinal cord and that knockout reduces inflammatory markers. Thus, antagonists of GPR34 are expected to have pharmaceutical applications for neuropathic pain.

LPS1 (GPR34) is a member of the P2Y family that accepts purine nucleotides. Recently, the structures of P2Y1 and P2Y12 have been reported. 2.6 Å crystal structure of P2Y12 in complex with antagonist revealed that P2Y12 is a GPCR with an α-helix at ECL2. The 2.2 Å and 2.7 Å crystal structures of the complex of P2Y1 and two antagonists revealed that P2Y1 is a GPCR with a β-sheet structure in ECL2. In addition, the antagonist BPTU-bound state was unique in that BPTU was bound outside of TM2 and TM3 of P2Y1. However, since P2Y1 and P2Y12, which have high sequence homology with GPR34, both accept purine nucleotides, it was unclear from the crystal structures of the two receptors how the LysoPS receptor accepts lipids.

1.4 Overview of this thesis

1.4.1 Ligand selectivity of ETRs (in Chapter 2)

Contrary to the diversity of amino acid sequences, ET-1, ET-3, IRL1620, and S6b all had the peptide C-terminus deeply penetrating the binding pocket in the center of the transmembrane region, and ECL2 formed the lid. The residues in the binding pocket were highly conserved between ETAR and ETBR, while those in ECL2 were less conserved, suggesting that structural differences in ECL2 contribute to the type B selectivity. Comparison of the binding modes of various agonists to ETBR, analysis of mutants of ECL2 residues, and construction of a model based on the structure of S6b-bound ETBR suggest that the binding of ligand to ECL2 of ETAR affects the type B selectivity.

1.4.2 Activation mechanism of ETBR (in Chapter 2)

In Class A GPCRs, Na^+ coordinates near the receptor core D147 and forms Na^+-H_2O clusters, which contribute to the stability of the inactive state. 2.2 Å resolution structure of the antagonist K8794-bound ETBR shows that a water molecule-mediated network instead of Na+ stabilizes the inactive state in ETBR. In this study, I report the structure of ET-3 bound to the antagonist K8794. In this study, I determined the conformation of the ET-3 binding type at 2.0 Å. The water-mediated network in the vicinity of D147 is disrupted and reorganized upon agonist binding, leading to the migration of W336 and N378, which are highly conserved among GPCRs, and caused outward shift of the intracellular TM6.

1.4.3 Partial activation (in Chapter 2)

Because IRL1620 lacks the N-terminal seven residues, there is no binding to K346 of TM6, and the IRL1620-bound ETBR has a smaller inward shift of extracellular TM6 than the ET-3-bound structure. To investigate the effect of this, I compared the vicinity of D147 and found that the receptor core residue of IRL1620-bound ETBR was similar to that of the antagonist K8794-bound ETBR. The result suggested IRL1620 may have a low receptor activation ability, so I next evaluated the receptor activation ability of ET-3, thermostable mutations of IRL1620, and wild-type ETBR by TGFα shedding assay (in collaboration with Aoki Laboratory, Graduate School of Pharmaceutical Sciences, Tohoku University). The E_{max} of IRL1620 was 88% of that of ET-1, even though IRL1620 has the same binding capacity as endothelin. This study suggests that IRL1620 is a partial

agonist that cannot fully activate the receptor.

1.4.4 Lipid recognition of GPR34 (in Chapter 3)

The complex of GPR34 was successfully expressed, purified and obtained Cryo-EM density map, and its 3.6 Å structure is determined. The density of the putative ligand, unfortunately, could not be determined due to the middle resolution. However, there were holes from the positively charged moiety covered by ECL2 to the groove structure of TM4-5, suggesting that these pockets and grooves are involved in the recognition of polar and acyl groups of LysoPS, or in the pathway of the ligands into the receptor.

1.4.5 Thesis composition

In this study, I focused on the differences in the binding mode and activation mechanism of peptide hormone receptor type B and its various agonists, and tried to explain the mechanism by X-ray crystallography. The results showed that the structures of the ECL2 and ECL2TM5 linkers may differ among ETRs, suggesting that these structural differences determine the type B selectivity. The ET-3-bound structure also revealed that changes in the water molecule network near D147 are required for activation. The interaction near D147 suggested that IRL1620 is a partial agonist, which was demonstrated by functional analysis. This study is the first example of a partial activation mechanism in terms of structure in Class A GPCR. In the next step, I will perform single-particle structural analysis of the complex of ETAR, agonist, and G protein by cryo-electron microscopy, and discuss the B-type selectivity by comparing the structures of ETAR and ETBR. The structure of the LysoPS receptors has not been previously reported. The structural information on GPR34 revealed in this paper will greatly advance our understanding of the orthosteric site and the access pathway to the orthosteric site of the ligand, which has not been clarified for lipid-sensing class A GPCRs.

Figures and Tables of Chapter 1

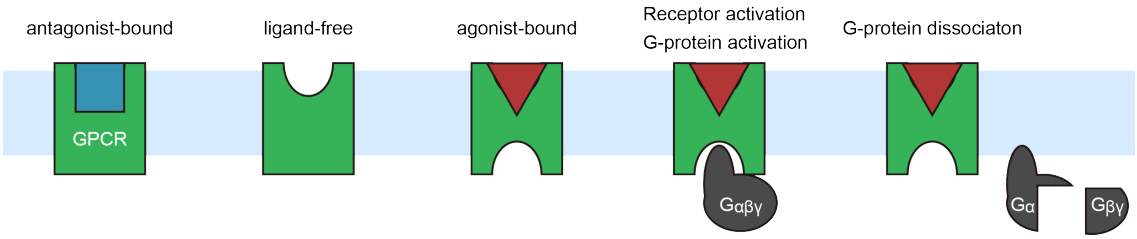


Figure 1-1 Common GPCR signaling

When GPCRs receive agonists, the flexibility of intracellular TM6 and TM7 increases, facilitating the binding of G-protein; G-protein-bound GPCRs have reduced flexibility of intracellular TM6 and TM7 and are change to the active state. The G-protein bound to the GPCR is able to bind GTP, and the Gα subunit is phosphorylated and separate into Gα and Gβγ subunits. The GPCR-bound G protein is able to bind GTP and diverge into Gα and Gβγ subunits by phosphorylation of the Gα subunit, and each diverged subunit activates a different signaling. Antagonist prevents the agonist from binding to the GPCR or inhibits the change of the GPCR to an activated state.

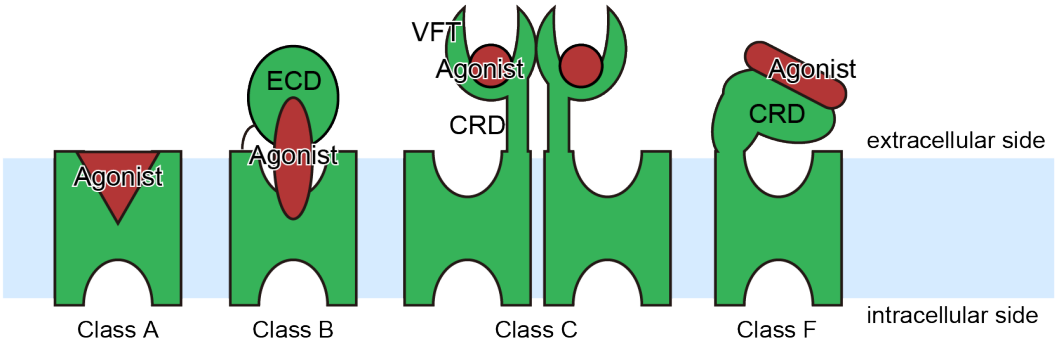


Figure 1-2 Cartoon models of vertebrate GPCR structures in different classes

The green objects show the models of GPCRs for each class, and the red region shows the agonist binding site. Common class A GPCRs have the agonist binding site in the TM, while classes B, C, and F have a large extracellular domain and accept the agonist by the extracellular domain or conjugation with TMs domain.

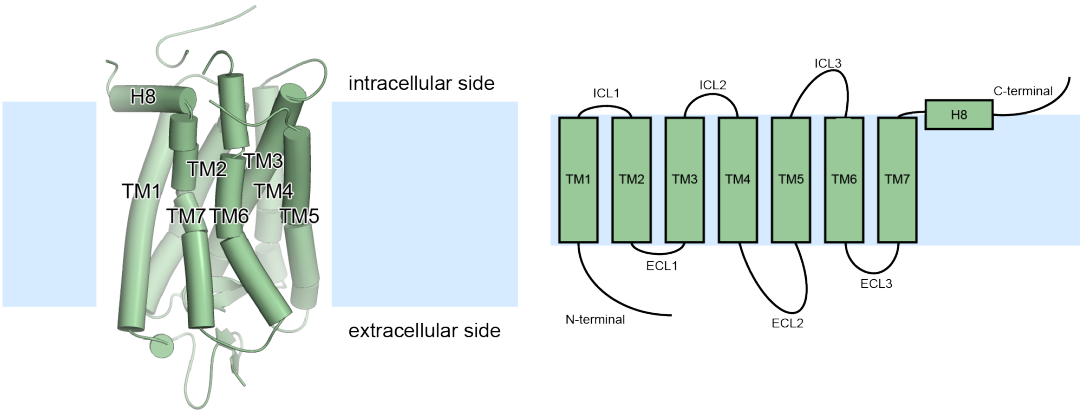


Figure 1-3 Crystal structure of rhodopsin

The crystal structure of bovine rhodopsin (PDBID: 1F88). α-helices and β-sheet are shown in cartoon and sheet, respectively. As is tradition, the structure of Rhodopsin is usually written with the inside of the cell facing up.

Figure 1-4 Crystal structure of β2AR-Gs complex

The crystal structure of β2AR-Gs complex (PDBID: 3SN6), viewed parallel to the membrane plane. The β2AR (green) is coupled with $G_{\alpha s}$ (yellow), G_{β} (blue), and G_{γ} (purple). The nanobody (Nb35, Dark gray) and T4 lysozyme (Gray) facilitates crystallization.

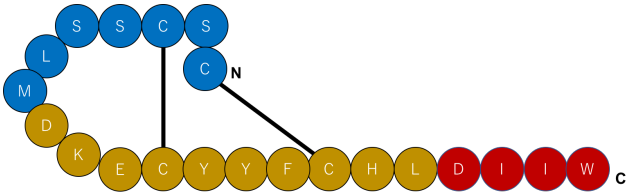


Figure 1-5 Amino acid sequence and structure of Endothelin-1

ET-1 is a peptide hormone consisting of 21 amino acid residues. Endothelin has a cyclic structure with two disulfide bonds in the sequence. Residues 1-7 are known as the N-terminal region (blue), 8-17 as the α-helical region (orange), and 18-21 as the C-terminal region (red).

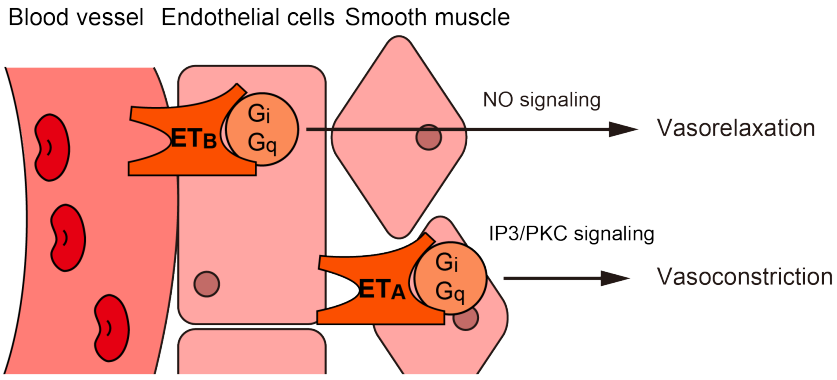


Figure 1-6 ETRs signaling

Table 1-1 Endothelin receptor agonists and its selectivity

	N-terminal		α-helical		C-terminal		
Agonists	1	7	8	17	18	21	ETR selectivity
ET-1	CSCSSLM		DKECVYFCHL		DIIW		Non-selective
ET-3	CTCFTYK		DKECVYYCHL		DIIW		ETB ~ 100 fold
IRL-1620	Suc		DEEAVY		DIIW		ETB ~100,000 fold
Sarafotoxin S6b	CSCKDMT		DKECLYFCHQ		DVIW		Non-selective
Sarafotoxin S6c	CTCNDMT		DEECLNFCHQ		DVIW		ETB ~ 100-10,000 fold

There are three types of endothelin, ET-1, ET-2 and ET-3. Only ET-3 has 100-fold lower affinity for ETAR. IRL1620 is the smallest artificial peptide agonist for ETBR. IRL1620 is 100,000-fold more selective for type B and binds little to ETAR. There are four types of sarafotoxin, a-d, with S6b being the most potent. Sarafotoxin s6c is often used in functional analysis as a B-type selective agonist along with IRL1620.

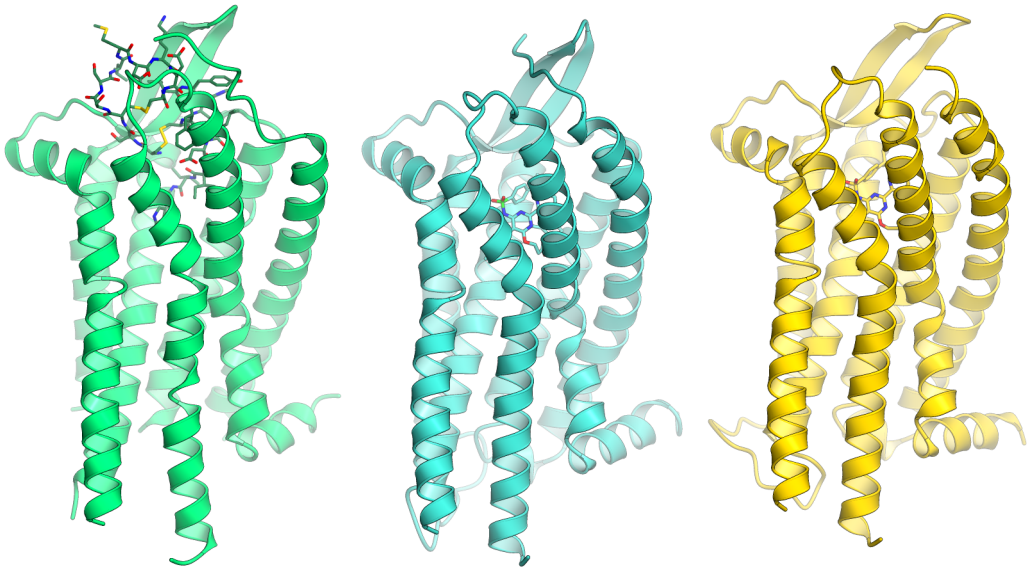


Figure 1-7 Overall structures and Ligand binding of ETBRs

Crystal structures of peptide ligand ET-1-bound ET_B (a), small molecule antagonist K8794-bound ET_B (b), and bosentan-bound ET_B (c). The receptors are indicated by ribbons, and the endothelin and the antagonists are indicated by sticks.

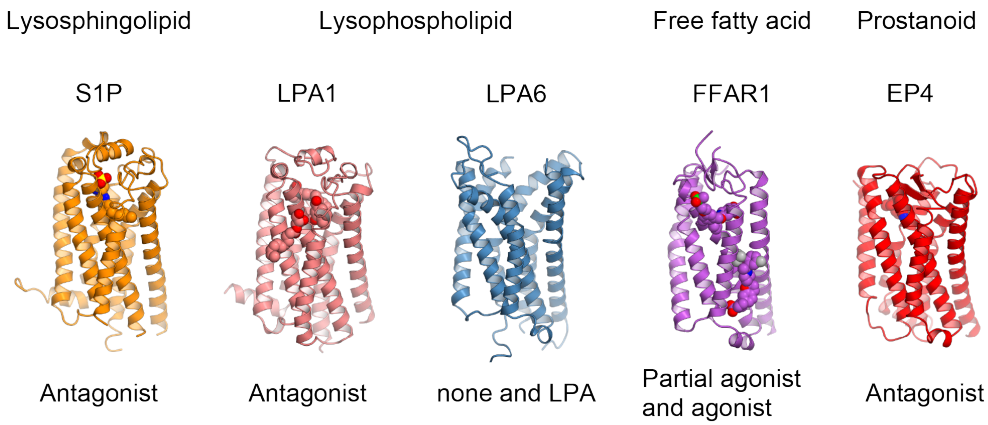


Figure 1-8 Overall structures of lipid-sensing GPCRs

The crystal structures of S1P (PDBID: 3V2Y), LPA1 (PDBID: 4Z36), LPA6 (5XSZ), FFAR (5TZY), EP4 (5YWY). The receptor and ligands are shown in ribbon and cpk, respectively.

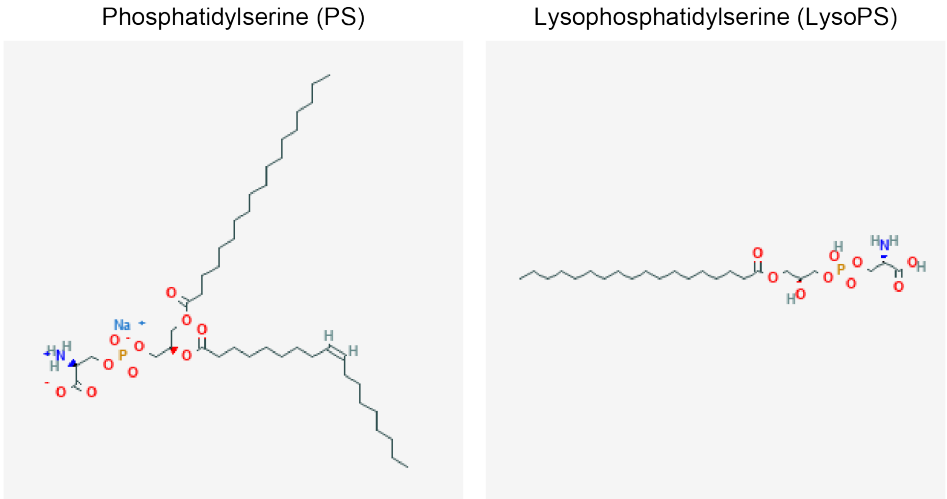


Figure 1-9 Chemical structural formula of PS and LysoPS.

2D structures of PS (18:0 - 18:1) (PubChem ID: 46891787) and LysoPS (18:0) (PubChem ID: 42607474) from PubChem.

Chapter 2: Structural analysis of human Endothelin receptor type B

2.1 Introduction (study purpose)

As described above, Endothelin receptor is Class A GPCR, which is activated by peptide agonist, endothelin (ET). Especially, Endothelin receptors type B has vasorelaxation effects and may improve hypertension, Alzheimer's disease and drug delivery system. The agonist that can also bind to endothelin receptor type A induces vasoconstriction of type A. This led to the development of ET_B-selective agonists. Inspite of the importance of binding mechanism of B type-selective agonist, little is known about the selectivity and activation mechanism of this artificially designed agonist peptide, although the ET_B structures in complex with ET-1 and antagonists have been determined. Also, binding modes and selectivities of endothelin-like peptides, sarafotoxins are remained to be elucidated, despite it is often used in functional analysis.

In this study, I report the crystal structures of the ET_B receptor in complex with two ET_B-selective ET variant agonists, ET-3 and IRL1620, and endothelin-like peptides sarafotoxin S6b. Together with their detailed pharmacological characterization, the structures reveal the different activation mechanisms of these agonists.

2.2 Material and methods

2.2.1 Expression and purification

I used the thermostabilized ET_B receptor ET_B-Y5-T4L. as previously established. In brief, the haemagglutinin signal peptide followed by the Flag epitope tag was added to the N-terminus, and a TEV protease recognition sequence was introduced between G57 and L66. The C-terminus was truncated, and three cysteine residues were mutated to alanine (C396A, C400A, and C405A). To facilitate crystallogenesis, T4 lysozyme (C54T and C97A) was introduced into intracellular loop 3 (Figure 2-1).

