THE UNIVERSITY OF CHICAGO

THE EXTRACELLULAR REGIONS OF ADHESION G PROTEIN-COUPLED RECEPTORS REGULATE LIGAND BINDING AND INTRACELLULAR SIGNALING

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

BIOPHYSICAL SCIENCES

BY

GABRIEL SIMON SALZMAN

CHICAGO, ILLINOIS

AUGUST 2017

DEDICATION

This work is dedicated to my ever-loving and supportive parents, Robin Flam and Neil Salzman. Your intelligence, dedication, and compassion will always be my inspiration, both in academia and the real world.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I extend my greatest thanks to Drs. Demet Araç and Shohei Koide for taking me on as a dual-mentored student. Together, their support and mentorship have been invaluable to my training. The extra responsibilities undertaken by each of them to accommodate me have always been noticed and deeply appreciated. I am proud to have been a part of both the Araç and Koide labs, and I will always think of my training with sincere fondness.

I thank my thesis committee chair, Dr. Tobin Sosnick, for his unwaveringly biophysical perspective and experienced mentorship. Dr. Sosnick wrote two letters supporting my NIH F30 fellowship application, one as a referee, and one expressing departmental support of my training. I also thank Dr. Gopal Thinakaran, a member of my thesis committee, whose neurobiological perspective was critical for the design of several important experiments throughout my training.

I thank the members of the Koide and Araç Labs for contributing to my training, scientifically and socially. From the Araç Lab, I thank Katherine Leon, Yue Lu, Olha Nazarko, Celia Fernandez, Jingxian Li, Chen Ding, Amanuel Kibrom, and Shu Zhang. From the Koide Lab, I thank Fern Sha, Irena Dimentiva, Takamitsu Hattori, Kohei Kurosawa, Ankit Gupta, Rob Wells, Raphael Eguchi, Chris Risley, Jay Goodman, and Akiko Koide. I extend my most heartfelt thanks to Akiko, who was never too busy to discuss my latest experiments or the new restaurants in Chicago or New York City. Along with Akiko's world-class practical instruction, her mentorship has made me a better scientist. I will always look up to Akiko and be grateful for the privilege to work with her at the bench.

The administrations of the Medical Scientist Training Program (MSTP) and Biophysical Sciences Program have supported me throughout my training, at several times vitally. Specifically, I thank Drs. Marcus Clark, Elise Covic, Alison Anastasio, Julie Feder, and of course Michele Wittles, whose kindness, guidance, and wise advice will always be profoundly appreciated.

I acknowledge the members of my cohort: Jeff Bunker, Justin Chew, David Koren, Rajiv Nadadur, Ramya Parameswaran, Sylvia Ranjeva, Ian Roundtree, Nicelio Sanchez-Luege, and Katherine Zhao, as well as Hector Acaron Ledesma and Joshua Riback. I am thankful to be a member of such a close-knit group of smart and fun people. I also thank Alan Hutchison for blazing the trail for MSTP students joining the Biophysical Sciences graduate program, and sharing his insights throughout his experience.

I thank Engin Özkan and the staff at the Advanced Photon Source at Argonne National Labs, specifically Craig Ogata, at GM/CA, for their help with x-ray crystallography. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. I also thank Navraj Pannu for his assistance in implementing the CRANK2 software package for experimental phasing. I thank the A. Kossiakoff and C. He labs for the use of their luminescence plate readers and the T. Sosnick lab for the use of their CD spectrometer. Supported by Brain Research Foundation (D.A.), Big

Ideas Generator (D.A.), and NIH grants U54-GM087519 (S.K.), R01-GM120322 (D.A.), F30-GM116455 (G.S.S.), F31-NS087801 (S.D.A.), R01-NS079445 (K.R.M.), and T32GM007183.

ABSTRACT

G protein-coupled receptors (GPCRs) have emerged as incredibly successful drug targets. Members of the adhesion GPCRs (aGPCR) family are characterized by diverse extracellular regions (ECRs), which play roles in cell adhesion, and mediate a subset of aGPCR functions *in vivo*. Though these receptors are implicated in myriad disease processes, there are no aGPCR-targeted therapeutics to date, due in large part to both the absence of well-behaving ligands as well as the technical challenges associated with mechanistic studies of aGPCR ECRs. We present a structural and functional study of the aGPCR GPR56/ADGRG1, a receptor critical for neurodevelopment and leukemia progression. To overcome many of the challenges mentioned above, we generated over thirty synthetic protein ligands, termed monobodies, that bind diverse epitopes across the ECR. Using a monobody crystallization chaperone, we solved the structure of the full ECR of GPR56, a first for any aGPCR, revealing the domain boundaries in the ECR as well as the identity and unique fold of the previously undefined N-terminal domain. We showed this domain regulates signaling and natural ligand binding *in vitro*, is deleted via alternative splicing, and mediates myelination *in vivo*. Additionally, we developed monobodies with stimulatory and inhibitory functions, demonstrating that ECR-targeted ligands can directly regulate aGPCR signaling, and are therefore valuable experimental reagents as well as lead-molecules for therapeutic development. Our results suggest an intricate, ECR-mediated molecular mechanism underlying aGPCR regulation. With the ultimate goal of combating aGPCR-mediated diseases, our findings will pave the way for targeted therapeutic development.

CHAPTER I

Introduction

Cell surface receptors

The physical separation of the contents of a cell from its environment is paradoxically an indispensable component of cellular life, while also a significant hindrance on cellular vitality. Of course, without such a physical barrier, known as the cell membrane or plasma membrane, a cell could not exist, as its contents would merely contribute to the environmental milieu. However, if the cell membrane were impenetrable, the cell would be isolated from its surroundings, blind to critical environmental signals, and multicellular organisms, which rely on a complex spatiotemporal network of intercellular signals, could not have evolved. As such, the cell membrane is “semi-permeable”, allowing some signaling molecules, such as steroid hormones, to freely diffuse through it (1). However, the biophysical properties of many important signaling molecules prohibit them from diffusing directly across the cell membrane into the cytoplasm.

To enable cellular recognition of signaling molecules that are unable to freely diffuse across the cell membrane, proteins are incorporated into membrane that recognize and transduce specific signals from the environment, through the membrane, into the cytoplasm. For example, ion channels or transporters recognize a specific ion or molecule and allow it access into the cytoplasm, where it may directly interact with intracellular components and effect function. In contrast, other types of transmembrane proteins, including some cell surface receptors, transduce information into the cytoplasm without allowing matter to cross the cell membrane (i.e. only the information

that the extracellular signaling molecule exists is a signal communicated to the cytoplasm). This type of signal transduction is responsible for countless pathophysiological and pharmaceutical processes (Figure I.1).

Some proteins that are traditionally classified as cell surface receptors do allow matter into the cell, (e.g. ion channel receptors, like acetylcholine receptor (2), or receptors that primarily mediate endocytosis of a particular molecule, like low-density lipoprotein receptor (3)). Furthermore, some transmembrane proteins initially classified as simple adhesion molecules (i.e. cadherins (4)) were later found to play key, receptor-like roles in intracellular signaling (5). Thus, though cell surface receptors have traditionally been classified based on protein architecture and signal transduction mechanism, the literature contains inconsistent nomenclature.

For clarity, in this thesis all transmembrane proteins that primarily transduce information across the cell membrane, rather than matter, will be referred to as cell surface receptors, and classified as one of the following: single-pass enzyme-linked receptors, extracellular matrix (ECM) receptors, and G protein-coupled receptors (GPCRs). Though this thesis is focused on the adhesion GPCR (aGPCR) family, understanding their biological context is critical. Thus, I will provide a brief overview of the mechanisms of signal transduction utilized by other types of cell surface receptors before discussing aGPCRs.

A common conceptual framework of signal transduction bridges the various cell surface receptor superfamilies, though the underlying mechanisms vary significantly. Three sequential events constitute cell surface receptor-mediated signal transduction: first, an extracellular signaling molecule, known as a ligand, interacts with a region of

the receptor exposed to the extracellular space; second, a binding-induced conformational change is allosterically communicated to the intracellular region (via the membrane-spanning region); and third, cytoplasmic effectors interact with the ligand-bound state of the intracellular region, resulting in an intracellular response to the extracellular signal.

Single-pass enzyme-linked receptors

Single-pass enzyme-linked receptors, commonly known as enzyme-linked receptors or catalytic receptors, make up a large and intensely-studied class of cell surface receptors. These receptors bind myriad ligands ranging from hormones to cytokines to pathogen-associated molecular patterns. Including the receptor tyrosine kinase (RTK), cytokine receptor, and toll-like receptor (TLR) families, enzyme-linked receptors are characterized by a key enzyme-catalyzed reaction during signal transduction (6).

The domain architecture of RTKs, including members of the ErbB (e.g. epidermal growth factor receptor) and epherin receptor families, comprises an extracellular ligand-binding domain, a single transmembrane helical domain, and an intracellular kinase domain. Ligand-mediated activation results in receptor dimerization, facilitating kinase domain-mediated phosphorylation of Tyr residues on the opposing monomers (7). Cytokine receptors have a similar domain architecture with an extracellular ligand binding domain and single membrane-spanning helix. Additionally, cytokine receptors, as with RTKs, are canonically activated by ligand-mediated dimerization. However, in contrast to RTKs, cytokine receptors do not have a catalytic cytoplasmic domain.

Rather, their cytoplasmic domains recruit cytoplasmic enzymes, which catalyze reactions upon activation (6, 8). A classic example of one of these enzymes is the Janus kinase (JAK), which is recruited by the inactive state of cytokine receptors. Upon receptor activation and subsequent dimerization, JAK phosphorylates the receptor, resulting in a downstream signaling cascade, classically involving signal transducers and activators of transcription (STAT) proteins (9).