The ET_B-Y5-T4L was subcloned into a modified pFastBac vector, with the resulting construct encoding a TEV cleavage site followed by a GFP-His_{10} tag at the C-terminus. The recombinant baculovirus was prepared using the Bac-to-Bac baculovirus expression system (Invitrogen). Sf9 insect cells (Invitrogen) were infected with the virus at a cell density of 3.5-4.0 $\times 10^6$ cells per millilitre in Sf900 II medium, and grown for 48 h at 27 °C. The cells were disrupted by sonication, in buffer containing 20 mM Tris–HCl, pH 7.5, 200 mM NaCl, and 20% glycerol. The membrane fraction was collected by

ultracentrifugation, and solubilized in buffer, containing 20 mM Tris–HCl, pH 7.5, 200 mM NaCl, 1% DDM, 0.2% CHS, and 2 mg/ml iodoacetamide, for 1 h at 4 °C. The insoluble material was removed by ultracentrifugation at 180,000 × g for 20 min, and incubated with TALON resin (Clontech) for 30 min. The resin was washed with 10 column volumes of buffer, containing 20 mM Tris–HCl, pH 7.5, 500 mM NaCl, 0.1% LMNG, 0.01% CHS, and 15 mM imidazole. The receptor was eluted in buffer, containing 20 mM Tris–HCl, pH 7.5, 500 mM NaCl, 0.01% LMNG, 0.001% CHS, and 200 mM imidazole. TEV protease was added to the eluate, and the mixture was dialyzed against buffer, containing 20 mM Tris–HCl, pH 7.5, and 500 mM NaCl. The cleaved GFP and the protease were removed with Co^{2+}-NTA resin. The flowthrough was concentrated and loaded onto a Superdex200 10/300 Increase size-exclusion column, equilibrated in buffer containing 20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.01% LMNG, and 0.001% CHS (Figure 2-2). Peak fractions were pooled and concentrated to 40 mg ml^{-1} using a centrifugal filter device (Millipore 50 kDa MW cutoff), and frozen until crystallization. ET-3, IRL1620, Sarafotoxin S6b, and Sarafotoxin S6c were added to a final concentration of 100 μM, during the concentration.

2.2.2 Crystallization

The purified receptors were reconstituted into molten lipid (monoolein and cholesterol 10:1 by mass) at a weight ratio of 1:1.5 (protein:lipid). The protein-laden mesophase was dispensed into 96-well glass plates in 30 nl drops and overlaid with 800 nl precipitant solution, using an LCP dispensing robot (Gryphon, Art Robbins Instruments)[3435]. Crystals of ET_B-Y5-T4L bound to ET-3 were incubated at 20 °C in the precipitant conditions containing 25% PEG500DME, 100 mM MES-NaOH, pH 6.0, and 100 mM ammonium citrate tribasic. The crystals of ET_B-Y5-T4L bound to IRL1620 were incubated in the precipitant conditions containing 20–25% PEG500DME, 100 mM sodium citrate, pH 5.0, and 100 mM (NH4)2SO4 or NH4Cl. Crystals of ET_B-Y5-T4L bound to S6b were incubated at 20 °C in precipitant conditions containing 30% PEG600, 100 mM Citrate, pH 5.0, 100 mM ammonium sulfate. Crystals of ETB-Y5-T4L bound to S6c were incubated at 20 °C in precipitant conditions containing 30% PEG500dme, 100 mM Tris, pH 7.8, 100 mM potassium thiocyanate. All crystals were harvested directly from the LCP using micromounts (MiTeGen) or LithoLoops (Protein Wave) and frozen in liquid nitrogen, without adding any extra cryoprotectant.

2.2.3 Data collection and structure determination

X-ray diffraction data were collected at the SPring-8 beamline BL32XU with 1 × 10 to 8 × 25 μm^2 (width × height) micro-focused beams and an EIGER X 9M detector (Dectris). For the IRL1620 data, I manually collected 68 data sets (10°–180° per crystal), and the collected images were automatically processed using KAMO[36] (https://github.com/keitaroyam/yamtbx). Each data set was indexed and integrated with XDS[37] and subjected to a hierarchical clustering analysis based on the unit cell parameters using BLEND[38]. After the rejection of outliers, 46 data sets were finally merged with XSCALE[37]. From the ET-3-bound crystals, various wedge datasets (3–180°) per crystal were mainly collected with the ZOO system, an automatic data-collection system developed at SPring-8. The loop-harvested microcrystals were identified by raster scanning and subsequently analyzed by SHIKA[39]. The collected images were processed in the same manner, except that correlation coefficient-based clustering was used instead of BLEND, and finally 483 datasets were merged. The ET-3-bound structure was determined by molecular replacement with PHASER[40], using the ET-1-bound ET_B structure (PDB 5GLH). Subsequently, the model was rebuilt and refined using COOT[41] and PHENIX[42], respectively. The IRL1620-bound structure was determined by molecular replacement, using the ET-1-bound ETB structure, and subsequently rebuilt and refined as described above. The final model of ET-3-bound ETB-Y5-T4L contained residues 86–303 and 311–403 of ETB, all residues of T4L, ET-3, 12 monoolein molecules, a citric acid, and 175 water molecules. The final model of IRL1620-bound ETB-Y5-T4L contained residues 87–303 and 311–402 of ETB, all residues of T4L, IRL1620, 3 monoolein molecules, 4 sulfate ions, a citric acid, and 59 water molecules. The model quality was assessed by MolProbity[43]. Figures were prepared using CueMol (http://www.cuemol.org/en/)

Sarafotoxins-bound ET_B crystals were collectied and processed as ET-3- and IRL1620- bound ET_B. In brief, X-ray diffraction data of Sarafotoxins were collected at the SPring-8 beamline BL32XU, with 10 × 15 μm2 (width × height) micro-focused beams and an EIGER X 9 M detector (Dectris). I manually collected 32 data sets (10° per crystal), and the collected images were automatically processed with KAMO. Each data set was indexed and integrated with XDS and then subjected to a hierarchical clustering analysis based on the unit cell parameters using BLEND[38]. After the rejection of outliers, 16 data sets were finally merged with XSCALE. The S6b-bound structure was determined by

molecular replacement with PHASER, using the ET-3-bound ET_B structure (PDB code: 6IGK). Subsequently, the model was rebuilt and refined using COOT and PHENIX, respectively. The final model of the S6b-bound ETB-Y5-T4L contained residues 86–210, 214–303, and 311–403 of ET_B, 1–160 of T4L, and s6b. The model quality was assessed by MolProbity. Figures were prepared using CueMol (http://www.cuemol.org/ja/).

2.2.4 TGF α shedding assay

All TGF α shedding assay were performed by Francois Marie Ngako Kadji and Junken Aoki. The TGFα shedding assay[44], which measures the activation of G_q and G_{12} signalling, was performed as described previously[26]. Briefly, a pCAG plasmid encoding a codon-optimized ETB construct (Supplementary Table 2) with an internal FLAG epitope tag or a full-length human ETA construct with an internal FLAG epitope tag cloned from cDNA was prepared. Double mutant L252H I254T was generated by Quickchange PCR (Supplementary Table 3). These plasmids were transfected, together with a plasmid encoding alkaline phosphatase (AP)-tagged TGFα (AP-TGFα), into HEK293A cells by using a polyethylenimine (PEI) transfection reagent (1 μg ETR plasmid, 2.5 μg AP-TGFα plasmid and 25 μl of 1 mg/ml PEI solution per 10-cm culture dish). After a one day culture, the transfected cells were harvested by trypsinization, washed, and resuspended in 30 ml of Hank's Balanced Salt Solution (HBSS) containing 5 mM HEPES (pH 7.4). The cell suspension was seeded in a 96 well plate (cell plate) at a volume of 90 μl per well and incubated for 30 min in a CO2 incubator. Test compounds, diluted in 0.01% bovine serum albumin (BSA) and HEPES-containing HBSS, were added to the cell plate at a volume of 10 μl per well. After a 1 h incubation in the CO_2 incubator, the conditioned media (80 μl) was transferred to an empty 96-well plate (conditioned media (CM) plate). The AP reaction solution (10 mM p-nitrophenylphosphate (p-NPP), 120 mM Tris–HCl (pH 9.5), 40 mM NaCl, and 10 mM $MgCl_2$) was dispensed into the cell plates and the CM plates (80 μl per well). The absorbance at 405 nm (Abs405) of the plates was measured, using a microplate reader (SpectraMax 340 PC384, Molecular Devices), before and after a 1-h incubation at room temperature. AP-TGFα release was calculated as described previously[26]. The AP-TGFα release signals were fitted to a four-parameter sigmoidal concentration-response curve, using the Prism 7 software (GraphPad Prism), and the pEC_{50} (equal to $-\mathrm{Log}_{10}\,EC_{50}$) and E_{max} values were obtained.

2.2.5 β-arrestin recruitment assay

All β-arrestin recruitment assay were performed by Francois Marie Ngako Kadji and Junken Aoki. For the NanoBiT-β-arrestin recruitment assay[45], a receptor construct was designed to fuse the small fragment (SmBiT) of the NanoBiT complementation luciferase to the C-terminus of the ETR construct with a 15-amino acid flexible linker (GGSGGGGSGGSSSGG). A PCR-amplified ETR fragment and an oligonucleotide-synthesized SmBiT were assembled and inserted into the pCAGGS mammalian expression plasmid (a kind gift from Dr. Jun-ichi Miyazaki, Osaka University), using a NEBuilder HiFi DNA Assembly system (New England Biolabs). A β-arrestin construct was generated by fusing the large fragment (LgBiT), with nucleotide sequences gene-synthesized with mammalian codon optimization (Genscript), to the N-terminus of human β-arrestin1 (βarr1) with the 15-amino acid linker. The R393E and R395E mutations, which were shown to abrogate AP-2 binding and thus are defective in internalization[46], were introduced into LgBiT-βarr1 to facilitate the formation of the ETR-βarr1 complex. The plasmid encoding the ET_B-SmBiT construct or the ET_A-SmBiT construct was transfected, together with the plasmid encoding the internalization-defective LgBiT-βarr1, into HEK293A cells by the PEI method (1 μg ETR-SmBiT plasmid, 0.5 μg LgBiT-βarr1 plasmid, and 25 μl of 1 mg/ml PEI solution per 10-cm culture dish). After a one-day culture, the transfected cells were harvested with EDTA-containing Dulbecco's phosphate-buffered saline and resuspended in 10 ml of HBSS containing 5 mM HEPES and 0.01% BSA (BSA-HBSS). The cell suspension was seeded in a 96-well white plate at a volume of 80 μl per well and loaded with 20 μl of 50 μM coelenterazine (Carbosynth), diluted in BSA-HBSS. After an incubation at room temperature for 2 h, the background luminescent signals were measured using a luminescent microplate reader (SpectraMax L, Molecular Devices). Test compounds (6×, diluted in BSA-HBSS) were manually added to the cells (20 μl). After ligand addition, the luminescent signals were measured for 15 min at 20-s intervals. The luminescent signal was normalized to the initial count, and the fold-change values over 5–10 min after ligand stimulation were averaged. The fold-change β-arrestin recruitment signals were fitted to a four-parameter sigmoidal concentration–response, and the pEC_{50} and E_{max} values were obtained as described above.

2.2.6 Simulation system setup of ET_B-sarafotoxins and ET_A-sarafotoxins

All MD simulations were performed by Dr. Hirotake Miyauchi. The initial structures of S6b and S6c-bound receptors were prepared by using the program MODELLER[47] and the N-and C- terminus were truncated as observed in the crystal structure of the ET_B-S6b complex. All MD simulations were performed with NAMD2.12[48]. The simulation system was set at the condition of 96 × 96 × 96 $Å^3$ with 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane bilayer, 150 mM NaCl, and TIP3 water molecules. The missing hydrogen atoms were modified by the psfgen plugin of VMD[49]. The disulfide bonds and protonation states were properly set as the physiological condition. The net charge was neutralized by adding Cl-ions. The molecular topology and force field parameter from Charmm36[5051] were used for all simulations. The simulation systems were energy minimized for 1,000 steps with fixed positions of the non-hydrogen atoms. After minimization, another 1,000 steps of energy minimization with 10 kcal mol^{-1} restraints for the non-hydrogen atoms, except for the lipid molecules within 5.0 Å from the proteins. Next, equilibrations were performed for 0.1 ns under NVT conditions, with 10 kcal mol^{-1} $Å^{-2}$ restraints for the heavy atoms of the proteins. Finally, equilibration was performed for 1.0 ns under NPT conditions with the 1.0 kcal mol^{-1} $Å^{-2}$ restraints for all Cα atoms of the proteins. The production runs of the equilibrium simulations were performed for 50 ns without restraints, maintaining the constant temperature at 310 K using Langevin dynamics and the constant pressure at 1 atm using Nosé-Hoover Langevin piston[52]. The long-range electrostatic interactions were calculated by the particle mesh Ewald method[53].

2.3 Results and discussion

2.3.1 Functional characterization of ET-3 and IRL1620

I first investigated the biochemical activities of ET-3 and IRL1620 for the human endothelin receptors, by TGFα shedding (G-protein activation, specifically the G_q and the G_{12} families) and β-arrestin recruitment assays. The EC_{50} and E_{max} values of ET-3 for the ET_B receptor were similar to those of ET-1 in both assays, while the EC_{50} value for ET_A was about 5-fold lower (Figure 2-3, Figure 2-4). These data indicate that ET-3 functions as a full agonist for the endothelin receptors, with moderate ET_B-selectivity. The EC_{50} values of IRL1620 for the ET_B receptor were almost the same as those of ET-1 in both assays. In contrast, a 320 nM concentration of IRL1620 did not activate ET_A in the TGFα shedding assay (Figure 2-3). These data showed that IRL1620 is ET_B-selective by over

3000-fold, in excellent agreement with previous functional analyses17,29. However, despite its sub-nanomolar affinity, the E_{max} values of IRL1620 for the ET_B receptor were 88% (TGFα shedding assay) and 87% (β-arrestin recruitment assay) of the E_{max} value of ET-1 (Figure 2-3, FIgure 2-4), indicating that IRL1620 functions as a partial agonist for the ET_B receptor.

To obtain mechanistic insights into the different actions of these agonists, I performed X-ray crystal structural analyses of the human ET_B receptor in complex with ET-3 and IRL1620. For crystallization, I used the previously established, thermostabilized ET_B receptor (ET_B-Y5)27,30. IRL1620 also functions as a partial agonist for the thermostabilized receptor, as the E_{max} values for ET_B-Y5 were lower than those of ET-1 in both assays (84% and 85% in the TGFα shedding assay and the β-arrestin recruitment assay, respectively), while the EC_{50} values of IRL1620 were increased for ET_B-Y5 by about 9- and 6-fold in the TGFα shedding assay and β-arrestin recruitment assay, respectively, as compared to the wild type receptor (Figure 2-3, 2-4).

2.3.2 Crystallization of ET_B-ET-3 and ET_B-IRL1620

To facilitate crystallization, I replaced the third intracellular loop (ICL3) of the receptor with T4 Lysozyme (ET_B-Y5-T4L), and using in meso crystallization, I obtained crystals of ETB-Y5-T4L in complex with ET-3 and IRL1620 (Figure 2-5). In total, 757 and 68 datasets were collected for the ET-3- and IRL1620-bound receptors, respectively, and merged by the data processing system KAMO31. Eventually, I determined the ET_B structures in complex with ET-3 and IRL1620 at 2.0 and 2.7 Å resolutions, respectively, by molecular replacement using the ET-1-bound receptor (PDB 5GLH, Table 2-1). The datasets for the ET-3 bound receptor were mainly collected with an automated data-collection system, ZOO, which allowed the convenient collection of a large number of datasets. The electron densities for the agonists in both structures were clearly observed in the $F_o - F_c$ omit maps (Figure 2-6).

Using in meso crystallization, I obtained crystals of ETB-Y5-T4L in complex with S6b or S6c. The crystals of the S6c-bound receptor were tiny and diffracted X-rays to a maximum of only 7.0 Å resolution (Figure 2-5). The crystals of the S6b-bound receptor grew to their full size ($20 \times 20 \times 20$ μm^3) and diffracted X-rays to a maximum of 3.0 Å resolution (Figure 2-5). Therefore, 32 datasets were collected for the S6b-bound receptor and 16 datasets were merged by the data processing system KAMO. Eventually, I

determined the ET_B structures in complex with S6b at 3.0 Å resolution, by molecular replacement using the ET-3-bound receptor (PDB 6IGK, Table 2-1).

2.3.3 Overall structures

ETBR-ET-3

I first describe the ET_B structure in complex with ET-3. The overall structure consists of the canonical 7 transmembrane helices (TM), the amphipathic helix 8 at the C-terminus (H8), two antiparallel β-strands in the extracellular loop 2 (ECL2), and the N-terminus that is anchored to TM7 by a disulfide bond (Figure 2-7), and is similar to the previous ET-1-bound structure (overall R.M.S.D of 1.0 Å for the Cα atoms). Similar to ET-1, ET-3 adopts a bicyclic architecture comprising the N-terminal region (residues 1–7), the α-helical region (residues 8–17), and the C-terminal region (residues 18–21), and the N-terminal region is attached to the central α-helical region by the intrachain disulfide bond pairs (C1–C15 and C3–C11). The amino acid residues of the α-helical and C-terminal regions are highly conserved between ET-1 and ET-3 (Table 2-1), and the agonist peptides superimposed well (Figure 2-6).

Previous studies demonstrated that the N-terminal residues of the ETB receptor have a critical role in the virtually irreversible binding of the endothelins. As in the ET-1-bound structure, the N-terminal tail is anchored to TM7 via a disulfide bond between C90 and C358 in the ET-3-bound structure, constituting a lid that prevents agonist dissociation. The high-resolution ET-3-bound structure allowed more accurate tracing of the elongated N-terminal residues (Figure 2-8, Figure 2-9), as compared with the ET-1-bound structure, and revealed more extensive interactions with the agonist peptide. P88, I94, $Y247^{ECL2}$, and $K248^{ECL2}$ form a lid over ET-3, which is stabilized by a water-mediated hydrogen bonding network among the carbonyl oxygen of P93, the side chains of $Y247^{ECL2}$ and $K248^{ECL2}$, and D8 of ET-3 (Figure 2-8). In addition, three consecutive prolines (P87, 88, 89) stretch over the N-terminal region of ET-3, and two of them form van der Waals interactions with ET-3. Moreover, ECL1, 2 and the N-terminal residues form an extended water-mediated hydrogen bonding network over ET-3. These extensive interactions strongly prevent the agonist dissociation.

Accordingly, these regions form similar interactions with the receptor in both structures (Figure 2-25). In contrast, all of the residues, except for the disulfide bond-forming C1 and C3, are replaced with bulkier residues in ET-3 (Table 2-2). Despite these

sequence differences, the N-terminal regions are similarly accommodated in the orthosteric pocket of ECL2-TM5 linker in both structures, because these bulky residues are exposed to the solvent and interact poorly with the receptor (Figure 2-10). These structural features explain the similar high affinity binding of ET-3 to the ET_B receptor, as compared with ET-1.

ETBR-IRL1620

Next, I describe the ETB structure in complex with the partial agonist IRL1620, a linear peptide analogue of ET-117 (Table 2-3). Previous mutant and structural studies revealed that the N-terminal region contributes to the stability of the overall bicyclic structure by the intramolecular disulfide bonds, and thus facilitates the receptor interaction. IRL1620 completely lacks the N-terminal region, and consists of only the α-helical and C-terminal regions (Figure 2-11). Two cysteines in the α-helical region are replaced with alanines, and negative charges are introduced into the N-terminal end of the helix, by replacing lysine with glutamic acid (E9) and modifying the N-terminal amide group with a succinyl group (Fig. 3c). The consequent cluster of negative charges on the N-terminal end of IRL1620 (succinyl group, D8, E9, and E10) reportedly has an essential role in IRL1620 binding to the ET_B receptor. This cluster electrostatically complements the positively charged ET_B receptor pocket, which includes $K346^{6.58}$ and $R357^{ECL3}$ (Figure 2-12). Moreover, this negative cluster probably stabilizes the α-helical conformation of IRL1620, by forming a hydrogen-bonding cap at the N-terminally exposed amines. Due to these effects, IRL1620 adopts a similar helical conformation, even without the intramolecular disulfide bonds (Figure 2-13, Figure 2-14), and forms essentially similar interactions with the receptor, as compared with the endogenous agonists, ET-1 and ET-3 (Figure 2-25). These structural features are consistent with the high affinity of IRL1620 to the ET_B receptor, which is comparable to those of ET-1 and ET-3 (Figure 2-7). Nevertheless, the N-terminal side of the α-helical region of IRL1620 is less visualized in the electron density, suggesting its higher flexibility as compared to those of ET-1 and ET-3 (Figure 2-12), probably due to the lack of the N-terminal region. In contrast to the quasi-irreversible binding of ET-1, IRL1620 binding is reportedly reversible35. Such structural differences may account for the different dissociation properties of these agonists.