TLRs have similar domain architecture to cytokine receptors. However, unlike RTKs and cytokine receptors, ligand-mediated dimerization is unnecessary for TLR activation, though it has been reported (10). Similar to cytokine receptors, the cytoplasmic region of TLRs serves as a scaffold for adaptor proteins, like MyD88 (6, 11). Upon ligand-mediated activation, these adapter proteins recruit downstream signaling effector proteins, like members of the interleukin-1 receptor associated kinase (IRAK) family, which catalyze adapter protein phosphorylation, dissociation from the TLR, and downstream signaling (10).

The T-cell receptor (TCR) is a hyper-specialized type of cell-surface receptor that, like the TLRs, plays key roles in immunity. Though it is traditionally classified on its own, it may also be classified as a single-pass enzyme-linked receptor. The TCR is composed of one α, one β, and two ζ chains, each of which spans the membrane. The α and β chains each have a large extracellular ligand-binding domain and a short cytoplasmic domain. The ζ chains have a short extracellular domain and a long intracellular domain, which recruits contains immunoreceptor tyrosine-based activation motifs (ITAMs). Additionally, the TCR is found in complex with multiple co-receptors including the CD3 ε, γ, and δ chains which also contain ITAMs (12). During T-cell

maturation, the TCR undergoes somatic recombination and hyper-mutation, such that each T-cell expresses a unique TCR variant. Each unique TCR is activated by a unique set of peptide antigens presented by members of the major histocompatibility complex (MHC) protein family by a neighboring cell. Upon peptide-MHC binding, ITAMs are phosphorylated by SRC family kinases, which are believed to be recruited by TCR co-receptors CD4 and CD8. Upon ITAM phosphorylation, the downstream effector ZAP70 is recruited and activated via phosphorylation, resulting in downstream signaling (13).

Though conventionally studied in the context of immunological effector function, the TLR and TCR are extremely similar to the other single-pass enzyme-linked receptors in the context of protein architecture and signal transduction. The signaling pathways regulated by single-pass enzyme-linked receptors are thus involved in diverse biological processes.

Extracellular matrix receptors

The dynamic meshwork of macromolecules that forms the extracellular matrix (ECM) is constantly remodeled by cells. Information flow between a cell and its environment is vital, especially during development, and such communication is facilitated by ECM receptors. Originally thought to function primarily as adhesive anchors, proteins from the integrin, cadherin, selectin, and Ig-like cell adhesion molecule (IgCAM) families have been shown to transduce signals into the cell. Furthermore, some of these proteins have been found to participate in so-called “inside-out” signaling, whereby cytoplasmic information is transduced through the receptor to ECM proteins or neighboring cells (5, 14).

Integrins are heterodimeric receptors composed of an α chain and a β chain, leading to combinatorial diversity (15). Though the two chains are structurally distinct, each has a single transmembrane helix. Macroscopically, the extracellular region (ECR) of the heterodimer consists of a large globular “head” region and a thin “tail” region. Relative to the ECR, the transmembrane region and cytoplasmic region are small. Upon ligand binding, integrin ECRs undergo large-scale conformational changes, resulting in the separation of the transmembrane and cytoplasmic regions by ~70Å within the heterodimer (15). While the cytoplasmic domains of integrins primarily interact with cytoskeletal proteins, like focal adhesion kinase (FAK), that regulate cell shape and migration, the extracellular domains interact with extracellular matrix proteins, including collagen, or adhesion molecules on neighboring cells. Thus, integrins are able to transduce extracellular signals such as mechanical force into appropriate cellular responses, like migration (16).

Cadherins were first characterized as a physical scaffold protein involved in stable cell-cell contacts, particularly in adherens junctions (17). Like integrins, cadherins have an extracellular ligand binding domain, transmembrane domain, and cytoplasmic domain that interacts with cytoskeletal proteins, specifically catenins, which regulate small GTPase activity (e.g. Rho, Rac, and Cdc42) (5, 17). Also like integrins, the mature receptor is made up of a cadherin dimer. However, cadherins do not generally form heterophilic interactions with other protein ligands. Rather, cadherin dimers homophilically interact with cadherin dimers on neighboring cells, forming a classical tetrameric cell-cell attachment (18). Though cadherins undoubtedly play critical

structural roles, their ability to transduce extracellular signals classifies them as important cell surface receptors.

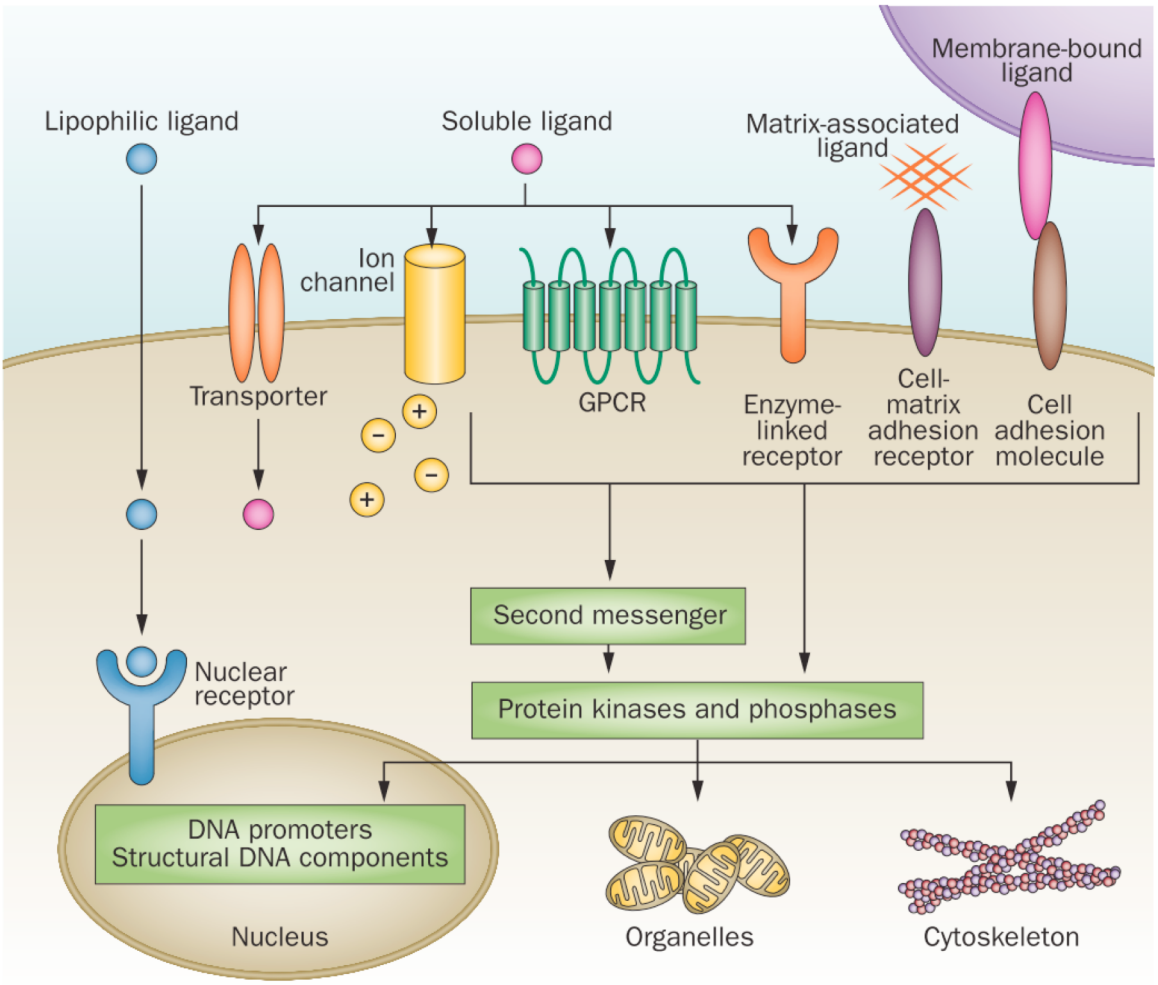


Figure I.1 Overview of cellular signaling mediated by extracellular ligands.

A lipophilic ligand may freely diffuse across the plasma membrane and interact with intracellular receptors. Lipophobic soluble ligands may enter the cell via transporters or channels or interact with the extracellular regions of cell-surface receptors and thereby mediate intracellular signaling. Matrix-associated or membrane-bound ligands may interact with extracellular matrix cell surface receptors to initiate intracellular signaling events. Figure adapted from Reiser, et al., 2014 (234).

Unlike integrins and cadherins, IgCAMs are a diverse family of proteins, defined only by their presumptive role in cell adhesion and the presence of one or more Ig domains in their ECR (19, 20). As such, their roles in signaling need not be conserved, requiring each IgCAM subfamily to be independently characterized. One of the most well-studied roles of IgCAMs involves endothelial cell signaling to migrating leukocytes. In this example, the endothelial cells express ICAM-1, an IgCAM that regulates Rho-like GTPase-dependant cytoskeletal changes upon receptor clustering (20). Migrating leuokyctes express particular integrin receptors, which bind ICAM-1, effectively serving as its ligand. Of course, depending on perspective, the ICAM-1 may be considered the ligand for the integrins as well. In addition to this interaction and its associated signal transduction, glycans displayed on the surface of the endothelial cells and leukocytes serve as ligands for L- and P-selectins, receptors expressed on leukocytes and endothelial cells, respectively. Upon glycan binding, selectins also transduce signals which generally manifest as kinase-mediated phosphorylation cascades and adapter protein recruitment to modulate the integrin signaling described above (21). This example illustrates the complex and overlapping roles played by cell-surface receptors involved in intercellular and cell-ECM signaling networks.

G protein-coupled receptors

With over one thousand members in the human genome GPCRs make up one of the largest protein superfamilies (22, 23). All GPCRs have an extracellular N-terminus, a hallmark seven-pass transmembrane helix bundle (7TM), and a cytosolic C-terminus. Though the length and architecture of the 7TM is relatively well-conserved, phylogenetic

analysis of 7TM sequences has resulted in the identification of five GPCR families: rhodopsin, adhesion, secretin, glutamate, and frizzled/TAS2. In addition to relatively subtle 7TM sequence variations between GPCR families, ECR and ICR length and architecture vary extensively across and even within families. The majority of mechanistic and functional studies of GPCRs have focused on receptors without large ECRs, particularly members of the rhodopsin family (24–26). This enormous body of work has elucidated the mechanisms by which these receptors transduce signals.

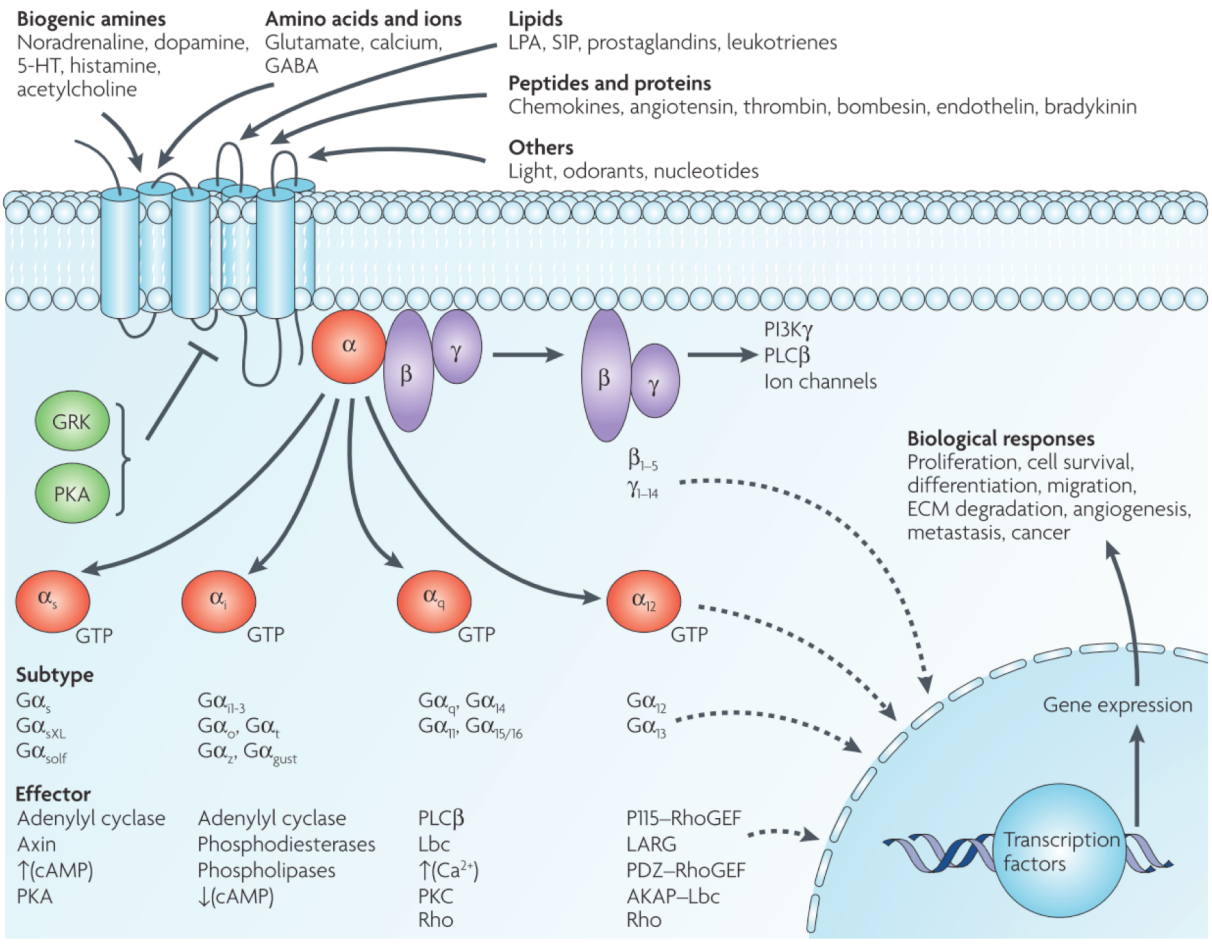


Figure I.2 Overview GPCR ligands and downstream signaling pathways.

Diverse GPCR ligands elicit downstream signaling through one or more G protein signaling pathways. Figure adapted from Dorsam and Gutkind, 2007. (235)

Though the cell surface receptors described in the previous sections utilize individual protein domains for extracellular ligand binding, membrane spanning, and intracellular effector function, canonical G protein-coupled receptors (GPCRs) are composed of a single domain, known as the seven-pass transmembrane helix bundle (7TM), which is responsible for all three events.

GPCR ligands include photons, ions, small molecules, lipids, and proteins. Canonically, GPCRs are activated when a ligand interacts with a specific region, known as the orthosteric ligand binding site, which is made up of the extracellular loops and face of the 7TM. Upon ligand binding, the 7TM undergoes a conformational change, usually associated with movement of the cytoplasmic end of transmembrane helix 6, resulting in an increased affinity for cytoplasmic heterotrimeric G proteins (24, 27, 28). When bound, the Gα subunit undergoes GDP-GTP exchange and dissociates from the Gβγ heterodimer. Both the activated Gα-GTP and the free Gβγ go on to initiate downstream signaling cascades.

There are four major classes of heterotrimeric G proteins, classically defined by their Gα subunit. Upon activation, $G\alpha_s$ positively regulates adenylyl cyclase (AC), resulting in increased cyclic adenosine monophosphate (cAMP) concentration. The inhibitory G protein, $G\alpha_{i/o}$ has the opposite effect on AC, resulting in decreased cAMP. $G\alpha_{q/11}$ positively regulates phospholipase C (PLC), resulting in increased diacyl glycerol (DAG) and inositol triphosphate (IP_3). $G\alpha_{12/13}$ activates RhoGEFs, resulting in activate RhoA. Upon dissociation from Gα, the Gβγ heterodimer also regulates cellular signaling, including affecting AC and PLC activity (27). Though these four Gα classes serve as a convenient framework, they fail to accurately describe the functions of many

G proteins. Furthermore, it has been repeatedly demonstrated that a single GPCR may couple to several different Gα subunits, and different ligands may introduce bias towards a particular one, a phenomenon known as “biased signaling” (29, 30). Thus, accurately modeling the downstream signaling pathway(s) of a particular GPCR in the cellular context is challenging (Figure I.2).

GPCRs have been incredibly successful pharmaceutical targets (31–33). As such, the field has established a precise and detailed nomenclature for classifying GPCR signaling as well as their ligands (34). In basic terms, even in the absence of a ligand, GPCRs may display low levels of activation, referred to as ‘basal activity’. A molecule that increases signaling above basal activity is known as an ‘agonist’ and a molecule that reduces the level of agonist-mediated activation is known as an ‘antagonist’. A molecule that decreases signaling below basal activity is known as an ‘inverse agonist’. Finally, a molecule that binds the GPCR but does not alter signaling is known as a ‘neutral ligand’. Indeed, agonists, antagonists, and inverse agonists are all classified as ‘ligands’ because they all bind the receptor. Furthermore, ligands may be classified as ‘orthersteric’ if they bind the orthosteric site. All non-orthosteric ligands are referred to as ‘allosteric’.

In order to determine the effect of a particular ligand on G protein signaling, a wide variety of assays have been developed; it is possible to measure signaling anywhere in the cascade. Though direct measurement of Gα activation, for example, is free from any cross-talk that may occur in downstream effectors, this readout is not amplified as the signal propagates downstream (35). As such, the low magnitude of the signal may be harder to detect. Conversely, transcription-based readouts are far

downstream and therefore undergo amplification, but suffer from hours of lag between receptor activation and readout response, and may be subject to cross-talk (36). Overall, the wealth of mechanistic and pharmacological insight into the biological roles of GPCRs has been invaluable for advancing studies of all types of cell surface receptors and cell signaling.

Adhesion GPCRs

The vast majority of studies focused on GPCRs have targeted receptors from the rhodopsin family, resulting in an in-depth understanding of how these receptors function. However, due in part to technical reasons, these studies have generally ignored GPCRs with prominent extracellular regions (ECRs), despite the pathophysiological relevance of these more complicated receptors. Adhesion, secretin, frizzled/taste2, and even some rhodopsin family members have one or more extracellular domains (22). Only recently have studies of GPCR ECRs began to elucidate their key regulatory roles. Such studies focues on secretin family GPCRs have elucidated the mechanism by which their ECRs (which consist of a relatively small hormone-binding domain) interact with the C-terminus of a peptide hormone ligand and thereby direct its N-terminus into the orthosteric site of the 7TM (37–39). Similar studies have proposed that secretin and frizzled/TAS2 family GPCR ECRs may regulate the receptor even in the absence of a ligand (40–44). However, in the case of the adhesion family, hundreds to thousands of amino acids comprise diverse extracellular regions (ECRs) consisting of multiple protein domains, many of which have been bioinformatically identified as adhesion-type domains (45, 46). Though some of these

domains have been shown to interact with other proteins on neighboring cells or in the ECM, the link between binding and signaling is not well understood.