ETBR-sarafotoxin S6b

The overall structure of Sarafotoxin S6b-bound ET_B receptor is essentially similar to

those of the ET-1- and ET-3-bound receptors (Figure 2-15). However, there is a remarkable difference on the intracellular side. In the S6b-bound receptor, the intracellular end of TM5 is opened up by 4 Å, as compared with those in the ET-1- and ET-3-bound receptors (Figure 2-16). A number of previous studies have shown that the intracellular portion of TM6 moves outwardly by 10–14 Å upon G-protein coupling. Moreover, in NTSR1 (neurotensin receptor type 1), agonist and G-protein binding induce the outward displacement of TM5 by 2–4 Å (Figure 2-16). Although T4L insertion affects the conformations of TM5 and TM6, agonist binding may cause the outward displacement of the TM5 in the ETB receptor, as in NTSR1.

2.3.4 Insight into B-type selectivity of IRL1620 and Sarafotoxins

IRL1620 does not bind to the ET_A receptor at all in the same concentration range, confirming its high selectivity for the ET_B receptor (Figure 2-3, Figure 2-4). To elucidate the mechanism of this selectivity, I compared the amino acid compositions of the IRL1620 binding sites between the ET_B and ET_A receptors (Figure 2-17a). While the transmembrane region is highly conserved, the residues in ECL2 are diverse. In particular, the hydrophobic residues L252 ^{ECL2} and I254 ^{ECL2} are replaced with the polar residues H236 and T238 in the ETA receptor, respectively (Figure 2-17b). These residues form extensive hydrophobic interactions with the middle part of IRL1620. However, the double mutation of L252H and I254T only reduced the potency of IRL1620 by 2-fold in the TGFα shedding assay (Figure 2-17c), suggesting that these residues are not the sole determinants for the receptor selectivity of IRL1620. Therefore, I focused on other residues of ECL2. In the ET_B receptor, $P259^{ECL2}$ and $V260^{ECL2}$ generate a short kink on the loop between the β-strand and TM5, but the ET_A receptor has a truncated loop region and completely lacks these residues (Figure 2-17b). In addition, the ET_A receptor has a proline (P228) in the first half of ECL2, which should disturb the β-strand formation as in the ET_B receptor (Figure 2-17b). These observations suggest that ECL2 adopts completely different structures between the ET_A and ET_B receptors. Moreover, ECL1 is also different between the two receptors, as the ET_A receptor has a five amino-acid elongation as compared to the ET_B receptor (Figure 2-18). Since ECL1 interacts with the β-strands in ECL2 (Figure 2-17a), this elongation could affect the orientation of the β-strands. Overall, the sequence divergences in the extracellular loops suggest that the ET_A and ET_B receptors have different extracellular architectures in these regions, which may account for their different selectivities for the isopeptides.

Sarafotoxin S6b and S6c have similar affinities towards the ET_B receptor, whereas S6c has reduced affinity for the ET_A receptor by about 100- to 10,000-fold. Between S6b and S6c, three residues are diverged (S2T, K9E, and Y13N) (Table 2-4). Moreover, a previous study showed that [E9]S6b functions as a ET_B-selective agonist and is approximately equipotent with S6c, suggesting that the charge of the 9th residues in the sarafotoxins determines the subtype selectivity of the sarafotoxins. To obtain structural insight into the subtype selectivity of the sarafotoxins, I performed molecular dynamics simulations of the human ET_B and ET_A receptors in complex with S6b or S6c. I focused on the receptor interactions with the 9th residues of the sarafotoxins, which are located at the N-terminal end of the α-helical regions. In the S6b-bound ET_B receptor, K9 forms a salt bridge with $D246^{ECL2}$ in ECL2, anchoring the α-helical region to ECL2 (Figure 2-19a). In the S6c-bound ET_B receptor, $D246^{ECL2}$ repels E9, displacing the α-helical region toward TM6-7 (Figure 2-19b), thereby E9 forms a hydrogen bond with $Y247^{ECL2}$. Nevertheless, S6b and S6c have similar affinities for the ETB receptor, suggesting that the polar interaction between the 9th residue and receptor is not critical for sarafotoxin binding. Instead, the extensive hydrophobic residues in ECL2 play an important role in sarafotoxin binding (Figure 2-19a and Figure 2-19b).

However, the hydrophobic residues $L252^{ECL2}$ and $I254^{ECL2}$ in the ET_B receptor are replaced with the polar residues $H236^{ECL2}$ and $T238^{ECL2}$ in the ET_A receptor, respectively. These replacements would reduce the hydrophobic interactions with the sarafotoxin α-helical region. Instead, in the S6b-bound ET_A receptor, $R232^{ECL2}$ forms a salt bridge with E10 (Figure 2-19c), and the N-terminal residues of the ET_A receptor cover the α-helical region. These interactions compensate for the weaker hydrophobic interactions with the α-helical region and allow the high-affinity binding of S6b. In the S6c-bound ETA receptor, E9 forms an ionic interaction with R232ECL2 together with E10 (Figure 2-19d), but still repels $E230^{ECL2}$. Due to the interaction, the N-terminal residues of the ET_A receptor cannot cover the α-helical region of S6c, thus opening up the ligand binding pocket. Overall, E9 in S6c may rearrange the interaction over the α-helical region, consequently destabilize the interaction with the ET_A receptor, and reduce the affinity.

2.3.5 ETR activation mechanism

A comparison of the intermembrane parts revealed further differences in the allosteric coupling between the orthosteric pocket and the intermembrane part. Previous studies

have shown that the agonist binding induces the disruption of the hydrogen-bonding network around D147$^{2.50}$, which connects TMs 2, 3, 6, and 7 and stabilizes the inactive conformation of the ET_B receptor (Figure 2-22). The present high-resolution ET-3-bound structure provides a precise mechanistic understanding of this rearrangement (Figure 2-22, Figure 2-23). In particular, the water-mediated hydrogen bonds involving D147$^{2.50}$, W336$^{6.48}$, and N378$^{7.45}$ in the inactive conformation collapse upon ET-3 binding, by the inward motions of TMs 2, 6, and 7. The W336$^{6.48}$ side chain moves downward by about 2.5 Å, resulting in the disruption of the water-mediated hydrogen bond with D147$^{2.50}$, and consequently, the D147$^{2.50}$ side chain moves downward by about 3 Å and forms hydrogen bonds with the N382$^{7.49}$ and N119$^{1.50}$ side chains. The N378$^{7.45}$ side chain also moves downward by about 1.5 Å and forms a hydrogen bond with the nitrogen atom of the W336$^{6.48}$ side chain. The downward movements of the W336$^{6.48}$ and N378$^{7.45}$ side chains consequently induce the outward repositioning of the F332$^{6.44}$ side chain and the middle part of TM6, by about 1 Å. W6.48 and F6.44 are considered to be the transmission switch of the class A GPCRs, which transmits the agonist-induced motions to the cytoplasmic G-protein coupling interface. Overall, our results show that the collapse of the water-mediated hydrogen-bonding network involving D147$^{2.50}$, W336$^{6.48}$, and N378$^{7.45}$ propagates as the structural change in the transmission switch, and probably induces the outward displacement of the cytoplasmic portion of TM6 upon G-protein activation (Figure 2-23).

2.3.6 partial activation mechanism of IRL1620

To elucidate the mechanism of the partial activation by IRL1620, I compared the IRL1620-bound structure with the full-agonist ET-3-bound structure (Figure 2-20). IRL1620 forms essentially similar receptor interactions to those of the α-helical and C-terminal regions of ET-3 (Figure 2-18). The intracellular portions of the receptors are quite similar between the ET-3- and IRL1620-bound structures, in which TM7 and H8 adopt active conformations, while the remaining parts of the receptors still represent the inactive conformation of GPCRs (Figure 2-20, Figure 2-21). On the extracellular side, IRL1620 induces similar conformational changes to those observed in the ET-1- and ET-3-bound structures; namely, the large inward motions of TM2, 6, and 7, which are critical for receptor activation (Figure 2-20a). However, the extent of the inward motion of TM6-7 is smaller by about 1 Å in the IRL1620-bound structure, as compared with that in the

ET-3-bound structure, due to the different ligand architectures between IRL1620 and ET-3. Since IRL1620 lacks the N-terminal region, the orthosteric pocket of the receptor has more space, and consequently the α-helical region of IRL1620 is tilted differently toward TM6 (Figure 2-20c). In addition, while the N-terminal region of ET-3 interacts with TM6 of the receptor, by forming a hydrogen bond between the carbonyl oxygen of T2 and K346$^{6.58}$, IRL1620 lacks this interaction, resulting in the different orientation of TM6-7. As TM6 has an especially important role for the cytoplasmic G-protein binding, this difference is probably related to the partial agonist activity of IRL1620.

IRL1620 induces a similar but slightly different rearrangement of the hydrogen bonding network in the intermembrane part (Figure 2-23). Due to the smaller inward shift of the extracellular portion of TM6, the downward shift of the W336$^{6.48}$ side chain is smaller in the IRL1620-bound structure, and it still forms a water-mediated hydrogen bond with the D147$^{2.50}$ side chain. Consequently, the D147$^{2.50}$ side chain forms a direct hydrogen bond with N378$^{7.45}$, thereby preventing the downward motion of N378$^{7.45}$ and the hydrogen bond formation between N378$^{7.45}$ and W336$^{6.48}$. Overall, the downward motions of the W336$^{6.48}$ and N378$^{7.45}$ side chains are only moderate, as compared to those in the ET-3-bound structure, and the hydrogen-bonding network involving D147$^{2.50}$, W336$^{6.48}$, and N378$^{7.45}$ is partially preserved in the IRL1620-bound structure (Figure 2-22, Figure 2-23). Accordingly, in the IRL1620-bound structure, the position of the "transmission switch" residue F332$^{6.44}$ is in between those of the active (ET-3-bound) and inactive (K-8794-bound) structures (Figure 2-23). This intermediate position of F332$^{6.44}$ should partly prevent the outward displacement of the cytoplasmic portion of TM6 that is required for G-protein activation. Overall, the smaller inward shift of the extracellular potion of TM6 and the preserved interactions at the receptor core account for the partial agonistic activity of IRL1620 (Figure 2-24).

Figures and Tables of Chapter 2

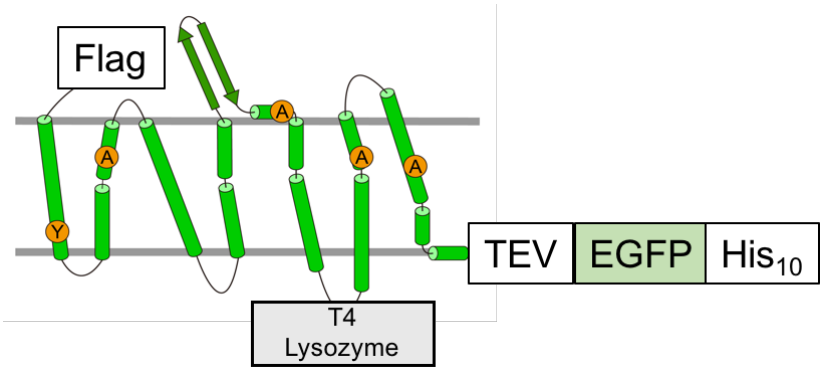


Figure 2-1 Construct of ET_B-Y5-T4L

As in the crystal structure of the ET-1-bound ET_B receptor, a Flag tag is added to the N-terminus of the extracellular and a His tag to the C-terminus of the intraceller in the thermos stablized mutant ET_B-Y5-T4L. In addition, a T4 Lysozyme was inserted into ICL3 for structural stabilization.

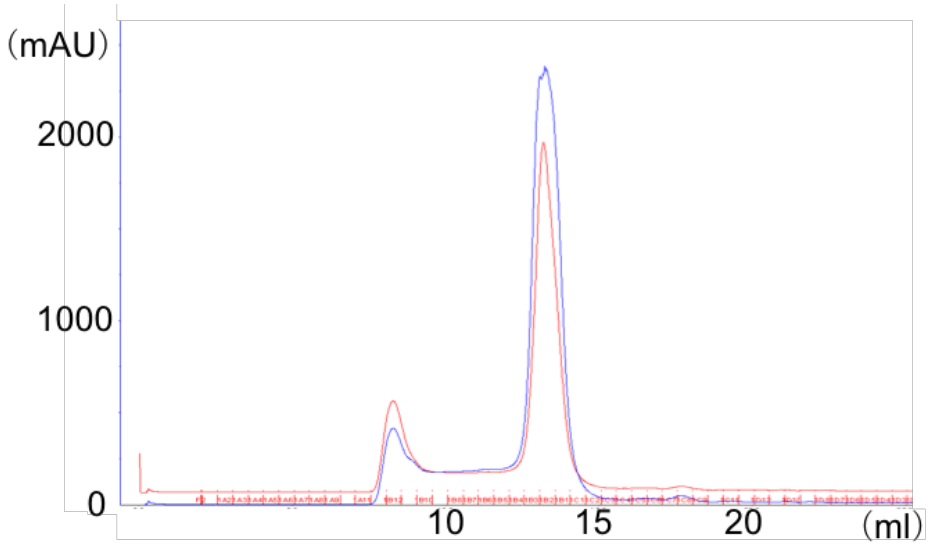
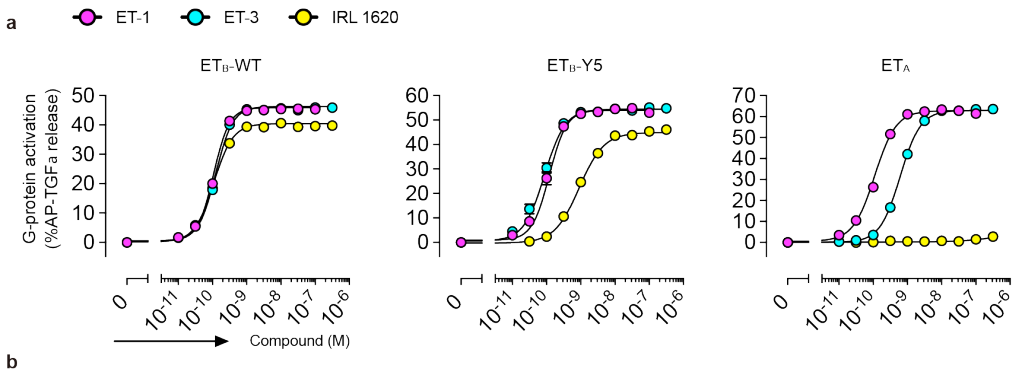


Figure 2-2 Gel filtration of ET_B-Y5-T4L

Size-exclusion chromatogram of the ET_B-Y5-T4L. Blue and red lines indicate absorbances at 280 nm and 260 nm, respectively.

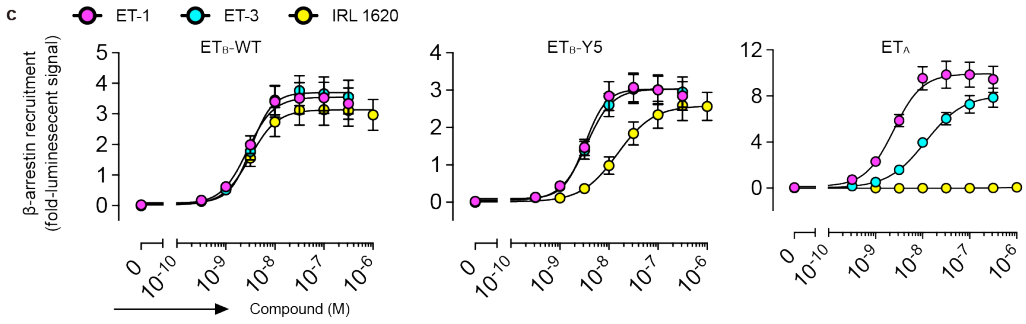


b

Ligand	ET_B (n=7) EC_{50}, nM (pEC_{50} ± SEM)	E_{max} ± SEM, %	ET_B-Y5 (n=7) EC_{50}, nM (pEC_{50} ± SEM)	E_{max} ± SEM, %	ET_A (n=5) EC_{50}, nM (pEC_{50} ± SEM)	E_{max} ± SEM, %
ET-1	0.11 (9.94 ± 0.04)	100	0.099 (10.01 ± 0.07)	100	0.13 (9.89 ± 0.04)	100
ET-3	0.13 (9.89 ± 0.04)	100.5 ± 0.6 (NS)	0.077 (10.12 ± 0.06)	100.9 ± 0.3 (NS)	0.65 (9.19 ± 0.03)	101.2 ± 0.8 (NS)
IRL 1620	0.11 (9.97 ± 0.03)	87.6 ± 1.0***	0.92 (9.04 ± 0.04)	83.6 ± 1.4***	NA	NA

Figure 2-3 TGF α shedding assay

Concentration-response curves of AP-TGFα release in the ET-1, ET-3, and IRL1620 treatments of HEK293 cells expressing the indicated ET_A and ET_B receptors. Symbols and error bars are means and s.e.m. (standard error of the mean) of five or seven independent experiments, each performed in triplicate. Note that the error bars are smaller than the symbols for most data points. All TGF α shedding assay were performed by Francois Marie Ngako Kadji and Junken Aoki.



d

Ligand	ET_B (n=4) EC_{50}, nM (pEC_{50} ± SEM)	E_{max} ± SEM, %	ET_B-Y5 (n=4) EC_{50}, nM (pEC_{50} ± SEM)	E_{max} ± SEM, %	ET_A (n=4) EC_{50}, nM (pEC_{50} ± SEM)	E_{max} ± SEM, %
ET-1	2.7 (8.56 ± 0.02)	100	3.3 (8.48 ± 0.02)	100	2.3 (8.63 ± 0.05)	100
ET-3	3.3 (8.49 ± 0.03)	104.8 ± 0.5 (NS)	3.5 (8.46 ± 0.02)	99.0 ± 1.9 (NS)	10.8 (7.97 ± 0.05)	81.1 ± 0.7***
IRL 1620	3.1 (8.51 ± 0.05)	87.3 ± 2.7***	15.1 (7.82 ± 0.08)	85.4 ± 3.3***	NA	NA

Figure 2-4 β-Arrestin recruitment assay

Concentration–response curves of β-arrestin recruitment in the ET-1, ET-3, and IRL1620 treatments of HEK293 cells expressing the ET_A and ET_B receptors. Symbols and error bars are means and s.e.m. of four independent experiments, each performed in duplicate. All β-arrestin recruitment assay were performed by Francois Marie Ngako Kadji and

Junken Aoki.

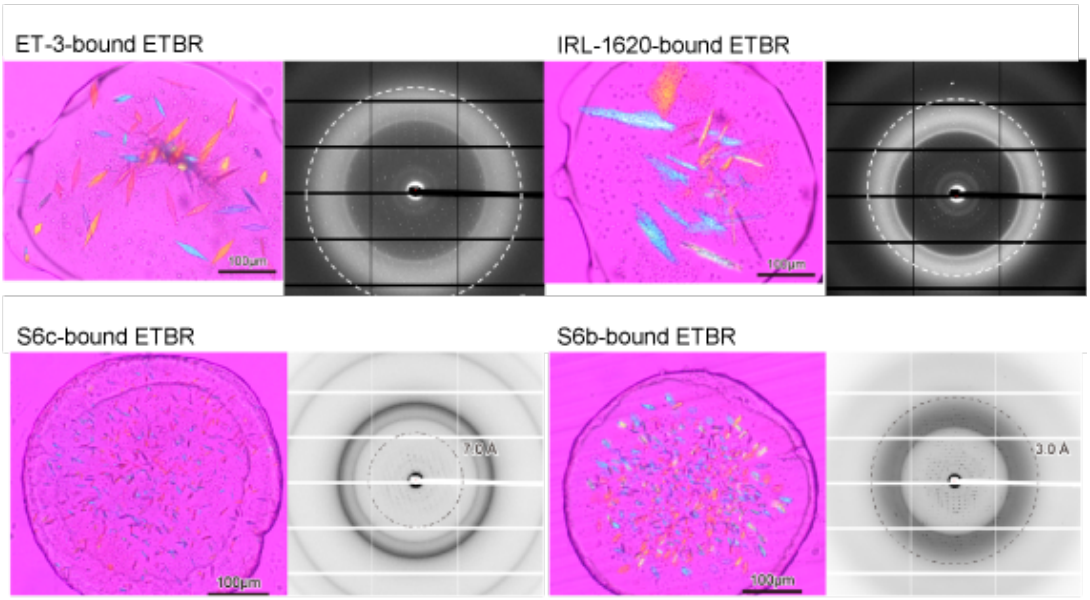


Figure 2-5 Crystals and Diffraction images

Crystals (left) and diffraction images (right) of agonist-bound ET_B receptors.