The sheer size of aGPCR genes has led to substantial technical obstacles when attempting genetic or biochemical studies. As such, despite their pathophysiological relevance (45, 47), the study of aGPCRs has lagged significantly behind the other GPCR families. However, in the past decade, genetic studies have repeatedly identified aGPCR as key players in numerous disease processes (48–54). These results,

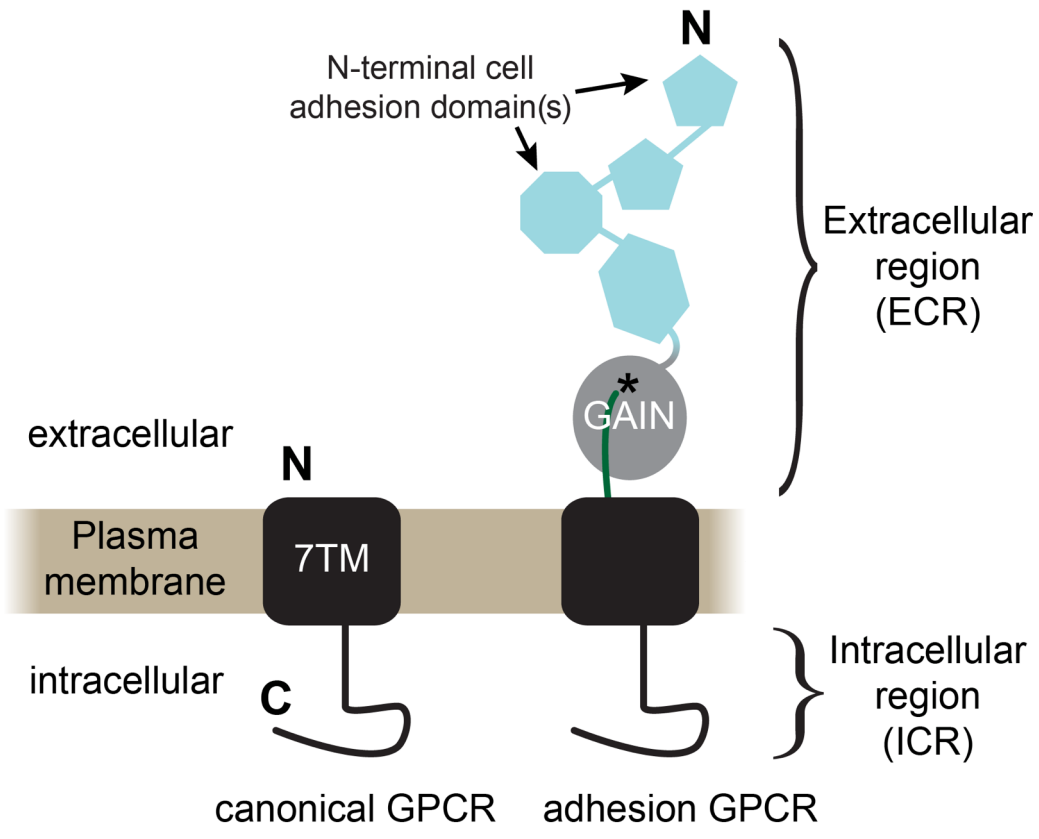


Figure I.3 Domain architecture of GPCRs.

GAIN domain cleavage site indicated by *. The C-terminal β-strand of the GAIN domain (i.e. *Stachel*) is colored green. Figure adapted from Salzman, et al. (93).

combined with technological advances in molecular biology, biochemistry, and genetics have jump-started a thorough investigation into aGPCR biology and pharmacology.

The domain architecture of aGPCR ECRs includes one or more N-terminal adhesion-type domains and a conserved juxtamembrane GPCR Autoproteolysis INducing (GAIN) domain (55). Importantly, during aGPCR maturation, an autoproteolytic event occurs within the GAIN domain, cleaving the receptor into two fragments: a soluble N-terminal fragment (NTF; composed of the N-terminal adhesion domains and the majority of the GAIN domain) and a membrane-anchored C-terminal fragment (CTF; composed of the C-terminal β-strand of the GAIN domain, the 7TM, and the intracellular region [ICR]; Figure I.3). Because the autoproteolysis site is within the core of the GAIN domain, the NTF and CTF do not dissociate upon cleavage. Rather, the folded GAIN domain ensures they remain noncovalently associated to form the mature receptor, which is trafficked to plasma membrane (55). While some aGPCR GAIN domains do not facilitate robust autoproteolysis (55, 56), of those that do, tissue-specific variations in cleavage efficiency have been observed (57). Furthermore, as some, but not all aGPCR functions require autoproteolysis (58), a unified mechanistic model detailing both the autoproteolysis-dependent and -independent biological roles played by aGPCRs has been elusive.

Latrophilins/ADRGL

The Latrophilin (LPHN) aGPCR family has four members: LPHN1-3 and ELTD1. LPHN1 was first identified as the receptor for α-latrotoxin, the toxin in black widow spider venom that stimulates massive exocytosis of synaptic vesicles (59). Two

decades of follow-up experiments have identified endogenous LPHN ligands and elucidated many pathophysiological roles of LPHN proteins, most of which are centered in neurobiology. Using LPHN ECRs as bait for pull-down/mass spectrometry experiments, Lasso/teneurin (60) and fibronectin leucine-rich repeat transmembrane (FLRT) (61) proteins were identified as LPHN ligands. Additionally, LPHNs were shown to regulate intercellular adhesion mediated by LPHN-neurexin binding (62). Teneurin proteins are found highly expressed in the brain and play roles in neurodevelopment, neurite outgrowth, axon guidance, neuronal connectivity, and synaptogenesis (60, 63, 64). Though the LPHN-tenurin interaction is regulated by alternative splicing of LPHN (65), and may mediate cytoskeletal remodeling and synaptic plasticity (66), its functional significance *in vivo* remains incompletely understood. Though FLRTs are also expressed in the central nervous system (CNS), they are found in other tissues too (61). Via their interaction with UNC5 proteins FLRTs have been shown to regulate axon guidance (67). Together LPHN and FLRT regulate glutamatergic synapse density and function (61).

The N-terminal adhesion domains of LPHN include an N-terminal lectin domain, followed by an olfactomedin and hormone-binding domain. The ECR also contains a characteristic aGPCR GAIN domain. Though FLRT was identified as an ECR-binding LPHN ligand, its precise binding site was unknown. Furthermore, as both LPHN and FLRT are transmembrane proteins, it was unclear if the LPHN-FLRT interaction occurs in *cis* or in *trans*. Finally, as FLRT also interacts with UNC5, it was unknown if LPHN and UNC5 could simultaneously interact with FLRT, or if binding was mutually exclusive. Answers to these questions are critical in order to understand the roles of

LPHN, FLRT, and UNC5 in brain development (68). The seventh chapter of this thesis is devoted to the structural characterization of the LPHN-FLRT interaction.

GPR56/ADGRG1

The aGPCR GPR56/ADGRG1/TM7XN1 was first identified in 1999 based on its sequence homology to secretin family GPCRs and its high expression in the thyroid gland (69). In the years since, GPR56 has become one of the most well-characterized aGPCRs. Its pathological relevance was first established in 2004 when point mutations in GPR56 were shown to cause bilateral frontoparietal polymicrogyria (BFPP), a nonlethal, yet debilitating brain malformation (70). Clinically, patients with BFPP exhibit severe intellectual disability, motor and language impairment, seizures, and shortened life expectancy (71). On MRI, the brains of BFPP patients exhibit two characteristic findings: loss of cortical gyri and reduction in white matter (70), suggesting GPR56 plays roles in both cortical development and myelination.

More recent studies have implicated GPR56 in an incredibly diverse array of developmental, biological, and pathological processes: cortical development (57, 72–76), oligodendrocyte development (77, 78), muscle cell development (79), cytotoxic lymphocyte function (80, 81), and the progression of various cancers including melanoma (82, 83), acute myelogenous leukemia (AML) (49, 84), and colorectal cancer (48). Furthermore, several GPR56 ligands have been reported including tetraspanins CD9 and CD81 (81, 85), collagen III (75, 86), tissue transglutaminase (TG2) (82, 87), heparin (88), and progastrin (48). Finally, using various G protein signaling assays, it

has been established that GPR56 signals through RhoA via $G\alpha_{13}$ (35, 57, 75, 86, 89–91) and reported that GPR56 may also signal through $G\alpha_q$ (92).

Despite all of these studies into the biological role of GPR56, its domain architecture remained a mystery: bioinformatics predications strongly suggested a GAIN domain with an unidentifiable N-terminal boundary, and an unknown N-terminal adhesion domain. Though three of the GPR56-binding ligands mentioned above were reported to bind to this unknown N-terminal adhesion domain (75, 87, 88), no effort had been made to identify it or characterize any role it may have in regulating receptor signaling. The fourth, fifth, and sixth chapters of this thesis detail a multidisciplinary, structural and functional characterization of the ECR of GPR56, including the identification of this N-terminal adhesion domain (93), and the development of a unified model for ECR-regulated autoproteolysis-dependent and -independent aGPCR signaling.

Synthetic protein scaffolds as biological tools[1]

Synthetic binding proteins are human-made proteins that have been tailored to bind to a target molecule of interest. The capability of the immune system to generate antibodies binding to virtually any antigens and the knowledge of the molecular

[1] Text in this section was taken verbatim (with minor changes) from:
Sha F, Salzman G, Gupta A, and Koide S. (2017) Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci* 26(5):910–924.
My contributions to this review article included writing and editing the text and performing the monobody concavity analysis shown in Figure 3 of the review, reproduced here in Chapter III. (233)

mechanisms underlying this capability have inspired the genesis and subsequent development of the field of the design and engineering of synthetic binding proteins. Analogous to how natural antibodies for diverse antigens are made by altering portions of the immunoglobulin molecule, synthetic binding proteins are most commonly generated by altering portions of a functionally inert protein, referred to as a protein scaffold. Synthetic binding protein systems are developed with the ultimate aim of generating binding proteins to diverse target molecules, rather than binding proteins to one specific target. These proteins are synthetic in that they have not been found in nature, although they are polypeptides consisting of natural amino acids and made using natural machinery for protein synthesis.

Synthetic binding proteins are usually generated by introducing multiple mutations, typically 10-20, to a protein scaffold (Figure I.4A). Directed evolution approaches, in particular those utilizing molecular display technologies, enables one to efficiently generate a vast ensemble ("library") of mutants and identify clones that bind to the target molecule of interest with high affinity. The starting scaffold systems are usually chosen with the hope of generating synthetic binding proteins with desirable functional and biophysical properties, including the ability to generate high-performance molecular recognition interfaces for diverse targets, small size, high stability, ease of production and ease of use as a building block in fusion proteins.

Choosing an appropriate starting scaffold is an important step, but it is equally important to carefully choose how portions of the scaffold are diversified in a combinatorial library. Many practitioners in the field originally thought that, given the capacity of molecular display methods to test billions of sequences, it should be

straightforward to produce high-performance binding proteins by introducing amino acid diversity using a random mixture of all possible codons, such as NNN and NNK where N is a mixture of A, T, G and C and K is a mixture of T and G, at casually chosen positions. However, they quickly found that this was not the case. Binding proteins generated from such libraries often had low affinity and low specificity. A major breakthrough came from the work of Sidhu and colleagues on synthetic antibody libraries (94–97). They established that the utilization of a highly biased distribution of amino acids (with particular enrichment of Tyr) in synthetic libraries is highly effective in generating potent and specific antibodies. Parallel studies demonstrated that the equivalent approach is effective even in a much smaller synthetic scaffold, Monobody (see below). The reader is referred to reviews dedicated to this topic (98, 99).

Among synthetic binding protein platforms, the most established systems include Affibodies, Anticalins, Monobodies and DARPins (Figure I.4B). Affibodies are based on the Z domain of protein A from *Staphylococcus aureus*. They contain three α-helices, no disulfides, and are among the smallest synthetic binders (~6 kDa) that have been well characterized (100–103). Anticalins, based on lipocalins, have a β-barrel architecture with an attached α-helix. While some lipocalins do contain disulfides, they are exploited due to their natural ability to bind to small molecules using their barrel and loops, and this mode of binding has been exploited for Anticalin libraries (104–107). Monobodies are based on the fibronectin type III (FN3) domain that has an immunoglobulin fold, but no disulfide bonds (108). Following successes of Monobodies and their equivalence in the industry, Adnectins, several "Monobody mimics" have been successfully developed (109, 110), demonstrating the robustness of the FN3 scaffold for

generating synthetic binding proteins. Designed ankyrin-repeat proteins (DARPins) exploit repetitive structural units to form an extended binding surface (111). DARPins also lack disulfide bonds yet exhibit high thermodynamic stability (112, 113). Although these platforms are based on proteins with distinct folds, they have all produced high-performance synthetic binding proteins against diverse targets. These numerous successes clearly show that the synthetic binding protein field has collectively established sufficient knowledge and technologies for developing a scaffold system.

Structural analyses of Monobody-target complexes revealed that in addition to the intended mode of target interaction mediated by the diversified loops, a distinct mode was observed in which (unmutated) residues on the β-sheet surface ("side" of the scaffold when we place the diversified loops at the "top") contributed to target recognition. Inspired by this observation, a new library was constructed in which residues on a β-sheet were diversified (114). Monobodies from the new "side" library presented a concave surface for recognition, as opposed to convex surfaces typically found for Monobodies from the original "loop" libraries, therefore expanding the diversity of binding site topography. As intended by the designs, these two distinct libraries show preferences toward differently shaped surfaces. The loop library tends to prefer binding into a concave epitope, whereas the side library prefers a flatter surface. For example, in an unbiased library selection experiment against the Abl SH2 domain, that is, a selection that did not involve a step that steer binders to a specific epitope, a dominant Monobody clone from the loop library bound to the concave, peptide-binding groove, whereas a dominant clone from the side library bound to a flat surface on the opposite side of the SH2 domain (115, 116). These results illustrate the possibility of expanding

the efficacy of a scaffold system by the use of distinct surfaces for presenting amino acid diversity and thereby expanding the types of epitopes that can be effectively recognized.

The use of antibody fragments as crystallization chaperones has made important contributions to the successes of challenging structural biology projects. Crystallization

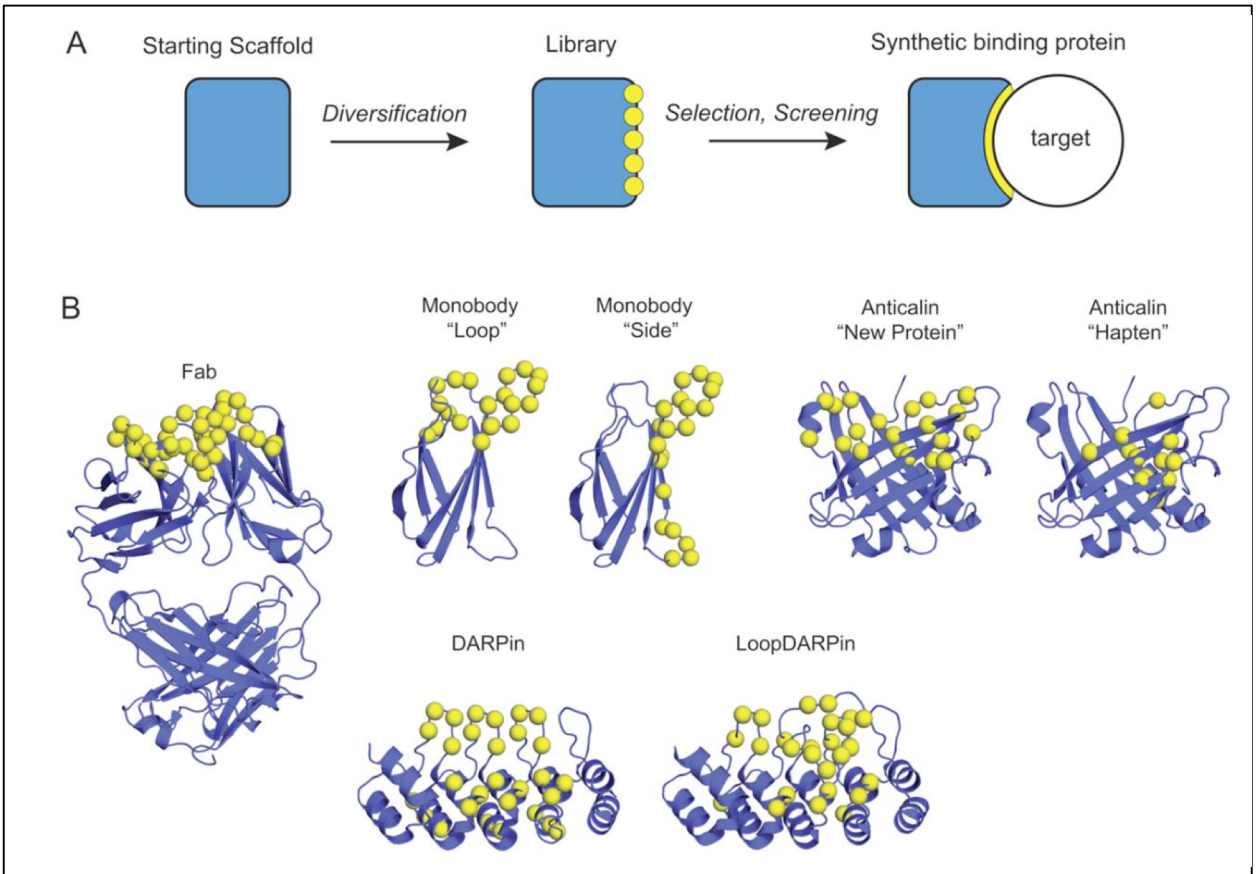


Figure I.4 Generation of synthetic binding proteins using molecular scaffolds.

(A) A schematic representation of the processes for generating synthetic binding proteins. The rectangle at the left indicates an inert scaffold. The yellow circles denote sequence diversity at chosen positions. The right is a binding protein-target complex with the optimized interface shown in yellow. (B) The three-dimensional structures of representative scaffold architectures, including the natural antibody for which only the Fab portion is shown. Only those scaffolds for which structure-guided design of libraries have led to improved performance are shown. References are given in the main text. Figure adapted from Sha et al., 2017 (233).

chaperones can increase the likelihood of producing macromolecular crystals suitable for diffraction studies through several potential mechanisms including (i) reducing the fraction of disordered regions (ii) reducing the conformational heterogeneity and (iii) providing surfaces that are conducive to forming crystal contacts (117, 118). Although antibody fragments such as Fab and Fv are still the most common crystallization chaperones, the ability to produce large quantities of stable, high-affinity binding proteins in *E. coli* has made synthetic binding proteins attractive alternatives. Unlike Fab that exhibits substantial hinge bending motions between the variable and constant domains (119), synthetic binding proteins and also nanobodies (single-domain antibody fragments derived from the camelid heavy chain-only antibodies) are single-domain proteins and thus do not have such internal mobility. This attribute seems to contribute to the ability of these single-domain chaperones to help produce higher-resolution structures. A combination of a Monobody chaperone and linking of heterodimer into a single-chain construct was used to determine the structure of an otherwise ill-behaving Prdm14-Mtgr1 complex (120). Furthermore, their small sizes may be important for crystallizing integral membrane proteins using the lipid cubic phase method, because of the limited size of cavities that can accommodate water-exposed portions of the protein system, i.e. the water-exposed portion of the target protein plus the chaperone (24, 121, 122).