Table 2-1 Data collection and refinement statistics

	ET-3	IRL-1620	sarafotoxin S6b
Data collection			
Space group	$C222_1$	$C222_1$	$P22_12_1$
Cell dimensions			
a, *b*, *c* (Å)	65.5, 172.3, 121.3	100.0, 303.9, 60.2	61.12 82.28 166.78
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)*	50-2.00 (2.12-2.00)	50-2.70 (2.80-2.70)	49.3 - 3.0 (3.107 - 3.0)
R_{meas}*	0.860 (18.057)	0.499 (5.826)	0.2802 (3.504)
R_{pim}*	0.060 (1.277)	0.108 (1.250)	
$\langle I/\sigma(I)\rangle$*	20.7 (0.95)	9.0 (1.4)	9.06 (0.91)
$CC_{1/2}$*	0.998 (0.559)	0.981 (0.382)	0.994 (0.545)
Completeness (%)*	100.0 (100.0)	99.99 (100.0)	99.73 (99.94)
Redundancy*	204.6 (199.2)	21.5 (21.5)	12.3 (12.6)
Refinement			
Resolution (Å)	50-2.00	50-2.70	49.3-3.0
No. reflections	46,785	25,764	17530
R_{work} / R_{free}	0.1834 / 0.2271	0.2066 / 0.2335	0.2697 / 0.3036
No. atoms			
Protein	3914	3871	3708
Ligand/ion	296	108	176
Water	175	59	
Averaged *B*-factors (Å^2)			
Protein	44.5	64.8	100.58
Ligand/ion	79.2	80.8	93.35
Water	49.7	40.8	
R.m.s. deviations from ideal			
Bond lengths (Å)	0.007	0.002	0.003
Bond angles (°)	0.862	0.517	0.56
Ramachandran plot			
Favored (%)	98.6	97.3	96.27
Allowed (%)	1.2	2.5	3.52
Outlier (%)	0.2	0.2	0.21

*Values in parentheses are for highest-resolution shell.

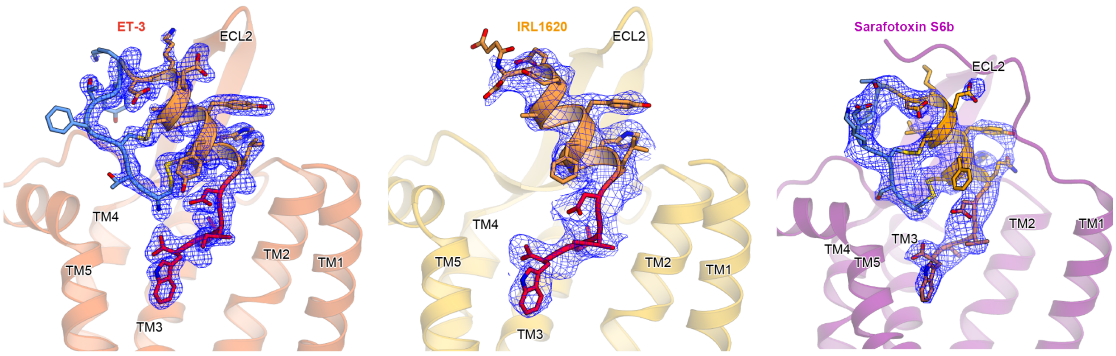


Figure 2-6 Electron densities

Fo-Fc omit maps for ET-3, IRL1620, and Sarafotoxin S6b. All structures are contoured at 2.0 σ .

Table 2-2 Comparison of the amino acid sequences of ET-1 and ET-3

	N-terminal		α-helical		C-terminal	
ET subtype	1	7	8	17	18	21
ET-1	CSCSSLM		DKECVYFCHL		DIIW	
ET-3	CTCFTYK		DKECVYYCHL		DIIW	

N-terminal, alpha-helical, and C-terminal regions are coloured blue, Yellow, red, respectively. Different residues between ET-1 and ET-3 are labeled in black.

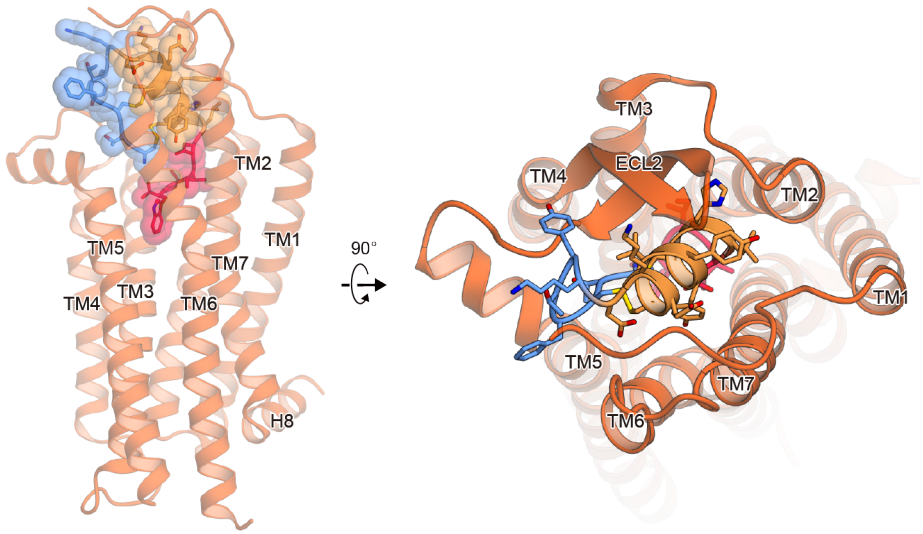


Figure 2-7 The overall structure of the ET-3-bound ETB receptor

The side view (left) and top view (right) of the ET-3-bound ET_B receptor. The receptor is shown as an orange-red ribbon model. ET-3 is shown as a transparent surface representation and a ribbon model, with its N-terminal region coloured cyan, α-helical region orange, and C-terminal region deep pink. The side chains of ET-3 are shown as sticks.

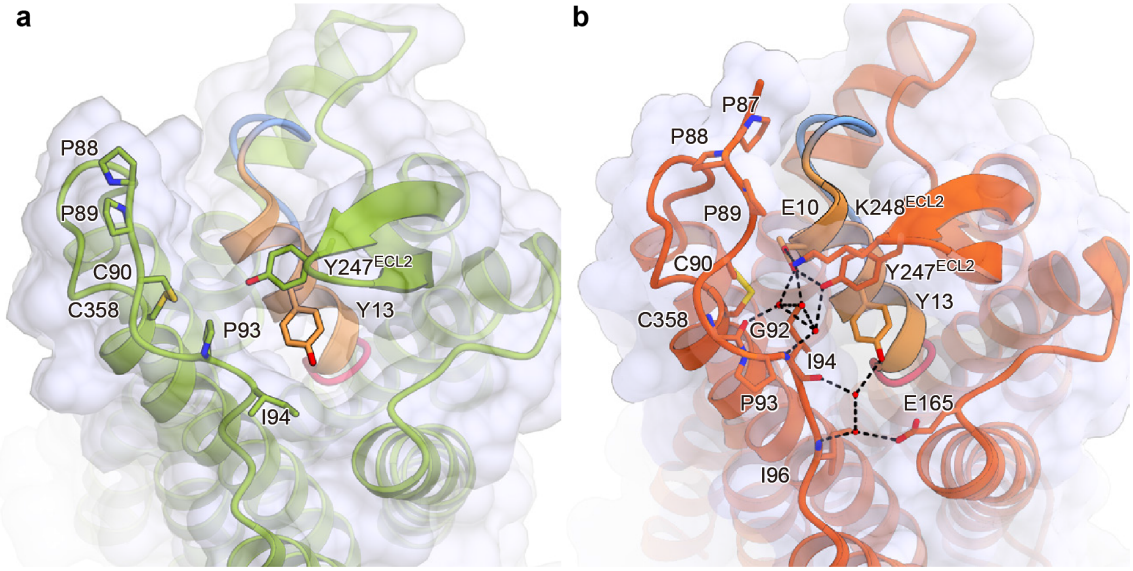


Figure 2-8 Comparison of the N-terminus and ECL2 structure in ET-1 and ET-3
ET-1- (a) and ET-3- (b) bound structures. The receptors are shown as transparent surface representations and ribbon models. Endothelins are shown as ribbon models. The residues involved in the irreversible binding of the endothelins are shown as ballsticks.

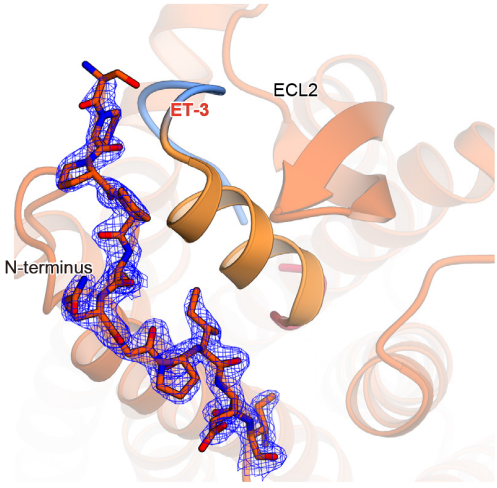


Figure 2-9 $2F_o - F_c$ map of the N-terminus in the ET-3-bound structure

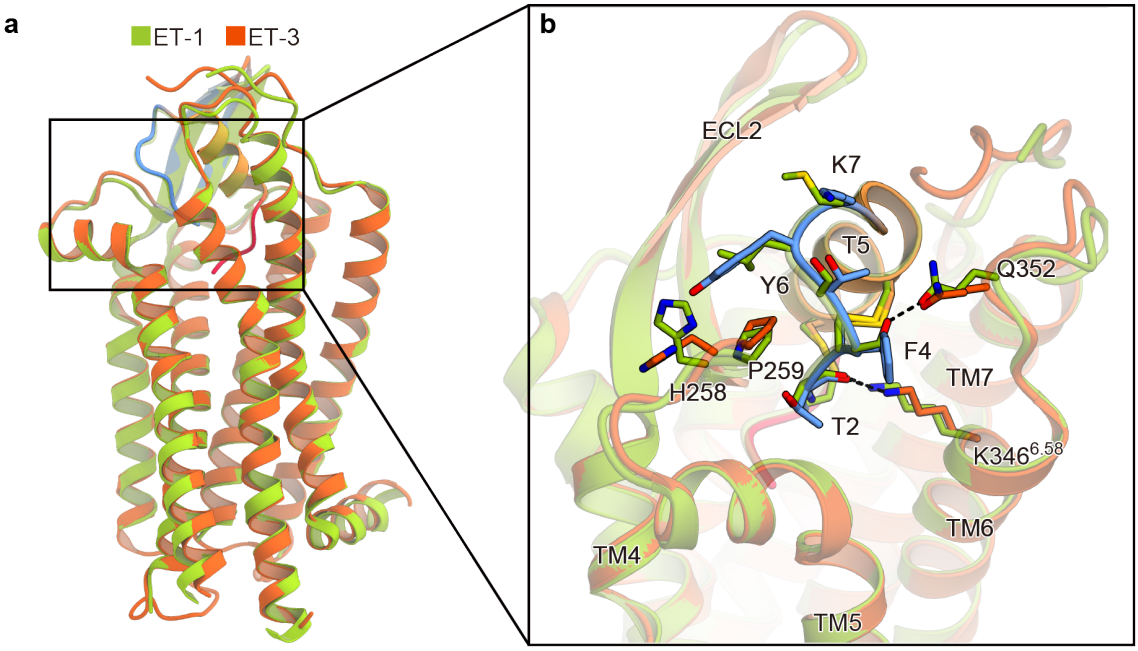


Figure 2-10 Comparison of ET-1- and ET-3-bound ET_B structure

Superimposition of the ET-3 and ET-1-bound ET_B receptors, coloured orange-red and green, respectively, viewed from the membrane plane **(a)**, and from the extracellular side **(b)**. The side chains of ET-1 and ET-3 are shown as sticks.

Table 2-3 Comparison of the amino acid sequences of ET-1 and IRL1620

	N-terminal		α-helical		C-terminal	
ET subtype	1	7	8	17	18	21
ET-1	CSCSSLM		DKECVYFCHL		DIIW	
IRL1620			SucDEEAVYFAHL		DIIW	

N-terminal, alpha-helical, and C-terminal regions are coloured blue, Yellow, red, respectively. Different residues between ET-1 and IRL1620 are labeled in black.

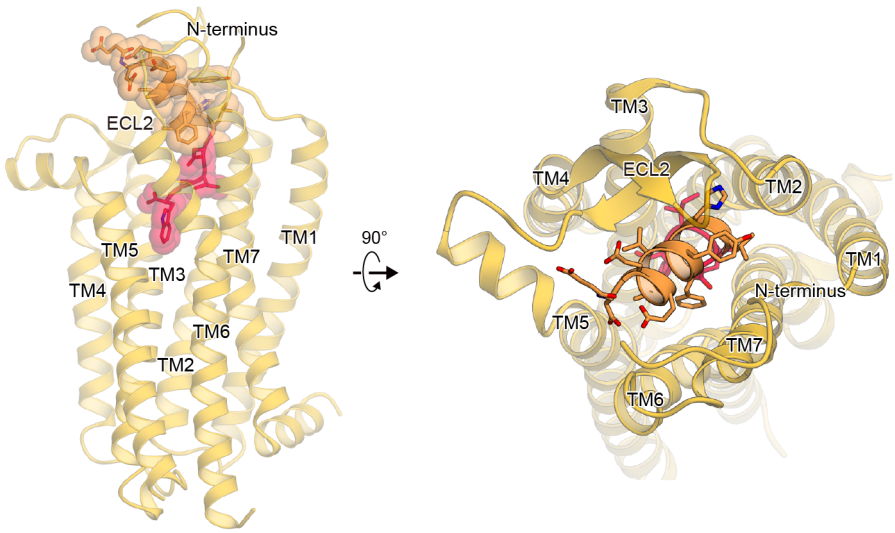


Figure 2-11 The overall structure of the IRL-1620-bound ETB receptor

The side view (left) and top view (right) of the IRL1620-bound ET_B receptor. The receptor is shown as an yellow ribbon model. IRL1620 is shown as a transparent surface representation and a ribbon model, with the α-helical and C-terminal regions coloured orange and deep pink, respectively.

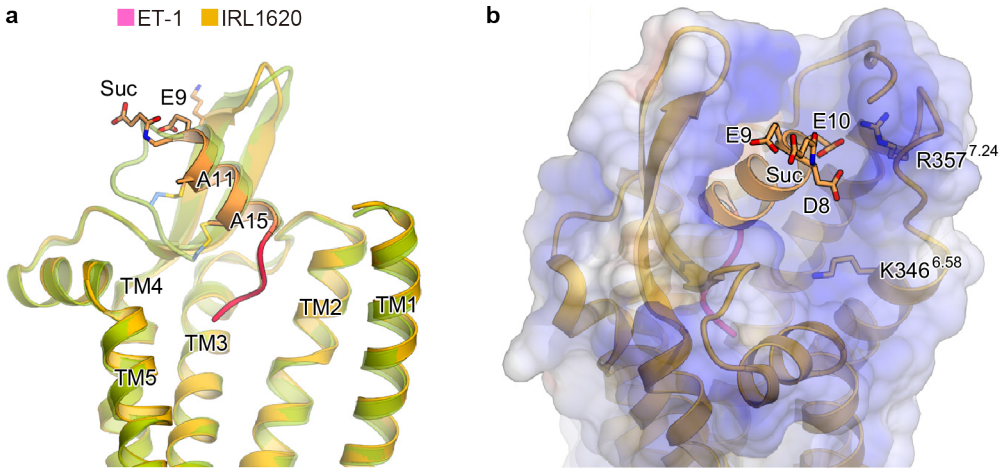


Figure 2-12 N-terminus truncated agonist IRL1620

(a) Superimposition of the IRL1620- and ET-1-bound receptors viewed from the membrane plane, coloured yellow and green, respectively. The ET_B receptors and agonists are shown as ribbon models. The different residues between ET-1 and IRL1620 are shown as sticks. **(b)** Electrostatic surfaces of the IRL1620-bound ET_B structure. The

negatively charged moieties on the N-terminal end of IRL1620 and the positively charged residues in the extracellular side of TMs 6 and 7 are shown as sticks.

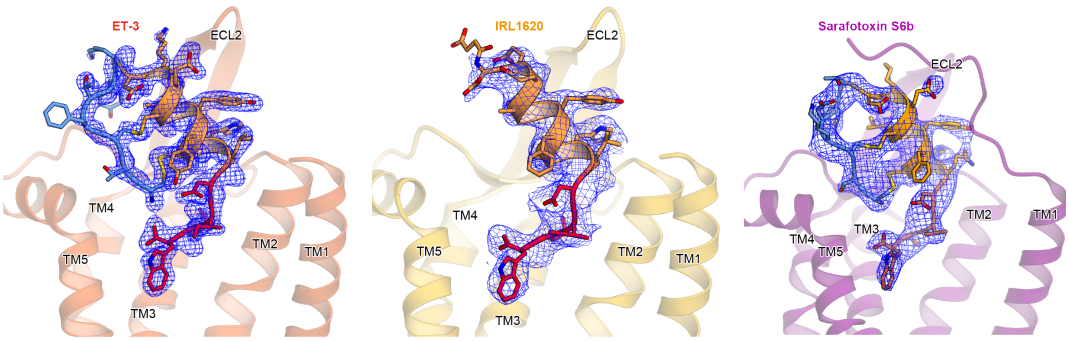


Figure 2-13 Comparison of ET-1 and IRL1620 structures in the ETBR complex.
The structures of ET-3- (left), IRL1620- (center), and S6b (right). $F_o - Fc$ omit maps for each ligands are shown in mesh.

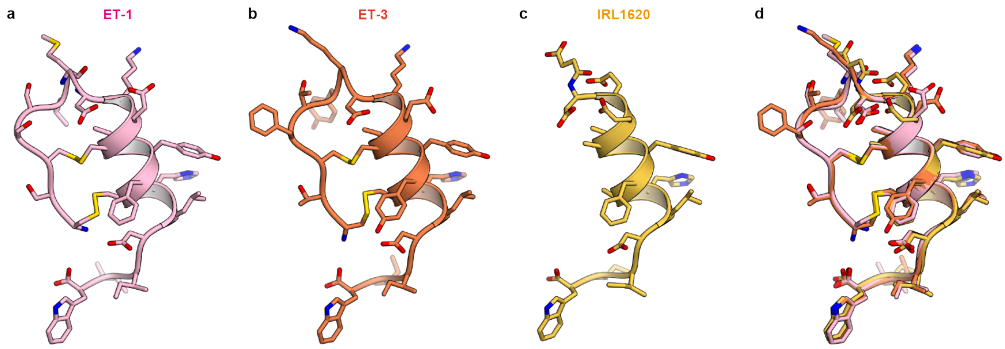


Figure 2-14 Comparison of agonist structures
Structures of ET-1 (a), ET-3 (b), IRL1620 (c), Sarafotoxin s6b in the complex structures, coloured pink, orange-red, yellow, purple, respectively. The agonists are shown as ribbon and stick models.

Table 2-4 Sequences of Sarafotoxin S6b and S6c

	N-terminal		α-helical		C-terminal	
Agonists	1	7	8	17	18	21
ET-1	CSCSSLM		DKECVYFCHL		DIIW	
Sarafotoxin S6b	CSCKDMT		DKECLYFCHQ		DVIW	
Sarafotoxin S6c	CTCNDMT		DEECLNFCHQ		DVIW	

N-terminal, alpha-helical, and C-terminal regions are coloured blue, Yellow, red, respectively. Different residues between ET-1 and sarafotoxins are labeled in black.

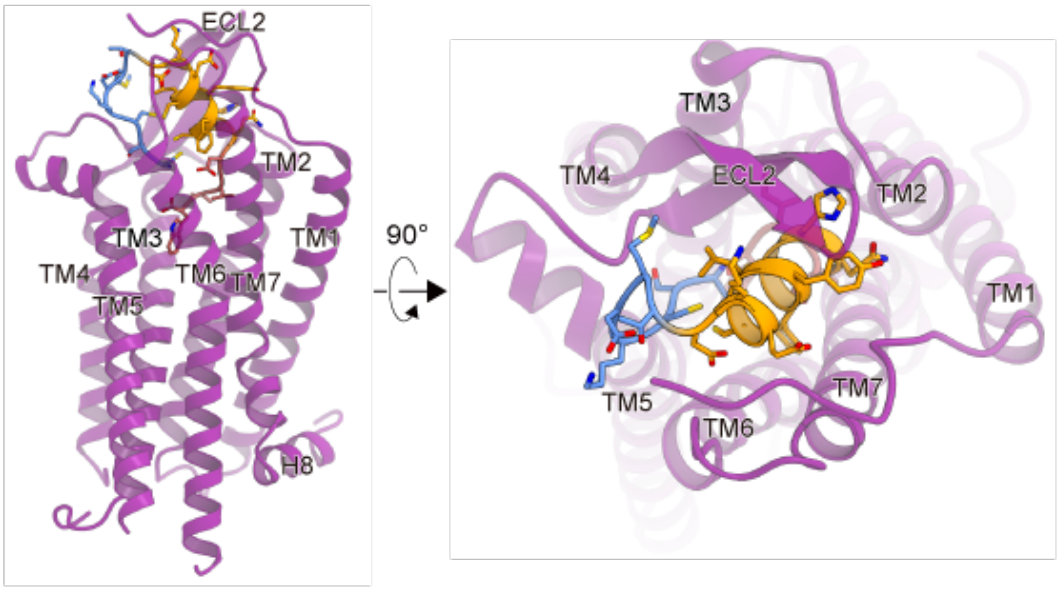


Figure 2-15 Overall structure of sarafotoxin S6b-ETBR

The side view (left) and top view (right) of the S6b-bound structure. The receptor is shown as an purple ribbon model. S6b is shown as a transparent surface representation and a ribbon model, with its N-terminal region coloured cyan, α-helical region orange, and C-terminal region deep pink. The side chains of S6b are shown as sticks.

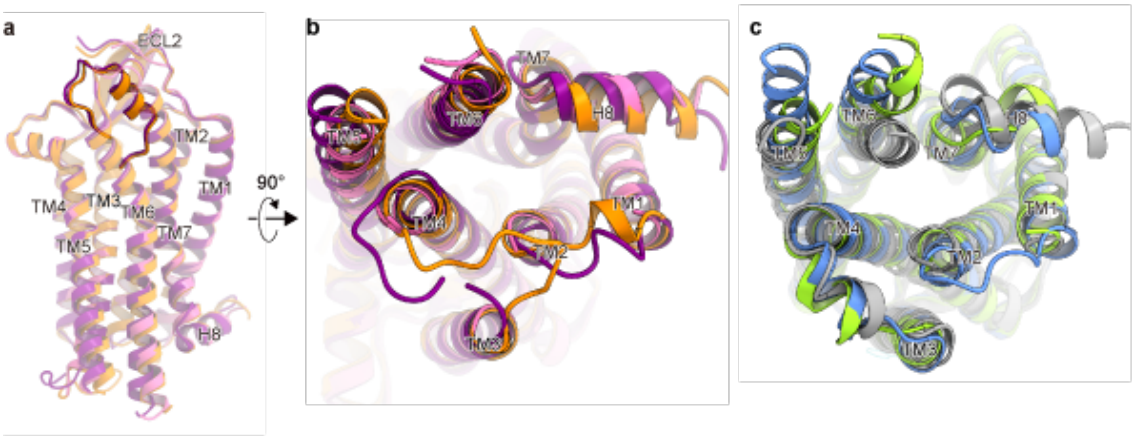


Figure 2-16 Intracellular TM5 outward shift of S6b-bound ETBR

(a), (b), Superimposed ETB structures in complex with S6b (purple), ET-1 (pink), and ET-3 (orange), viewed from the membrane plane **(a)** and intracellular side **(b). (c),** Intracellular views of the superimposed structures of the C-state hNTSR1 in complex with Gi1 (blue), active-state rNTSR1 (green), and inactive-state rNTSR1 (grey).

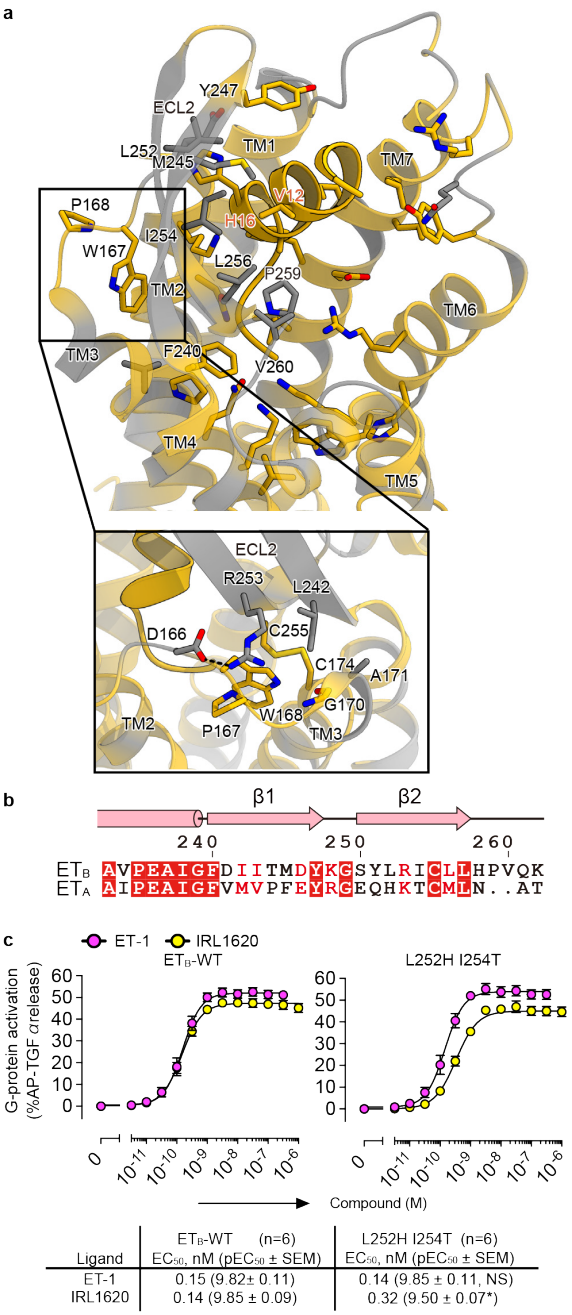


Figure 2-17 Conservation of the IRL1620 binding site

(a) Sequence conservation between ETRs, mapped onto the IRL1620-bound structure. The upper panel shows the conservation of the residues constituting the IRL1620 binding site. Conserved and non-conserved residues are coloured yellow and grey, respectively. The ET_B receptor is shown as ribbons, and the residues involved in IRL1620 binding are shown as sticks. Moreover, W167 and P168 in ECL1 are shown as sticks. The lower panel shows the conservation of the ECL1-ECL2 interface. **b** Alignment of the amino acid

sequences of the human ET_B and ET_A receptors, focused on ECL2. **c** Concentration-response curves of AP-TGFα release response upon ET-1 or IRL1620 treatments of HEK293 cells expressing the wild-type (WT) ET_B receptor or the L252H I254T double mutant. Symbols and error bars are means and s.e.m. of six independent experiments, each performed in triplicate.

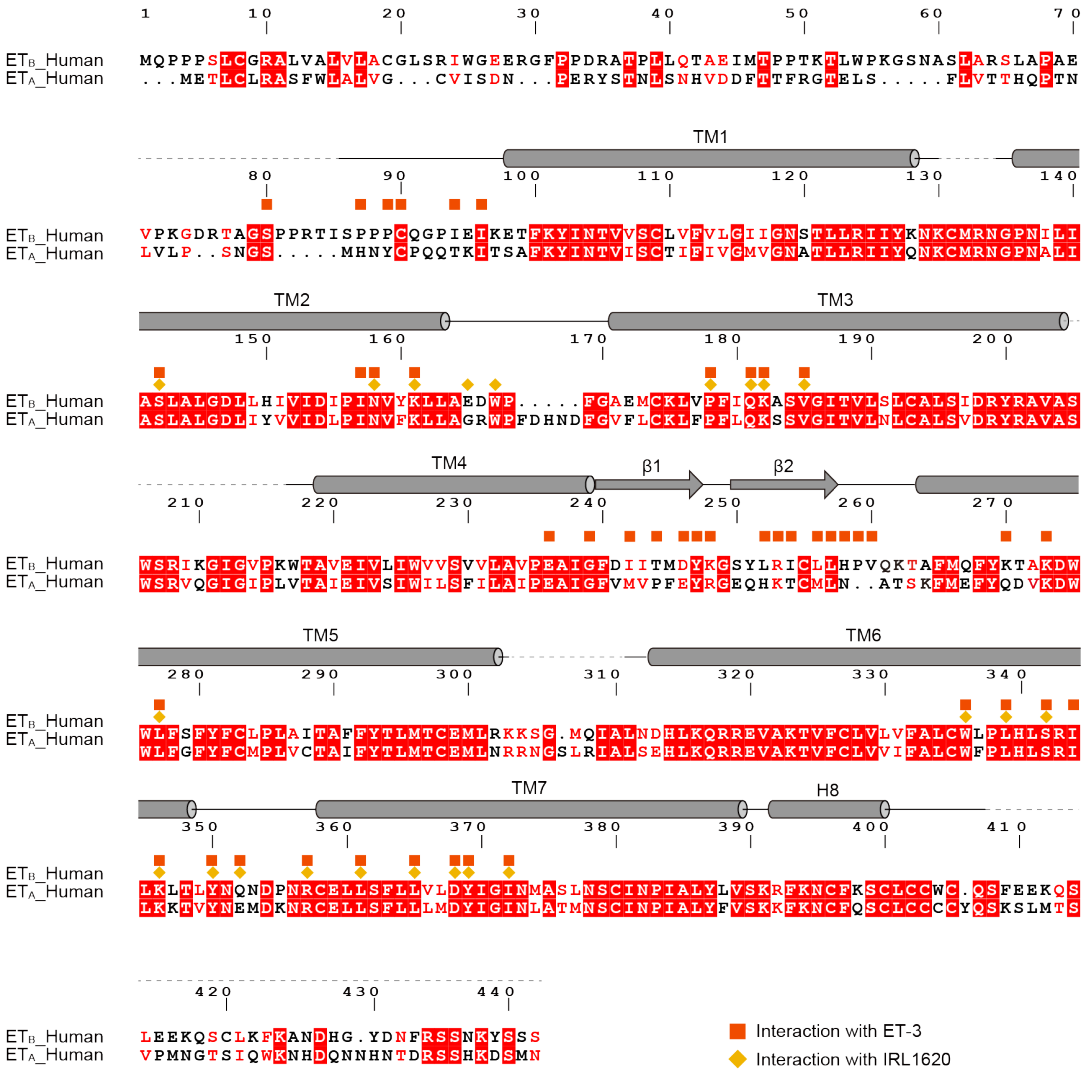


Figure 2-18 Alignment of human ET_A and ET_B receptor sequences

Alignment of the amino acid sequences of the human ET_B receptor (UniProt ID: P24530) and human ET_A receptor (P25101). Secondary structure elements for α-helices and β-strands are indicated by cylinders and arrows, respectively. Conservation of the residues between ET_A and ET_B is indicated as follows: red panels for completely conserved; red letters for partially conserved; and black letters for not conserved. The residues involved

in the ET-3 and IRL1620 binding are indicated with squares and diamonds, respectively.

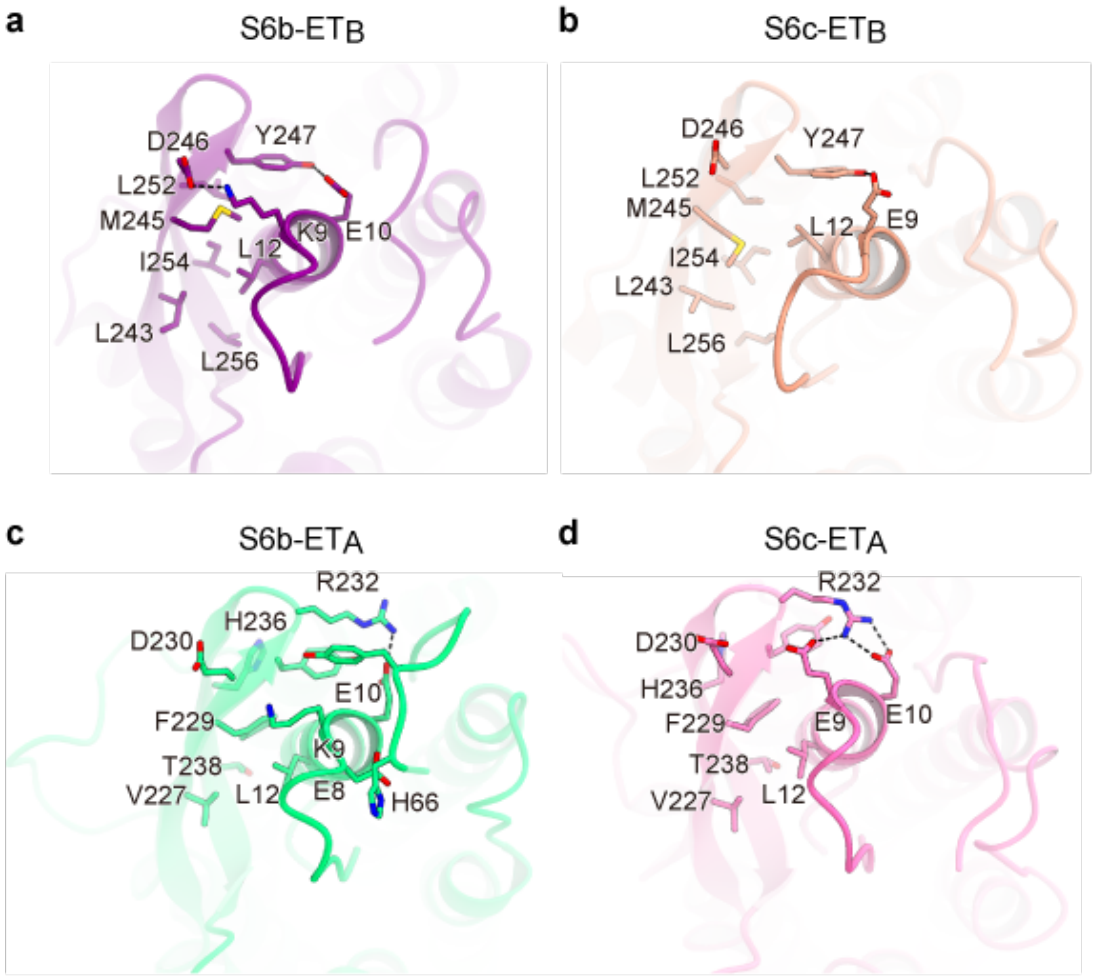


Figure 2-19 Molecular dynamics simulation of human ETA and ETB receptors in complex with S6b and S6c

a-d, Models of the human ET_A and ET_B receptors in complex with S6b or S6c, obtained by molecular dynamics simulations. Interactions between sarafotoxins and ECL2 and N-terminus of the receptors are shows as sticks. All MD simulations were performed by Dr. Hirotake Miyauchi.

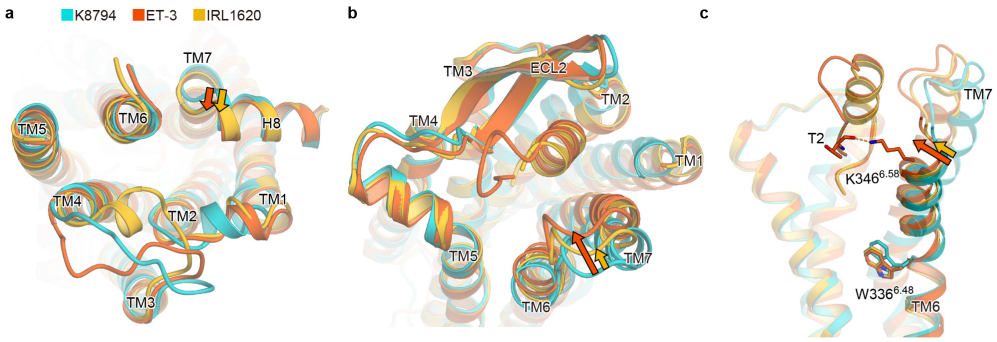


Figure 2-20 Comparison of the K8794, ET-3, and IRL1620-bound structures

Comparison of the K-8794-bound inactive (PDB 5X93), IRL1620-bound partially active, and ET-3-bound active ETB structures, coloured turquoise, orange, and orange-red, respectively. The ETB receptors and agonist peptides are shown as ribbon models, viewed from the intracellular side **(a),** the extracellular side **(b)**, and the membrane plane **(c)**. The cytoplasmic cavity for the G-protein binding is still hindered by the inwardly-oriented TM6 in the ET-3 and IRL1620-bound receptors, representing the inactive conformations. This is consistent with the notion that the fully active conformation is only stabilized when the G-protein is bound, as shown in the previous nuclear magnetic resonance (NMR) and double electron–electron resonance (DEER) spectroscopy study

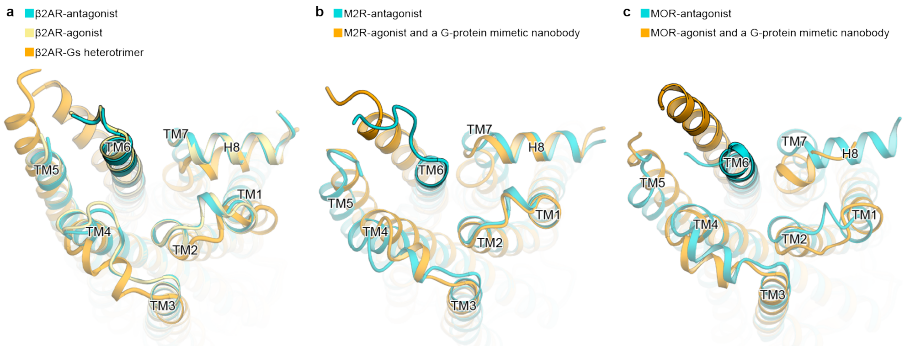


Figure 2-21 Comparison of the intracellular sides of class A GPCRs.

a, Structures of β2 adrenergic receptor in complex with an antagonist (PDB 2RH1), an agonist (PDB 3PDS), and G-protein heterotrimer (PDB 3SN6), coloured turquoise, khaki, and orange, respectively. **b**, Structures of M2 receptor bound to an antagonist (PDB 3UON), and in complex with an agonist and G-protein mimetic nanobody (PDB 4MQS), coloured turquoise and orange, respectively. **c**, Structures of μ-opioid receptor bound to an antagonist (PDB 4DKL), and in complex with an agonist and a G-protein mimetic

nanobody (PDB 5C1M), coloured turquoise and orange, respectively. Upon G-protein binding, the intracellular side of TM7 moves inward by 1 to 2 Å, and that of TM6 moves outward by 10 to 14 Å. These movements are a common structural feature among the class A GPCRs. The intracellular side of TM6 is not opened up in the agonist-bound structure of the β2 adrenergic receptor (PDB 3PDS). This outward movement of TM6 is highly dependent on the G-protein binding, and agonist binding only induces structural heterogeneity on the intracellular side of TM6.