Monobodies bind functional sites[2]

Although these synthetic binding protein systems have been developed originally for the purpose of generating simple affinity reagents, ensuing research has revealed that many of them, particularly Monobodies, have a strong tendency to bind to a functional surface on the target molecule. This attribute makes them modulators of biological functions. Combined with high specificity, high affinity, simple design and ability to function regardless of redox potential of the environment, Monobodies offer unique capabilities beyond "just" affinity reagents. The following examples illustrate this capability that has contributed to advancing mechanistic understanding.

In a typical project of synthetic binding protein generation, many clones are available at the end of the selection campaign, and the "best" clones among the candidates are chosen based on their affinity, specificity and amino acid sequences. However, these clones are chosen without the knowledge of where within the target molecule they bind (epitope). Although it is technically straightforward to direct binding proteins to a specific surface, such an approach is taken only in a project that starts with a detailed mechanistic understanding of the target molecule and clear descriptions of the desired properties of binding proteins. Despite this unbiased selection in terms of epitopes, synthetic binding proteins, particularly Monobodies, are found to bind to a

---

[2]Text in this section was taken verbatim (with minor changes) from:
Sha F, Salzman G, Gupta A, and Koide S. (2017) Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci* 26(5):910–924.
My contributions to this review article included writing and editing the text and performing the monobody concavity analysis shown in Figure 3 of the review, reproduced here in Chapter III. (233)

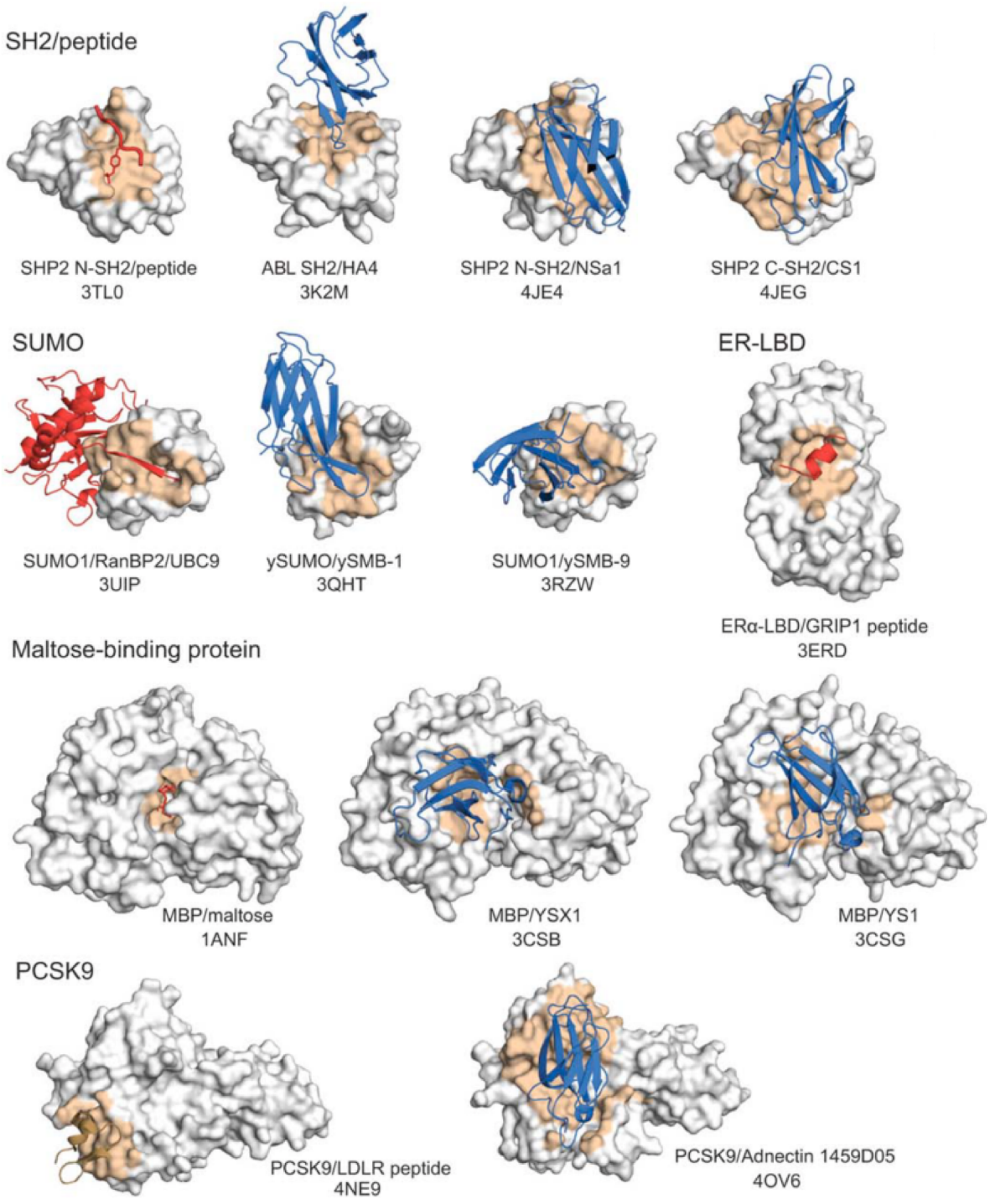


Figure I.5 Examples of Monobodies and Adnectins binding to a functional site within the target protein.

(continued on the following pages)

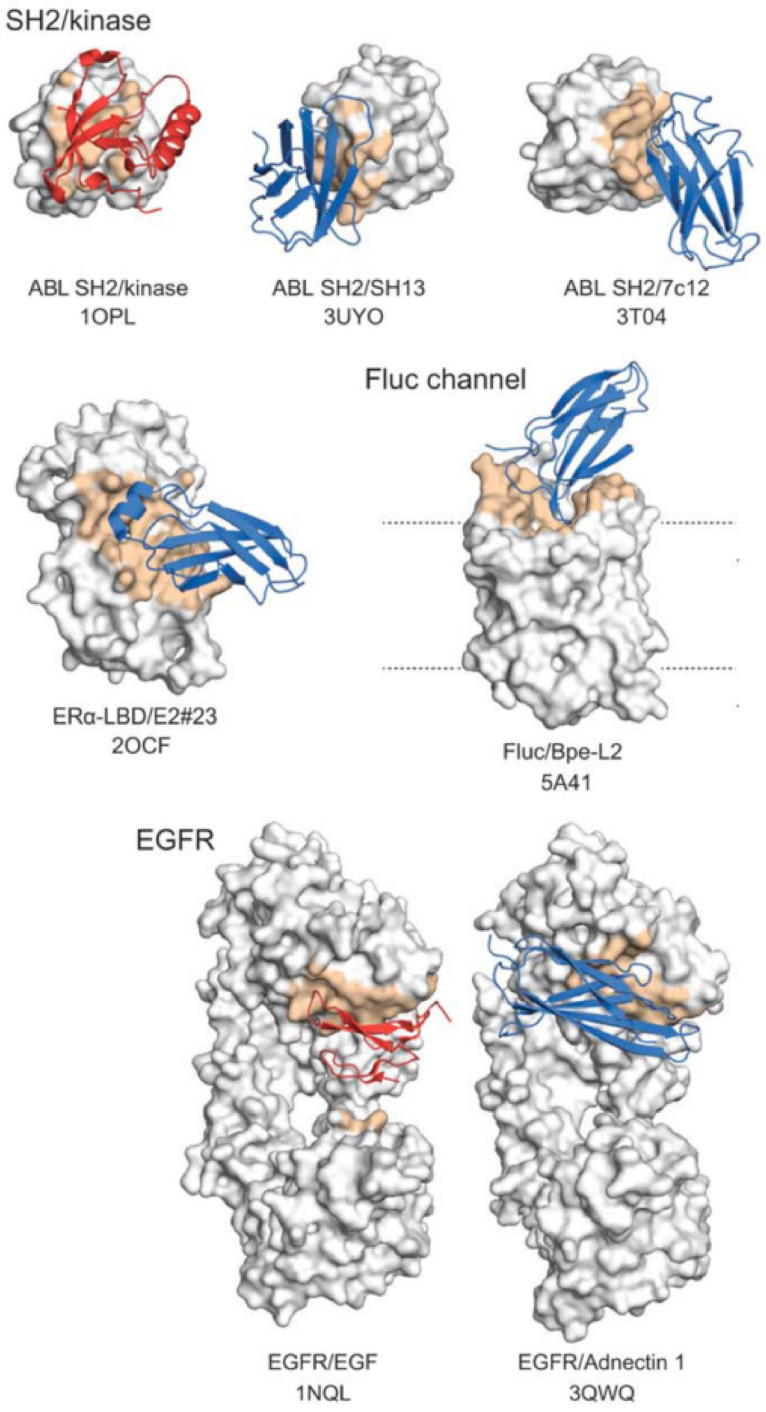


Figure I.5, continued.