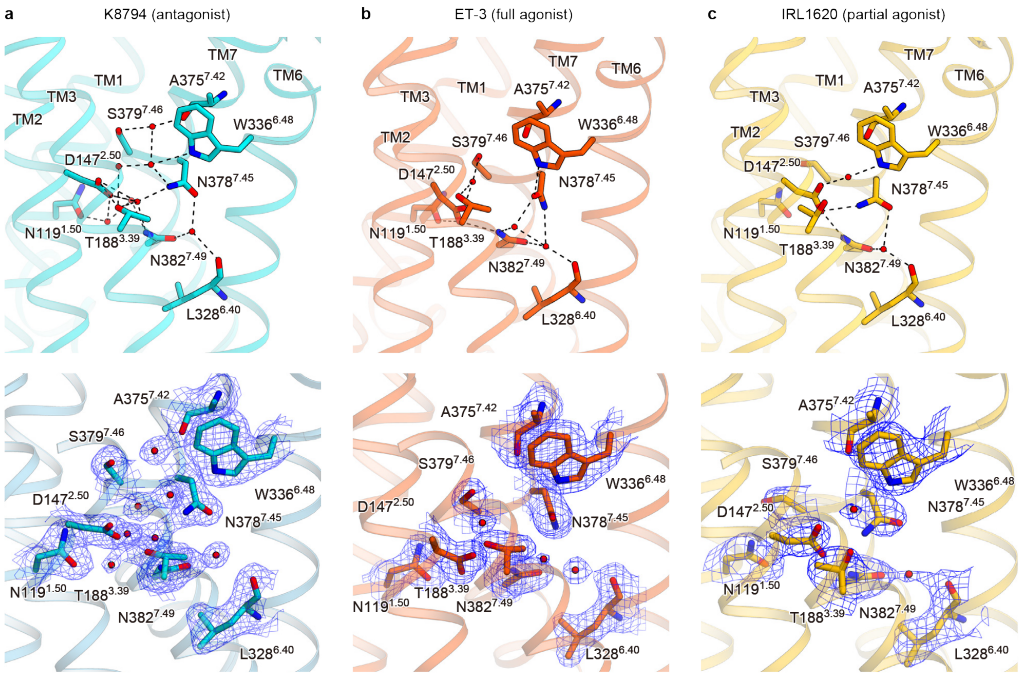


Figure 2-22 Water-mediated hydrogen-bonding network at the receptor core. Hydrogen-bonding networks in the intermembrane regions of the K8794- **(a)**, ET-3- **(b)**, and IRL1620- **(c)** bound structures, coloured turquoise, orange, and orange-red, respectively. The upper panels show the overall hydrogen-bonding interactions. Waters are shown as red spheres, and hydrogen bonding interactions are indicated by dashed lines. The lower panels show the $2F_o - F_c$ maps around $D147^{2.50}$ and $W336^{6.48}$, contoured at 1.0σ.

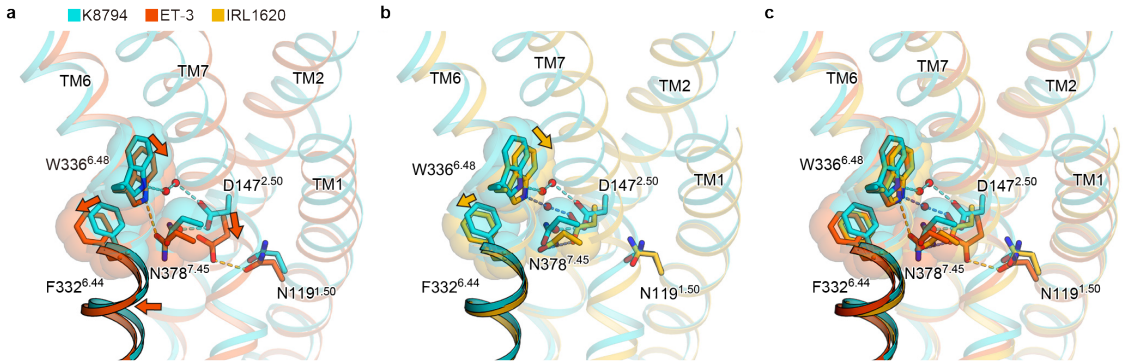


Figure 2-23 Comparison of the ET_B structures

Superimposition of the ET_B structures bound to K-8794 and ET-3 (a), K-8794 and IRL1620 (b), and K-8794, IRL1620, and ET-3 (c), focused on the intermembrane parts. The receptors are shown as ribbons, and the side chains of $N119^{1.50}$, $D147^{2.50}$, $F332^{6.44}$, $W336^{6.48}$, and $N378^{7.45}$ are shown as sticks with transparent surface representations. Waters are shown as red spheres. The dashed lines show hydrogen bonds coloured in the respective structures. The arrows show the structural changes on ligand binding.

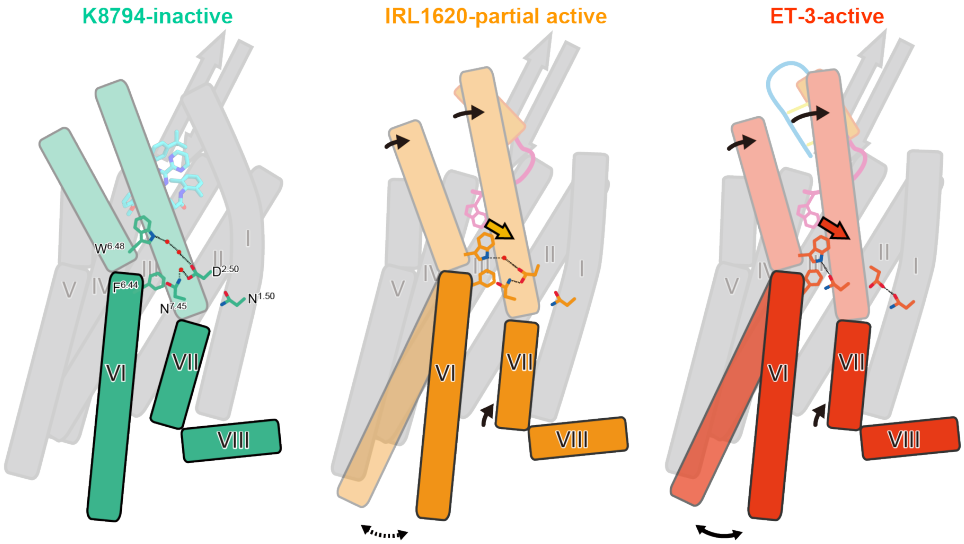


Figure 2-24 Activation and partial activation models of ET_B receptor

Receptor activation by ET-3 and partial activation by IRL1620. Schematic representations of receptor activation by ET-3 and partial activation by IRL1620. TM6, TM7, and H8 are highlighted. The residues involved in the signal transduction (N1.50, D2.50, F6.44, W6.48, and N7.45) are represented with sticks. Hydrogen bonds are indicated by black dashed lines. Arrows indicate the conformational changes in TM6 and TM7 upon ET-3 and IRL1620 binding. In the K-8794-bound structure (left), a water-

mediated hydrogen bonding network among D1472.50, W3366.48, and N3787.45 stabilizes the inactive conformation of the receptor. ET-3 binding disrupts this network and propagates the structural change of the transmission switch comprising F3326.44 and W3366.48 (middle), leading to the outward movement of TM6 upon G-protein coupling. In the IRL1620-bound structure (right), the water-mediated hydrogen-binding network is preserved, thus preventing the outward movement of TM6 upon G-protein coupling.

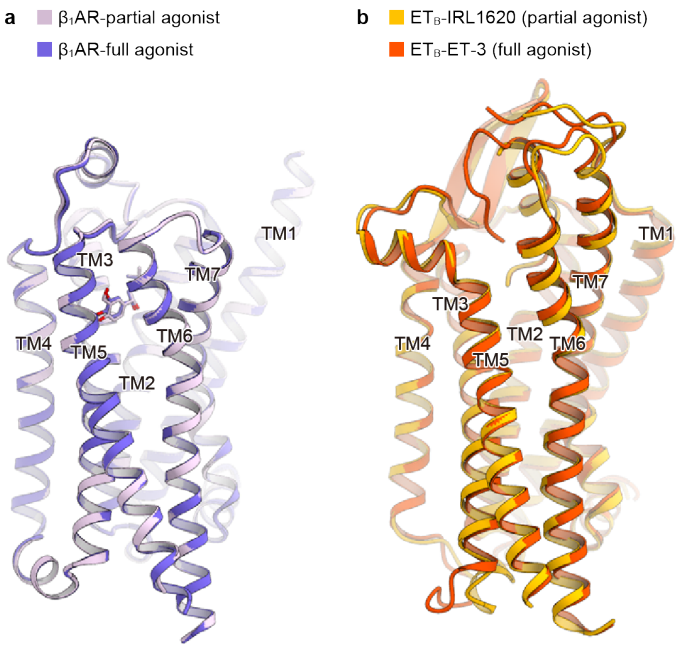


Figure 2-25 Comparison with β1-adrenergic receptor structures.

a, Superimposition of the β1-adrenergic receptor structures in complex with the full agonist isoprenaline (PDB 2Y03) and the partial agonist salbutamol (PDB 2Y04) (overall R.M.S.D of 0.25 Å for the Cα atoms). **b**, Superimposition of the ETB structures in complex with the full agonist ET-3 and the partial agonist IRL1620 (overall R.M.S.D of 0.91 Å for the Cα atoms).

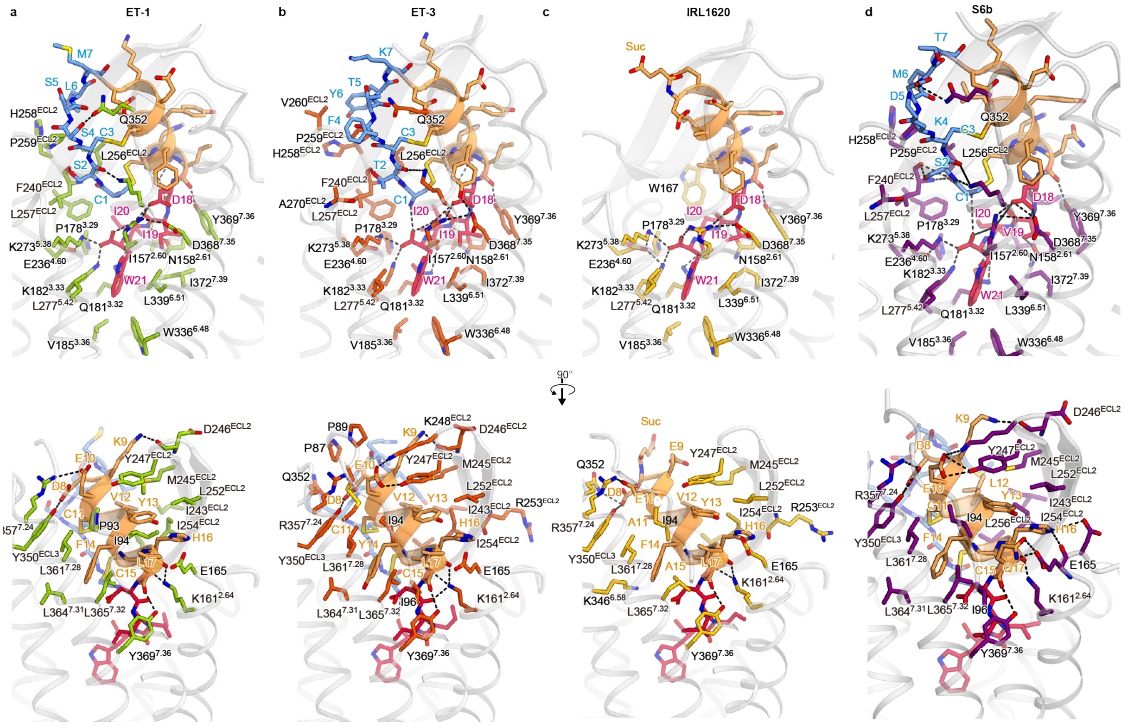


Figure 2-26 Comparison of receptor interactions of agonist peptides.

a-d, Comparison of the receptor binding interactions of ET-1 (a), ET-3 (b), IRL1620 (c), and sarafotoxin S6b. The upper panels show the interactions of the N-terminal and C-terminal regions, and the lower panels show those of the α-helical regions. The agonists are shown as ribbon and stick models. The structures of the receptors are shown as silver ribbons. The residues involved in the agonist binding are shown as sticks. I observed minor difference in the binding modes of ET-1 and ET-3. Three consecutive prolines (P88, P89, and P90) extend over the N-terminal region of ET-3 and form van der Waals interactions. In ET-1, E10 forms a salt bridge with $R357^{7.24}$, whereas in ET-3, E10 forms hydrogen-bonding interactions with $Y247^{ECL2}$ and $K248^{ECL2}$. F4 of ET-3 forms a van der Waals interaction with $Q352^{ECL3}$. The N-terminal amide and the C-terminal carboxylate of ET-3 form a direct hydrogen bond. I also observed differences in the binding modes of the full agonist ET-3 and the partial agonist IRL1620. Due to the truncation of the N-terminal region, IRL1620 completely lacks the interactions via this region. K9 is replaced with glutamic acid (E9) in IRL1620, and interactions at the N-terminal end of the α-helical region are slightly diverged.

K4, D5, and T7 form an intramolecular hydrogen-bonding network in S6b, and the $Q352^{ECL3}$ side chain participates in this network and form a hydrogen bond with D5. V12

in ET-1 and ET-3 is replaced with L12 in S6b, forming more extensive hydrophobic interactions with ECL2. L17 in ET-1 and ET-3 is replaced with the polar residue Q17 in S6b. Instead of the hydrophobic interaction, Q17 forms a hydrogen bond with $E165^{ECL1}$. I19 in ET-1 and ET-3 is replaced with V19, forming similar hydrophobic interactions. These differences are subtle, and thus the conserved residues among S6b, ET-1, and ET-3 form essentially similar interactions with the receptor.

Chapter 3: Structural analysis of human LysoPS receptor 1

3.1 Introduction (study purpose)

Lysophatidyl serine (LysoPS) is a type of lysophospholipid that induces inflammation as a lipid mediator. lysophosphatidyl serine (LysoPS) receptors include LPS1 (GPR34), LPS2 (P2Y10), and LPS3 (GPR174), all of which are class A GPCRs. LPS1 is the first LysoPS receptor to be discovered and the only Gi-coupled protein among the LysoPS receptors. However, the three-dimensional structure of the LysoPS receptor has not been reported, and the mechanism of ligand recognition and activation of LysoPS receptor is unknown. In this study, I have aimed to determine the three-dimensional structure of LPS1 in the Gi-bound state by cryo-EM analysis and to elucidate the binding mode of LysoPS and Gi.

3.2 Material and methods

3.2.1 Expression and purification

Receptor-GFP expression and purification

Wild type human GPR34 was modified to contain N-terminal FLAG tag (DYKDDDD). The GPR34 was subcloned into a modified pEG BacMam vector, with the resulting construct encording a TEV cleavage site followed by a GFP-His_8 tag at the C-terminus. The recombinant baculovirus was prepared using the Bac-to-Bac baculovirus expression system (Invitrogen). HEK293S GnT1$^-$ (N-acetylglucosaminyl-transferase I–negative) cells (American Type Culture Collection, catalog No. CRL-3022) were infected with the virus at a cell density of 3.0-4.0 $\times 10^6$ cells ml^{-1} in Freestyle 293 Expression Medium (Gibco, catalog No. 12338026), and grown for 37℃ for 12 h and supplemented with 10 mM sodium butyrate to boost protein expression. The cells were then grown for 30℃ for 48-60 h, collected by centrifugation (4000g, 10min, 4℃). The cells were disrupted by sonication in buffer, containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl and 10% Glycerol. The membrane fraction was collected by ultracentrifugation, and solubilized in buffer, containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2% LMNG, 0.2% CHS, 10% Glycerol and 10 μM LysoPS analog for 1 h at 4℃. The supernatant was separated from the insoluble material by ultracentrifugation, added 3 mM $CaCl_2$ and incubated with ANTI-FLAG M1 Agarose Affinity Gel (SiGMa, catalog No. A4596) for 1.5-3 h. The Gel was washed with 15 column volumes of buffer, containing 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 10 % Glycerol, 0.05% GDN, 1 μM LysoPS analog, 3 mM

$CaCl_2$ and 100 mM TCEP. The receptor was incubated with buffer, containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 10 % Glycerol, 0.03% GDN, 1 μM LysoPS analog, 5 mM EGTA, 0.1 mg ml^{-1} Flag peptide and 100mM TCEP for 30 min. the eluate was concentrated and loaded onto a Superdex200 10/300 Increase size-exclusion column, equilibrated in buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 10 % Glycerol, 0.03% GDN, 1 μM LysoPS analog and 100 mM TCEP. The peak fractions ware added final 10 μM LysoPS analog and frozen until GPR34-Gi complex formation.

Heterotrimeric G-proteins expression and purification

Wild type human $G_{\alpha i1}$ was subcloned into a modified pFastBac vector, with the resulting construct encording a TEV cleavage site followed by a GFP-His_8 tag at the C-terminus. Wild type $G_{\beta\gamma}$ was subcloned into a modified pFastBac vector. Each recombinant baculovirus was prepared using the Bac-to-Bac baculovirus expression system (Invitrogen). Sf9 insect cells were infected with the virus at a cell density of 3.0-4.0 × 10^6 cells per millilitre in Sf900 II SFM (ThermoFisher, catalog No. 11496-015), and grown for 48 h at 27 °C. The co-expressed cells are solubilized in buffer, containing 20 mM Tris-HCl (pH8.0), 150 mM NaCl, 1% DDM, 10% Glycerol, 50 μM GDP, 10 mM imidazole for 1h at 4℃. The supernatant was separated from the insoluble material by ultracentrifugation and incubated with Ni-NTA resin and rotated for 30 min. The complex was washed with 12 column volumes of buffer, containing 20 mM Tris-HCl (pH8.0), 500 mM NaCl, 0.02% DDM, 10% Glycerol, 10 μM GDP, 30 mM imidazole. The complex was eluted in buffer, containing 20 mM Tris-HCl (pH8.0), 150 mM NaCl, 0.02% DDM, 10% Glycerol, 10 μM GDP, 300 mM imidazole. The eluate was treated with TEV protease and dialysed against buffer, containing 20 mM Tris-HCl (pH8.0), 50 mM NaCl, 0.02% DDM, 10% Glycerol, 10 μM GDP, 2 μM β-ME. The cleaved GFP–His_8 tag and the TEV protease were removed with the Ni-NTA resin. The complex was concentrated and loaded onto a HitrapQ ion-exchange chromatography column equilibrated in buffer A (20 mM Tris-HCl (pH8.0), 50 mM NaCl, 0.02% DDM, 10% Glycerol, 1 μM GDP, 2 μM β-ME) and buffer B (20 mM Tris-HCl (pH8.0), 1 M NaCl, 0.02% DDM, 10% Glycerol, 1 μM GDP, 2 μM β-ME). The complex peak fractions were pooled, concentrated, and frozen until GPR34-Gi complex formation.

Complex formation of GPR34-Gi-scFv16

The purified receptor, Gi complex, and scFv16 (purified by Iwata laboratry) were

mixed together at a weight ratio of 1:3:1.5 (receptor:Gi:scFv16), added 1% GDN, 100 mU apyrase, 10 μM ligand, and incubated for 1h at 4℃. The protein was added TEV protease and incubate O/N. The protein was added ANTI-FLAG M1 Agarose Affinity Gel (SiGMa, catalog No. A4596) and 3 mM $CaCl_2$ and incubated for 3 h. The Gel was washed with 10 column volumes of buffer, containing 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 10 % Glycerol, 0.05% GDN, 1 μM LysoPS analog, 3 mM $CaCl_2$ and 100 μM TCEP. The receptor was incubated with buffer, containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 10 % Glycerol, 0.03% GDN, 1 μM LysoPS analog, 5 mM EGTA, 0.1 mg ml^{-1} Flag peptide and 100mM TCEP for 30 min. The eluate was concentrated and loaded onto a Superdex200 10/300 Increase size-exclusion column, equilibrated in buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 10 % Glycerol, 0.03% GDN, 1 μM LysoPS analog A and 100 μM TCEP. The peak fractions were pooled and concentrated to 9.4 $mgml^{-1}$ for making the cryo grid.

3.2.2 Grid preparation and cryo-EM image acquisition of LPS1-Gi-scFv16 complex

The purified complex was applied onto a freshly glow-discharged Quantifoil 300 mesh gold grids (R1.2/1.3), blotted for 10 s at 4 °C in 100% humidity, and plunge-frozen in liquid ethane by using a Vitrobot Mark IV. Data collections were performed on an FEI Titan Krios (FEI) EM, operated at an acceleration voltage of 300 keV, and equipped with a BioQuantum K3 imaging filter and a K3 direct electron detector (Gatan). EM images were acquired at a nominal magnification of ×105,000, corresponding to a physical pixel size of 0.83 Å $pixel^{-1}$, using the SerialEM software. Movies were dose-fractionated to 48 frames at a dose rate of 13.5 e^- $pixel^{-1}$ per second, resulting in a total accumulated dose of 50.8 e^- $Å^{-2}$.

3.2.3 Data processing

The movie frames were aligned in 5 × 5 patches and dose weighted in MotionCor2. Defocus parameters were estimated by CTFFIND 4.1. First, template-based auto-picking was performed with the two-dimensional class averages of a few hundred manually picked particles as templates. A total of 301,952 particles were extracted from 2,358 micrographs in 3.24219 Å $pixel^{-1}$. The initial model was generated in RELION-3.1. The particles were subjected to several rounds of 2D and 3D classifications, resulting in the best class from the 3D classification, which contained 179,882 particles. The particles

were then re-extracted with a pixel size of 1.44097 Å pixel^{-1}, and then subjected to 3D refinement. The resulting 3D mode and particle set was subjected to per-particle defocus refinement, Bayesian polishing, and 3D refinement. The final 3D refinement and postprocessing yielded a map with global resolutions of 3.57759 Å, with the gold standard Fourier Shell Correlation (FSC = 0.143) criteria.

3.2.4 Model building and refinement

Homology model of GPR34 was built by HHpred using Muscarinic acetylcholine receptor (M1R, PDBID: 6OIJ). Also, Gi and scFv16 model was builted using M1R-Gi-scFv structure. The model was docked into the EM density map using Chimera and cuemol followed by iterative manual adjustment in COOT and phenix.real_space_refine in Phenix. All structural figures were performed by cuemol.

3.3 Results and discussion

3.3.1 Co-expression check of GPR34-Gi-scFv16

First, I examined the Gi binding ability among LPS1 homologs. The figure shows the peak positions of the three homologs with and without Gi, which should be shifted to the high molecular weight side when complexed with Gi. I then examined the complex formation ability of 1-ethoxy-3-oleoyl-LysoPS, which had the best properties among the LysoPS homologs. After expressing human GPR34 on a small scale, I purified it by adding ligand and Gi. SDS-PAGE results at each step of purification indicated that the ligand is an agonist that forms the GPR34-Gi complex (Figure 3-1).

3.3.2 Overall structure of the complex

The EM-map shows the density of the GPR34-Gi-scFv16 complex (Figure 3-3). The final map is excluded detergent micelle density by RELION-3.1. The atomic model of GPR34-Gi-scFv16 is constructed. In this paper, I would like to focus on the receptor region.

The overall structure of GPR34 consists of the canonical 7 TMs, the amphipathic helix 8 (H8) at the C-terminus (Figure 3-4). The structure of Helix 8 is not found in P2Y1, but it is observed in P2Y12.

N-terminal 50 residues and several loops are flexible and hard to determine the structure. ECL2 has a short alpha-helix in a hairpin-like structure, and this structure is

also conserved in that of P2Y12 (Figure 3-5). Lipid-receiving GPCRs often have ECL2 or N-terminus or both forming the lid-like structure. In GPR34, only ECL2 forms the lid-like structure. In the EM-structure of EP4, the ligand PGE_2 does not interact electrostatically with ECL2, but is enclosed by ECL2. The closed extracellular domain may be important for lipid receptors (Figure 3-6).

Although GPR34-Gi forms a complex, the density of its ligand could not be determined. The inhibitor-binding site of P2Y12 is inside the lid of ECL2, and although the peripheral residues that recognize the inhibitor in P2Y12 and GPR34 are poorly conserved, there is a small density at a similar position in GPR34. The surrounding residues are predominantly hydrophilic, suggesting that they could recognize the polar group of LysoPS. In addition, the charge calculation of GPR34 using APBS shows that the ligand binding site is positively charged, which could bind to the negative charge of the polar group of LysoPS. Docking simulation and mutant analysis will reveal whether this region is the ligand binding pocket.

The charge calculation of GPR34 using APBS shows a gap with neutral charge at TM4-5 (Figure 3-7). This gap extends to the region of the presumed pocket of polar groups in LysoPS. This gap is not found in EP4, which has a similar ECL2 structure, but is found in several lipid receptors, including LPA6, which has a similar structure to the P2Y family. This gap may help lateral access where the acyl group slips into or the binding pocket of the acyl group.

Figures and Tables of Chapter 3

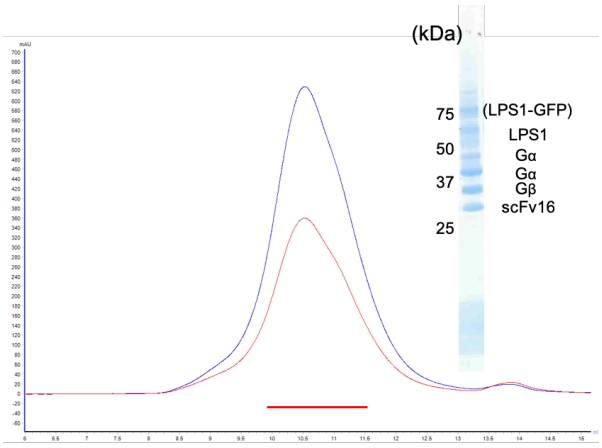


Figure 3-1 Gel filtration of GPR34-Gi-scFv16 complex

Size-exclusion chromatogram and SDS-PAGE of the GPR34-Gi-scFv16. Blue and red lines indicate absorbances at 280 nm and 260 nm, respectively.

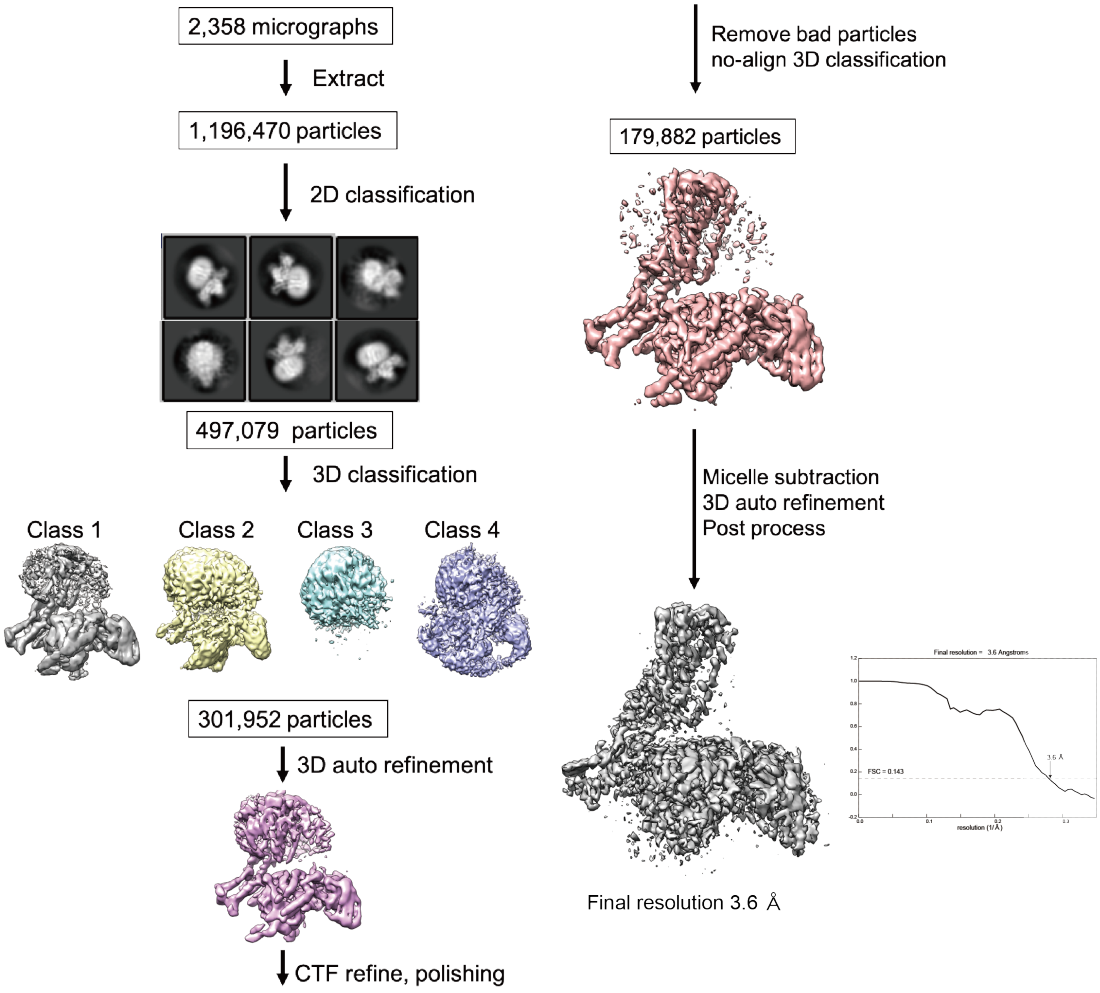


Figure 3-2 Structure determination of the GPR34-G_i-scFv16

Flow chart of cryo-EM data processing. Reprehensive 2D averages shows structural features of each GPR34-G_i-scFv16 complex domain. Gold standard FSC curves of Final post processed map indicates overall nominal resolutions of 3.6 angstroms using FSC = 0.143 criteria.

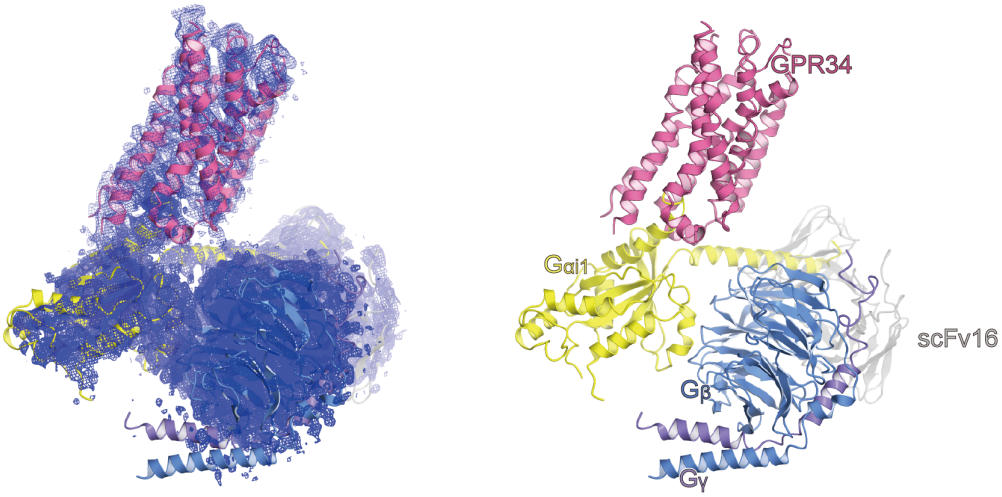


Figure 3-3 Structure of the GPR34-Gi-scFv complex in the cryo-EM map.
The model is shown as a ribbon representation (left), and with the transparent map (right). The GPR34 (hot pink) is coupled with $G_{\alpha i1}$ (yellow), G_{β} (blue), and G_{γ} (purple). The nanobody (Nb35, Dark gray).

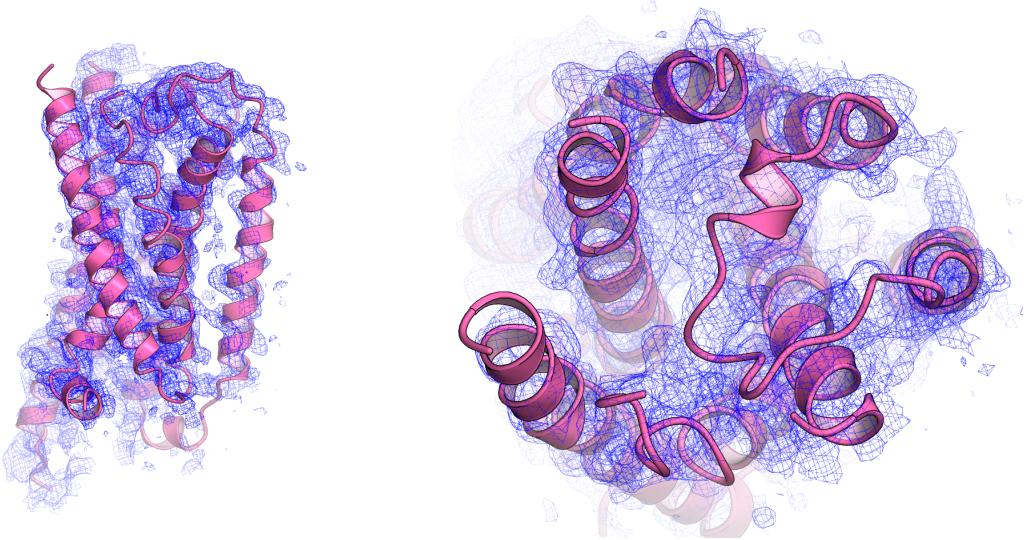


Figure 3-4 Overall structure of GPR34
Cryo-EM map and receptor atomic model of GPR34 in active state. The receptor (hot pink) is shown in membrane (left) and extracellular side (right).

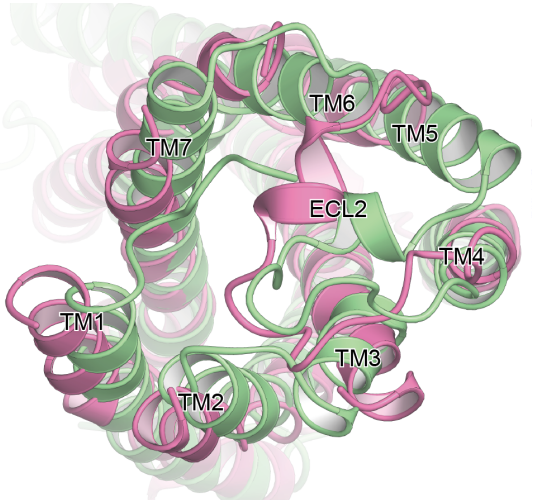


Figure 3-5 Comparison of ECL2 of GPR34 and P2Y12 structure

Superposed structures of GPR34 (hot pink) and P2Y12 (green). ECL2 topology of GPR34 is similar to that of P2Y12, despite of their sequence homology (32%).

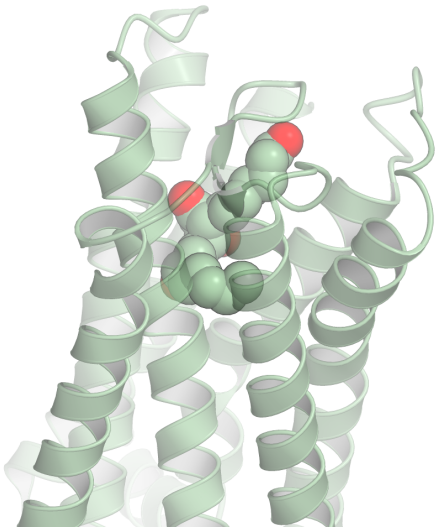


Figure 3-6 Ligand binding region of EP4

PGE_2-EP4-Gs-Nb35 complex is shown in extracellular side. Receptor (ribbon) and its ligand PGE2 (CPK) is coloured in dark green. PGE2 is covered by ECL2 and gap-like structure is seen in TMs 4 to 5.

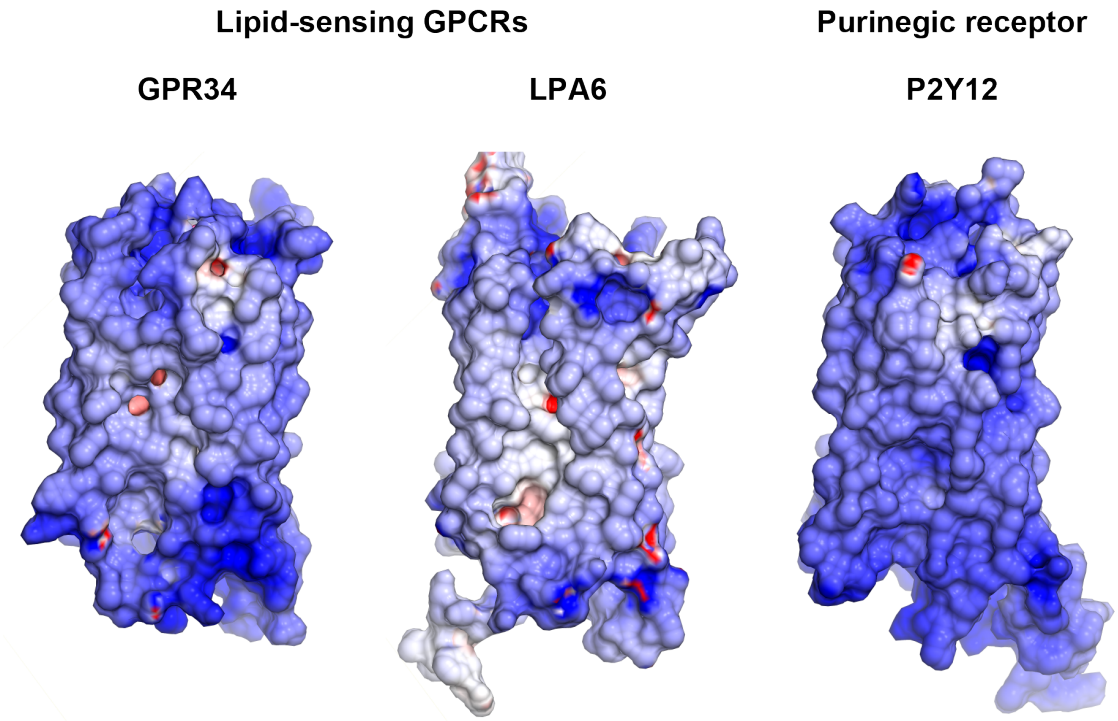


Figure 3-7 Electrostatic surfaces of GPR34, LPA6 and P2Y12.

Both surfaces are shown in TMs 4 to 5 side. GPR34 and LPA6 have a neutoral charge moiety in TMs 4 to 5 gap, while P2Y12, which receives nucleotide, has no neutoral charge moiety around TMs 4 and 5.