(legend on the following page)

functional site within the target molecule (Figure I.5). The third chapter of this thesis discusses a geometric basis for understanding the different binding site preferences of several commonly-used synthetic protein scaffolds.

Though aGPCRs seem combine the paradigms of classical GPCR signaling with ECM receptor signaling, their precise biological nice remains poorly understood. With the goal of understanding the molecular mechanisms by which aGPCR ECRs mediate pathophysiological processes, and ultimately combat aGPCR-mediated pathology, we set out to characterize the ECRs of GPR56 and LPHN3, with respect to their structure and function. Due to the paucity of natural and synthetic GPR56 ligands, we engineered monobodies to bind the ECR of GPR56 to aid our studies. Specifically, we hypothesized that these monobodies may 1) stabilize the ECR to aid crystallization, 2) serve as valuable detection tools, 3) perturb natural ligand binding, and 4) modulate receptor signaling directly. Together, the interdisciplinary studies discussed in this thesis represent a key step in the characterization of this understudied family of cell surface receptors.

Figure I.5, continued.

The target proteins are shown in gray with the epitope in orange. Natural ligands are in red, and Monobodies and Adnectins in blue. The identities of the target molecules and PDB entry codes are indicated. For the Fluc channel structure, the natural ligand, F^- ion, is not shown because of its small size. Figure adapted from Sha et al., 2017 (233).

CHAPTER II

Materials and Methods

Methods used in Chapter IV[1]

Cloning and purification of aGPCR extracellular fragments from insect cells

Extracellular regions (ECRs) of the following aGPCRs were cloned into pAcGP67a: mouse GPR56 (ADGRG1) full ECR (UniProt: Q8K209, residues S27-S392); mouse GPR56 GAIN domain (residues M176-S392); human GPR56 full ECR (UniProt: Q9Y653, residues G27-S392); zebrafish Gpr56 full ECR (UniProt: F1QZM9, residues T25-E358); human latrophilin 3 (LPHN3/ADGRL3) HormR+GAIN domains (UniProt: Q9HAR2, residues E496-S856); rat latrophilin 1 (LPHN1/ADGRL1) HormR+GAIN domains (UniProt: O88917, residues P460-I849); human brain angiogenesis inhibitor 3 (BAI3/ADGRB3) HormR+GAIN domains (UniProt: O60242, residues E498-E868); human GPR112 (ADGRG4) GAIN domain (UniProt: Q8IZF6), residues E2450-S2731).

[1] The text of this section is copied verbatim (with minor changes) from: Salzman GS, Ackerman SD, Ding C, Koide A, Leon K, Luo R, Stoveken HM, Fernandez CG, Tall GG, Piao X, Monk KR, Koide S, and Araç D. (2016) Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. Neuron 91(6):1292–1304. G.S.S., S.K., and D.A. designed all experiments, interpreted results, and wrote the manuscript. G.S.S. performed all experiments involving aGPCR ECR cloning, purification, MALLS, and crystallography (with assistance from C.D.); monobody engineering, purification, and characterization (with guidance from A.K.); mammalian cell expression; SRE signaling (with assistance from C.G.F.); bioinformatics; and mass spectrometry. A.K. and S.K. provided the monobody libraries; S.D.A. and K.R.M. designed and performed zebrafish experiments. R.L. and X.P. designed and performed RT-PCR experiments. H.M.S. and G.G.T. designed and performed direct G protein coupling experiments (with assistance from K.L.). All authors discussed the results and provided comments and revisions on the manuscript. (93)

C-terminal 6xHIS tags were added for affinity purification. C-terminal AVI-tags corresponding to the sequence GLNDIFEAQKIEWHE were added to aid biotinylation.

A baculovirus expression system was used for expression of proteins in High Five insect cells as previously described (55). The secreted, glycosylated proteins were purified using nickel-nitrilotriacetic agarose resin (Qiagen) and size-exclusion chromatography (Superdex 200 10/300 GL; GE Healthcare).

Monobody generation

Biotinylated mouse GPR56 ECR was used as a target for phage-display selection from a 'side and loop' monobody library as previously described (114). The naïve library contained ~10^9 different clones. Four rounds of selection were performed at target concentrations of 1) 100 nM (tetramerized), 2) 100 nM (monomeric), 3) 50 nM (monomeric), 4) 50 nM (monomeric). A yeast display library containing ~10^6 different clones was constructed from the output of phage display selection. A single round of positive sorting of the yeast display library was done using fluorescent-activated cell sorting (FACS) using dye-labeled GPR56 ECR to stain yeast. Binding assay for testing the affinity and specificity of individual monobody clones was performed using yeast surface display as described previously (123).

Purification of monobodies from E. coli

The genes encoding the identified monobodies were cloned into an expression vector, pHBT (123). Monobodies were expressed in *E. coli* via autoinduction at 37˚C for 20 hours. Monobodies were purified via an N-terminal 6xHIS tag using nickel-

nitrilotriacetic agarose resin (Qiagen), and refolded on the Ni-column using the β-cyclodextrin method (124). Refolded proteins were gel-filtered using a Superdex 200 10/300 GL column (GE Healthcare).

X-ray crystallography data collection

Purified mouse GPR56 ECR was mixed with purified monobody α5 and the complex was purified by size-exclusion chromatography (Superdex 200 10/300 GL; GE Healthcare) and concentrated to ~22 mg/mL protein complex. Crystals grew to >0.1mm in 80 mM sodium acetate pH 4.6, 19.5% glycerol, 16.9% PEG 600, 7.6% PEG 1000. To obtain phase information, some crystals were treated with KI_3 using vaporizing iodine labeling (125) Native and iodinated diffraction data were collected to 2.45 and 3.00Å, respectively at the Advanced Photon Source, beamline 23-ID-B.

X-ray crystallography data processing, phasing, and refinement

Data were processed using HKL2000. To obtain a heavy metal substructure and calculate experimental phases, the CRANK2 (ccp4) software package (126) was used, despite the weak anomalous signal. The model generated from the anomalous data was used for molecular replacement (phaser.mr, ccp4) into the native dataset. Phenix.refine (PHENIX) was used for all refinement.

Dual luciferase SRE reporter plasmid construction

Dual luciferase SRE reporter plasmid was constructed to have constitutively expressed renilla luciferase and SRE-activated firefly luciferase on the same plasmid.

The pmirGLO (Accession Number FJ376737) and pGL4.33 [luc2P SRE-Hygro] (Accession Number FJ773212) vectors were obtained from Promega. The SRE-firefly luciferase region of pGL4.33 was used to replace the PGK-firefly luciferase. The resulting reporter plasmid is referred to as dualLUC-SRE.

G protein signaling assay

HEK293T cells were seeded in 24-well plates (45,000 cells in 0.5 mL DMEM+10% FBS/well). After 12-18 hours, cells reached 40-50% confluence and were transfected with 11.3 ng *Gpr56* (WT or mutant) + 45.0 ng dualLuc-SRE + 0.23 µL FUGENE6 per well from a master mix. After 24 hours, media was aspirated and replaced with DMEM + 0% FBS. For monobody treatment, monobody was added to cells 6.5 hours after the start of serum starvation. After 12 hours total of serum starvation, media was aspirated. Cells were lysed using the Dual-Glo® Luciferase Assay System from Promega and firefly and renilla luciferase signals were measured using a Synergy^{TM} Neo luminescence plate reader. Signaling intensity in RLU (fold increase) is reported as: $(Firefly_{GPR56}/Renilla_{GPR56})/(Firefly_{EV}/(Renilla_{EV})$.

Flow cytometry

HEK293T cells were transiently transfected with FL or mutant mouse *GPR56* constructs (openbiosystems clone ID: 3709247) using Fugene6. After 48 hours, cells were detached and stained. Flow cytometry was performed using Guava® easyCyte as previously described (68).

Monobody α5 staining: To measure binding affinity, cells were stained primarily with biotinylated monomeric monobody at a range of concentrations, and secondarily with labeled neutravidin. To detect binding (as in Figure S3), cells were stained with 100 nM pre-tetramerized monobody on labeled neutravidin in order to increase avidity in a single staining reaction.

FLAG staining: Cells expressing N-terminally FLAG-tagged constructs were stained primarily with 1:1000 mouse anti-FLAG M2 (Sigma) and secondarily with 1:100 anti-mouse-FITC.

Streptavidin pull-down and western blot

Streptavidin pull-down and western blot were performed as previously described (35). Briefly, HEK293T cells were transiently transfected with GPR56 constructs using FUGENE6. After 48 hours, cells were treated with EZ-Link™ Sulfo-NHS-LC-Biotin (ThermoFisher), quenched, lysed, subject to pull-down with Streptavidin MagneSphere® Paramagnetic Particles (Promega), and subject to western blot using an antibody against the GPR56 CTF (Millipore Cat#: ABS1028, RRID: AB_2617058; 1:1000 dilution).

Zebrafish stocks and rearing conditions

Zebrafish (*Danio rerio*) were maintained in the Washington University Zebrafish Consortium facility (http://zebrafish.wustl.edu/), and all experiments were performed in compliance with Washington University's institutional animal protocols. WT (AB-

Tubigen) embryos were collected from harem matings and reared at 28.5˚C in egg water (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$). Embryos were staged in hours post-fertilization (hpf) as previously described (127). To prevent pigmentation in embryos > 1 dpf, egg water was supplemented with phenylthiourea to 0.003%.