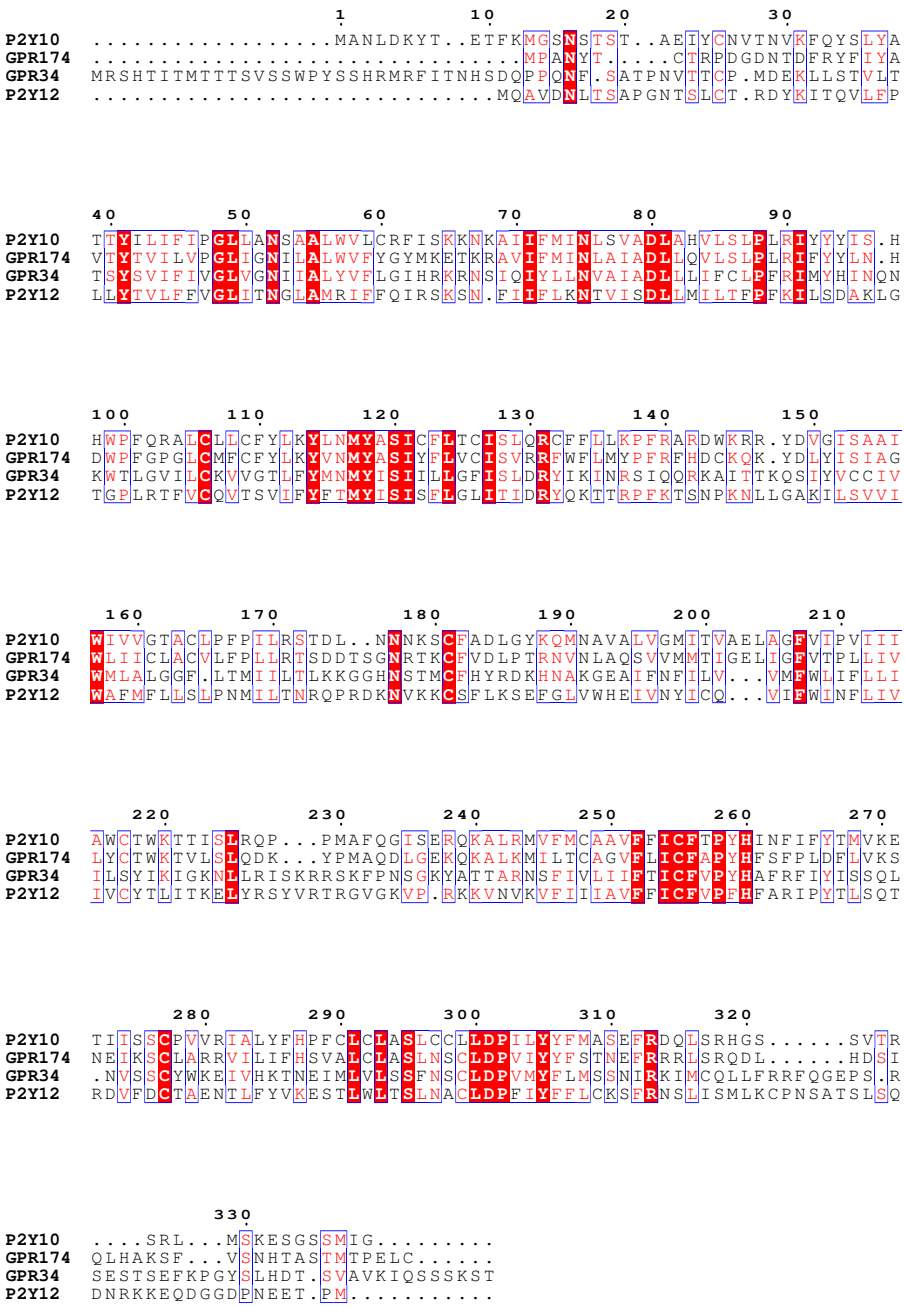


Figure 3-8 Sequence alignment of LysoPS receptor and its homolog P2Y12.

Chapter 4 General discussion

4.1 Summary of the thesis

4.1.1 Endothelin receptor type B

Previous studies have suggested that the α-helical region and C-terminal regions of endothelins are critical elements for receptor activation[545556]. whereas the N-terminal region is only responsible for the ETR selectivity[57]. Indeed, the N-terminal region-truncated analogue IRL1620 has similar EC_{50} values, as compared with most potent endothelin ET-1[18]. Our pharmacological experiments for the first time proved that IRL1620 functions as a partial agonist for the ET_B receptor, rather than a full agonist such as ET-1 and ET-3, suggesting the participation of the N-terminal region in the activation process of the ET_B receptor. To clarify the receptor activation mechanism, I determined the crystal structures of the human ET_B receptor in complex with ET-3 and IRL1620. The 2.0 Å high-resolution structure of the ET-3-bound ET_B receptor revealed that the large inward motions of the extracellular portions of TMs 2, 6, and 7 disrupt the water-mediated hydrogen bonding network at the receptor core (Figure 2-23). The IRL1620-bound ET_B structure revealed that the IRL1620-induced inward motions of extracellular TMs 6 and 7 are smaller by about 1 Å, as compared with those caused by ET-3 (Figure 2-20). Despite the lower resolution of the IRL1620-bound structure, the $2F_o - F_c$ map in the receptor core shows different rearrangement of water molecules and amino acid residues, in which the hydrogen-bonding network is partially preserved (Figure 2-22). This preserved network, together with the smaller inward motion of TM6, may prevent outward motion of cytoplasmic TM6 that occurs upon G-protein binding. These observations suggest that the interactions between the N-terminal regions of endothelins and TM6 is also effective in receptor activation, while the extensive interactions of the α-helical and C-terminal regions with the receptor primarily contribute to this process (Figure 2-26). This activation mechanism is different from that of the small-molecule activated GPCRs (e.g., β2 adrenaline and M2 muscarinic acetylcholine receptors), in which only a small number of hydrogen-bonding interactions between the agonist and the receptor induce receptor activation, by disrupting the receptor dynamics[5859].

$D^{2.50}$ is one of the most conserved residues among the class A GPCRs (90%). Recent high-resolution structures have revealed that $D^{2.50}$ is coordinated with a sodium ion and forms a water-mediated hydrogen bonding network in the intermembrane region, which stabilizes the inactive conformation of the receptor[60], and the networks collapse leads to

receptor activation. Previous 2.2 Å resolution structure of the K-8794-bound ET_B receptor revealed that a water molecule occupies this allosteric sodium site and participates in the extensive hydrogen-bonding network, instead of a sodium ion[26] (Supplementary Figure 7a), and this hydrogen-bonding network is collapsed in the 2.8 Å resolution structure of the ET-1-bound ET_B receptor, indicating its involvement in the receptor activation[25]. Nevertheless, the precise rearrangement of this water-mediated hydrogen bonding network remained to be elucidated, due to the limited resolution. The 2.0 Å resolution structure of the ET-3-bound ET_B receptor revealed that the collapse and rearrangement of the water-mediated interaction between $W336^{6.48}$ and $D147^{2.50}$ is critical for receptor activation (Fig. 5d). This network is partly preserved in the IRL1620-bound structure (Fig. 5e), thus preventing the transition to the fully active conformation upon G-protein coupling (Fig. 6). $W^{6.48}$ is also highly conserved among the class A GPCRs (71%)[61], and the association between $W^{6.48}$ and $D^{2.50}$ has a critical role in the class A GPCR activation process, as shown in the previous nuclear magnetic resonance (NMR) study of the adenosine A2A receptor[62]. Given the importance of $W^{3.36}$ and $D^{2.50}$ in the activation of GPCRs, our proposed model of the partial receptor activation by IRL1620-ET_B structure is consistent with the previous functional analyses of GPCRs. To date, the β1 adrenergic receptor is the only receptor for which agonist- and partial agonist-bound structures were reported[63]. However, these structures are both stabilized in inactive conformations by the thermostabilizing mutations and this may cause only slight differences (Figure 2-25). Therefore, our study provides the first structural insights into the partial activation of class A GPCRs.

Our current study further suggests possible improvements in clinical studies using ET_B-selective agonists. IRL1620 is the smallest among the ET_B-selective agonists, and thus is expected to be useful for the treatment of cancers and other diseases. While its effectiveness has been proven in rat experiments, a recent phase 2 study has failed[24], and thus further improvement of IRL1620 is required for clinical applications. Our cell-based assays and structural analysis revealed the partial agonistic effect of IRL1620 on the ET_B receptor in the G-protein coupling and β-arrestin recruitment activities, suggesting the possible tuning of its efficacy. The development of ET_B-selective agonists by fine-tuning their G-protein activation and/or β-arrestin recruitment activities might be beneficial for clinical applications.

4.1.2 GPR34

GPR34, the LysoPS receptor 1, has low sequence homology with its homologs, P2Y10 (27%) and GPR174 (29%). Also, there is low sequence conservation between P2Y10 and GPR174 (58%) indicated that it is more likely to assume that the LysoPS receptors have low sequence conservation rather than that only GPR34 has a unique amino acid sequence. It is common among lipid receptors that low sequence conservation among homologs that accept the same ligand. For example, EP4 has 37% sequence homology with the most similar receptor PGI receptor, and LPA6 has 59% sequence homology with the most similar one LPA4. In addition, LysoPS receptors belong to the P2Y family, but the ligand of the major P2Y family is ADP, not lipids. From these facts, it seems that structural features are essential for how lipid receptors accept lipids, rather than recognition by unique and highly conserved residues.

For example, LPA6 and S1P1, which accept lysophospholipids, are known to have gaps between TM4-5 and TM7-1, respectively. FFAR1, which accepts free fatty acids, also has a groove between TM3 and TM4. Although LPA1 does not have a gap structure, some receptor that receive lysophospholipids and compounds with acyl groups as ligands may require a gap structure during ligand entry or ligand binding. It is still not understood why GPCRs with low sequence homology, i.e., evolutionarily distantly related, sometimes accept the same lipids. This study may raise the hypothesis that the lipid receptor arises as a result of mutations that causes a gap structure somewhere between the TMs that connects to the original ligand recognition site. In this case, there are unfavorable points if the original ligand is allowed to access from the outside of the cell instead of lateral access. To prevent this, many lipid receptors may have a lid structure. While the lid structure of ETBR described in Chapter 2 allows irreversible binding of peptide hormones, the lid structure of GPR34 in Chapter 3 may be used to deny access from outside the cell.

Acknowledgements

First of all, I would like to express my gratitude to Prof. Osamu Nureki, the professor of my lab who helped me in my master and Ph.D. course. If it were not for him, I would not have been interested in structural biology and joined the Nureki lab. Even after I joined the laboratory, he provided me with a place to do research and ample opportunities to do so.

Next, I would like to thank Prof. Tomohiro Nishizawa and Dr. Wataru Shihoya for supervising my research. Dr. Tomohiro Nishizawa for his help in discussing the direction of my research theme. Assistant Professor Wataru Shihoya taught me almost all the research techniques used in the structural analysis and helped me in discussing the development of the paper. I would also like to thank Dr. Kunio Hirata and Dr. Keitaro Yamashita for conducting X-ray diffraction experiments at SPring-8 BL32XU. Without their ZOO system, it would not have been possible to collect more than 200 data sets of ET-3-ETBR crystals. In particular, I would like to thank Dr. Keitaro Yamashita for his help in analyzing the diffraction data and refining the protein model structures.

In addition, I would like to thank Prof. Junken Aoki, Prof. Aska Inoue, and Dr. Kadji at Tohoku University for their help in the experiments to evaluate the G-protein activation of GPCRs. I thank the OOwada laboratory for providing the LysoPS analog, an agonist of GPR34, and Iwata laboratory for the preparation of antibodies to stabilize Gi proteins. I also thank Dr. Hirotake Miauchi for his MD simulation of the binding mode between ETR and Sarafotoxin.

I thank Assistant Professor Tsukasa Kusakizako, Mr. Kazuhiro Kobayashi, and Mr. Hiroyuki Okamoto for their advice on the structural analysis using a cryo-EM software RELION-3.1. Mr. Kazuhiro Kobayashi and Mr. Hiroyuki Okamoto gave me advice based on their deep experience in the structural analysis of GPCRs. I would like to express my sincere gratitude to them.

I would like to express many thanks to Mr. Kazumasa Oda and Mr. Tatsuya Ikuta, with whom I spent five years of my life as a labmate.

Finally, I would like to express my deepest gratitude to my family and friends who have supported me throughout my life.

Tamaki Izume

References

1. Wu, F., Song, G., de Graaf, C. & Stevens, R. C. Structure and Function of Peptide-Binding G Protein-Coupled Receptors. *J. Mol. Biol.* **429**, 2726–2745 (2017).
2. Kamato, D. *et al.* Structure, Function, Pharmacology, and Therapeutic Potential of the G Protein, Gα/q,11. *Front. Cardiovasc. Med.* **2**, 1–11 (2015).
3. Suzuki, N., Hajicek, N. & Kozasa, T. Regulation and physiological functions of G12/13-mediated signaling pathways. *NeuroSignals* **17**, 55–70 (2009).
4. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature* **469**, 236–242 (2011).
5. Kolakowski, L. F. GCRDb: A G-protein-coupled receptor database. *Recept. Channels* **2**, 1–7 (1994).
6. Dixon, R. A. F. *et al.* Cloning of the gene and cDNA for mammalian β-adrenergic receptor and homology with rhodopsin. **321**, 75–79 (1986).
7. Unger, V. M., Hargrave, P. A., Baldwin, J. M. & Schertler, G. F. X. Arrangement of rhodopsin transmembrane α-helices. *Nature* **389**, 203–206 (1997).
8. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science (80-.).* **289**, 739–745 (2000).
9. Juan A., B. . H. W. . Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
10. Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *Elife* **8**, 1–31 (2019).
11. White, K. L. *et al.* Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling. *Structure* **26**, 259-269.e5 (2018).
12. Shihoya, W. *et al.* Activation mechanism of endothelin ET B receptor by endothelin-1. *Nature* **537**, 363–368 (2016).
13. Yanagisawa, M. *et al.* A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature* **332**, 411–415 (1988).
14. Arai, H., Hori, S., Aramori, I., Ohkubo, H. & Nakanishi, S. Cloning and expression of a cDNA encoding an endothelin receptor. *Nature* **348**, 730–732 (1990).
15. Sakurai, T. *et al.* Cloning of a cDNA encoding a non-isopeptide-selective

subtype of the endothelin receptor. *Nature* **348**, 732–735 (1990).

16. Remuzzi, G., Perico, N. & Benigni, A. New therapeutics that antagonize endothelin: Promises and frustrations. *Nat. Rev. Drug Discov.* **1**, 986–1001 (2002).
17. Hiyama, T. Y. *et al.* Endothelin-3 expression in the subfornical organ enhances the sensitivity of Nax, the brain sodium-level sensor, to suppress salt intake. *Cell Metab.* **17**, 507–519 (2013).
18. Takai, M. *et al.* A potent and specific agonist, Suc-[Glu9,Ala11,15]-endothelin-1(8-21), IRL 1620, for the ETB receptor. *Biochem. Biophys. Res. Commun.* **184**, 953–959 (1992).
19. Gulati, N. V. R., Matwyshyn, G. & Gulanti, A. IRL-1620, a tumor selective vasodilator, augments the uptake and efficacy of chemotherapeutic agents in prostate tumor rats. *Prostate* **67**, 701–713 (2007).
20. Rajeshkumar, N. V., Rai, A. & Gulati, A. Endothelin B receptor agonist, IRL 1620, enhances the anti-tumor efficacy of paclitaxel in breast tumor rats. *Breast Cancer Res. Treat.* **94**, 237–247 (2005).
21. Leonard, M. G., Briyal, S. & Gulati, A. Endothelin B receptor agonist, IRL-1620, provides long-term neuroprotection in cerebral ischemia in rats. *Brain Res.* **1464**, 14–23 (2012).
22. Briyal, S., Shepard, C. & Gulati, A. Endothelin receptor type B agonist, IRL-1620, prevents beta amyloid (Aβ) induced oxidative stress and cognitive impairment in normal and diabetic rats. *Pharmacol. Biochem. Behav.* **120**, 65–72 (2014).
23. Briyal, S., Nguyen, C., Leonard, M. & Gulati, A. Stimulation of endothelin B receptors by IRL-1620 decreases the progression of Alzheimer's disease. *Neuroscience* **301**, 1–11 (2015).
24. Kim, R. *et al.* Phase 2 study of combination SPI-1620 with docetaxel as second-line advanced biliary tract cancer treatment. *Br. J. Cancer* **117**, 189–194 (2017).
25. Shihoya, W. *et al.* Activation mechanism of endothelin ET B receptor by endothelin-1. *Nature* **537**, 363–368 (2016).
26. Shihoya, W. *et al.* X-ray structures of endothelin ETB receptor bound to clinical antagonist bosentan and its analog. *Nat. Struct. Mol. Biol.* **24**, 758–764 (2017).
27. Kloog, Y. & Sokolovsky, M. Similarities in mode and sites of action of

sarafotoxins and endothelins. *Trends Pharmacol. Sci.* **10**, 212–214 (1989).

28. Bdolah, A., Wollberg, Z., Fleminger, G. & Kochva, E. SRTX-d, a new native peptide of the endothelin/sarafotoxin family. *FEBS Lett.* **256**, 1–3 (1989).
29. Kochva, E., Bdolah, A. & Wollberg, Z. Sarafotoxins and endothelins: evolution, structure and function. *Toxicon* **31**, 541–568 (1993).
30. Lauer-Fields, J. L. *et al.* Engineered sarafotoxins as tissue inhibitor of metalloproteinases-like matrix metalloproteinase inhibitors. *J. Biol. Chem.* **282**, 26948–26955 (2007).
31. Serhan, C. N., Chiang, N. & Van Dyke, T. E. Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nat. Rev. Immunol.* **8**, 349–361 (2008).
32. Fukushima, N., Ishii, I., Contos, J. J. A., Weiner, J. A. & Chun, J. Lysophospholipid Receptors. *Annu. Rev. Pharmacol. Toxicol.* **41**, 34 (2001).
33. Nojima, S. *et al.* Cryo-EM Structure of the Prostaglandin E Receptor EP4 Coupled to G Protein. *Structure* 1–9 (2020). doi:10.1016/j.str.2020.11.007
34. Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**, 706–731 (2009).
35. Taniguchi, R. *et al.* Structural insights into ligand recognition by the lysophosphatidic acid receptor LPA 6. *Nature* **548**, 356–360 (2017).
36. Yamashita, K., Hirata, K. & Yamamoto, M. KAMO: towards automated data processing for microcrystals. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 441–449 (2018).
37. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr* **66**, 125–132 (2010).
38. Foadi, J. *et al.* Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1617–1632 (2013).
39. Ueno, G. et al. Remote access and automation of SPring-8 MX beamlines. *AIP Conf. Proc.* **1741**, 050021 (2016).
40. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
42. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for

macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, (2010).

43. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
44. Inoue, A. *et al.* TGFα shedding assay: An accurate and versatile method for detecting GPCR activation. *Nat. Methods* **9**, 1021–1029 (2012).
45. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **11**, 400–408 (2016).
46. Vrecl, M., Jorgensen, R., Pogačnik, A. & Heding, A. Development of a BRET2 screening assay using β-arrestin 2 mutants. *J. Biomol. Screen.* **9**, 322–333 (2004).
47. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779–815 (1993).
48. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
49. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
50. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
51. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
52. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
53. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
54. Saeki, T., Ihara, M., Fukuroda, T., Yamagiwa, M. & Yano, M. [Ala1,3,11,15]endothelin-1 analogs with ETB agonistic activity. *Biochem. Biophys. Res. Commun.* **179**, 286–292 (1991).
55. Saeki, T., Ihara, M., Fukuroda, T. & Yano, M. Structure-activity relationship for ETB agonism in truncated endothelin-1 analogs. *Biochem. Int.* **28**, 305—312 (1992).

56. Heyl, D. L. *et al.* Truncated analogues of endothelin and sarafotoxin are selective for the ETB receptor subtype. *Pept. Res.* **6**, 238—241 (1993).
57. Lättig, J., Oksche, A., Beyermann, M., Rosenthal, W. & Krause, G. Structural determinants for selective recognition of peptide ligands for endothelin receptor subtypes ETA and ETB. *J. Pept. Sci.* **15**, 479–491 (2009).
58. Manglik, A. *et al.* Structural Insights into the Dynamic Process of β2-Adrenergic Receptor Signaling. *Cell* **161**, 1101–1111 (2015).
59. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
60. Fenalti, G. *et al.* Molecular control of δ-opioid receptor signalling. *Nature* **506**, 191–196 (2014).
61. Holst, B. *et al.* A conserved aromatic lock for the tryptophan rotameric switch in TM-VI of seven-transmembrane receptors. *J. Biol. Chem.* **285**, 3973–3985 (2010).
62. Eddy, M. T. *et al.* Allosteric Coupling of Drug Binding and Intracellular Signaling in the A2A Adenosine Receptor. *Cell* **172**, 68-80.e12 (2018).
63. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **469**, 241–244 (2011).

Original papers related to this thesis

(1) Wataru Shihoya*, Tamaki Izume*, Asuka Inoue, Keitaro Yamashita, Francois Marie Ngako Kadji, Kunio Hirata, Junken Aoki, Tomohiro Nishizawa and Osamu Nureki

Crystal structures of human ET_B receptor provide mechanistic insight into receptor activation and partial activation

Nature communications **9** 4711 (9 November 2018)

(2) Tamaki Izume, Hirotake Miyauchi, Wataru Shihoya, Osamu Nureki

Crystal structure of human endothelin ET_B receptor in complex with sarafotoxin S6b

Biochemical and Biophysical Research Communications 23 July 2020 **528**-2 383-388