Zebrafish embryo synthetic mRNA injections

The full-length WT mouse *GPR56* cDNA clone (openbiosystems clone ID: 3709247) and all mutant derivatives of *GPR56* (H89A, S150A, H381S, C121S+C177S) were linearized with NotI, transcribed using the mMESSAGE mMACHINE® SP6 ULTRA kit (Ambion), combined with phenol-red dye and injected at a final concentration of 50 pg in 2 nl. To control for adverse side-effects resulting from mechanical stress during injection, we also injected zebrafish embryos with an equal volume of phenol-red diluted 1:5 in water.

Whole mount *in situ* hybridization

Whole-mount in situ hybridizations were performed as described previously (128). Briefly, embryos were fixed at 65 hpf in 4% paraformaldehyde at 4˚C overnight followed by dehydration in 100% methanol. Following dehydration, embryos were washed in 0.2% PBS-Tween (PBSTw), permeabilized in proteinase K (20mg/ul diluted 1:1000 in 0.2% PBSTw), and incubated with an *mbp* Digoxygenin-labeled riboprobe (129) overnight at 65˚C in hybridization buffer (50% formamide). Following overnight incubation, embryos were washed to remove formamide, blocked in 2% blocking

medium supplemented with 10% normal sheep serum and 0.2% Triton™, and incubated in primary antibody (Anti-Dig, Fab fragments (1:2000), Product # 11214667001, Roche) overnight in block. Embryos were then washed in Maleic Acid Buffer with 0.2% Triton™, and developed by alkaline phosphatase treatment. After colorimetric development was complete, embryos were post-fixed in 4% paraformaldehyde and stored long-term in 70% glycerol. Embryos were mounted on slides with glycerol and imaged at 10x with an AxioCam MRm on a light microscope (Zeiss AxioImager M2). Imaging was done in a blinded fashion such that the interpreter did not know whether the larvae were injected with WT or mutant mRNA. For further detail, please see previous studies which have employed similar *mbp* quantification protocols (78, 130).

Direct G protein-coupling assay using insect cell membranes

Protocol was implemented as previously described (35), the only difference being that for the present study, insect cell membranes were pre-incubated with G proteins for 5 minutes before starting the assay.

Matrix-assisted laser desorption time of flight (MALDI-TOF) mass spectrometry

Matrix was prepared by first preparing solvent: 50% acetonitrile + 50% (0.1% TFA in H_2O). 1mg sinapinic acid was dissolved in 0.1mL solvent to form saturated solution with precipitate. Samples for MALDI-TOF were prepared by spotting 0.5μL purified protein and 0.5μL matrix on the target plate. Samples were analyzed on a Bruker ultrafleXtreme MALDI-TOF/TOF instrument.

Circular Dichroism Spectroscopy

Purified mouse GPR56 ECR and GAIN domain were diluted to 200 nM and 250 nM, respectively in 100 mM NaCl, 25 mM sodium phosphate, pH 7.5. The sample was placed in a 1 cm path length cuvette and CD signal at λ=220±5 nm was observed as the temperature was increased. All measurements were taken on a Jasco J-715 CD spectrometer.

Methods used in Chapter V[2]

Cloning and purification of GPR56 extracellular domains from insect cells

The following constructs were prepared and cloned into pAcGP67a for expression in insect cells: human GPR56 (UniProt: Q9Y653) full ECR (residues G27-S392), PLL domain (residues G27-S160), and GAIN domain (residues M176-S392); and mouse GPR56 (UniProt: Q8K209) full ECR (residues S27-S392), PLL domain (residues G27-S160), and GAIN domain (residues M176-S392). The C121S and C177S mutations were introduced to the PLL and GAIN domain constructs, respectively, to remove the free cysteine residues that participate in the interdomain disulfide bond in

[2] The text of this section is copied verbatim (with minor changes) from:
Salzman GS, Zhang S, Gupta A, Koide A, Koide S, and Araç D. (2017) Stachel-independent activation and inhibition of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. Proceedings Nat Acad Sci USA: in revision. G.S.S, A.K., S.K., and D.A. designed experiments. G.S.S. cloned GPR56 constructs, purified GPR56 fragments, engineered monobodies, and performed signaling assays. G.S.S and S.Z. purified monobodies. G.S.S. and A.G. performed binding assays. A.K. generated the monobody library. G.S.S., S.K., and D.A. analyzed data and wrote the manuscript. (232)

the full ECR (93). A C-terminal His_6 tag was added for affinity purification. A C-terminal AVI-tag corresponding to the sequence GLNDIFEAQKIEWHE was added to aid biotinylation.

A baculovirus expression system was used for expression of proteins in High Five insect cells as previously described (55). The secreted, glycosylated proteins were purified using nickel-nitrilotriacetic agarose resin (Qiagen) and size-exclusion chromatography (Superdex 200 10/300 GL; GE Healthcare).

Monobody generation

Purified and biotinylated human and mouse full ECR, GAIN domain, and PLL domain were used as a targets for phage-display selection from a 'side and loop' monobody library as previously described (114). The naïve library contained ~10^9 different clones. Three rounds of selection were performed at target concentrations of 1) 100 nM (conjugated to streptavidin beads and thus in the tetravalent form), 2) 100 nM (monomeric), 3) 50 nM (monomeric). In some cases the species was altered for the second round of selection in an attempt to generate human and mouse cross-reactive clones. A yeast display library containing ~10^6 different clones was constructed from the output of phage display selection. Two rounds of positive sorting of the yeast display library was done using fluorescent-activated cell sorting (FACS) using the same GPR56 domains labeled with dye to stain yeast. Binding assays testing the affinity and specificity of individual monobody clones were performed using yeast surface display and M280 beads as described previously (93, 123).

Purification of monobodies from E. coli

The genes encoding the identified monobodies were cloned into an expression vector, pHBT (123). Monobodies were expressed in BL21(DE3) *E. coli* via IPTG induction at 18˚C for 20 hours. Monobodies were purified via an N-terminal 6xHIS tag using nickel-nitrilotriacetic agarose resin (Qiagen). β3 was refolded on the Ni-column using the β-cyclodextrin method (124). β1 and β7 were refolded in solution using the L-Arginine dilution method (131). β12 was purified from the soluble fraction *E. coli* and did not require refolding. (Re)folded proteins were further purified using a Superdex 200 10/300 GL column (GE Healthcare).

SRE-luciferase signaling assay

SRE-luciferase assay was performed as described previously (93) with several alterations. Briefly, HEK293T cells were seeded in 96-well plates (10,000 cells in 0.1 mL DMEM +10% FBS/well). After 12-18 hours, cells reached 40-50% confluence and were transfected with 10 ng *GPR56/Gpr56* (WT or mutant) + 20 ng dualLuc-SRE + 0.3 µL LipoD293 (SignaGen®) per well from a master mix. After 24 hours, media was aspirated and replaced with DMEM without FBS. For monobody treatment, monobody was added to cells 6.5 hours after the start of serum starvation. After 12 hours total of serum starvation, media was aspirated. Cells were lysed using the Dual-Glo® Luciferase Assay System from Promega and firefly and renilla luciferase signals were measured using a Synergy^{TM} HTX luminescence plate reader. Signaling intensity in RLU (fold increase) is reported as: ($Firefly_{GPR56}$/$Renilla_{GPR56}$)/($Firefly_{EV}$/($Renilla_{EV}$).

Flow cytometry

HEK293T cells were transiently transfected with WT or mutant mouse *GPR56* constructs using LipoD293. After 48 hours, cells were detached with citric saline and co-stained with blocked β3-neutravidin-DyLight®488 precomplex and blocked α5-neutravidin-DyLight®650 precomplex in a single 20 min staining reaction in PBS +2% BSA at room temperature. Flow cytometry was performed using Accuri® C6 flow cytometer.

Surface Plasmon Resonance

The biotinylated "ligand" was mixed with neutravidin in a 1:5 molar ratio and immobilized on the biotin chip of the SensiQ. Free biotin was flowed to block remaining biotin binding sites on the neutravidin. The analyte was then flowed and binding response units (RU) was measured over association time and, upon removal of analyte, dissociation time. All experiments were performed at 25˚C in 50mM HEPES pH 7.4 + 150mM NaCl + 0.01% TWEEN20. All curve fitting was done using QDat.

Methods used in Chapter VI

Cloning, expression and purification of TG2 fragments

Full length mouse TG2 was provided as kind gift from Lei Xu (University of Rochester). The full-length TG2 and the TG2 D3D4 constructs (residues T471-A686) were cloned into the vector pVL1393. N-terminal 6xHIS and AVI-tags were added to each construct to facilitate purification and biotinylation, respectively. Baculoviruses were generated for cytosolic insect cell expression as previously described (55, 93). Large-scale Hi5 insect

cell cultures were infected with baculovirus and grown for 48 hours before the cell pellet was harvested by centrifugation and frozen at -80˚C. Cell pellets were thawed and cells were lysed at 4˚C in a manual homogenizer in 10mM HEPES pH 7.2 + 150mM NaCl + 1mM TCEP + 2mM PMSF. The supernatant was collected by centrifugation (37,000 x g for 1 hour) and incubated with Ni-NTA sepharose resin for 3-5 hours at 4˚C with constant stirring. The beads were washed and incubated with purified BirA biotin ligase + biotin + ATP for 1 hour at 27 ˚C in 50mM Bicine pH 8.3 + 150mM NaCl + 10mM Mg acetate. The biotinylated TG2 was then eluted with in the same buffer with the addition of 200mM imidazole. The eluent was filtered and injected into a Superdex 10/300 Gel Filtration column. Peak fractions were pooled and frozen in liquid nitrogen.

Flow Cytometry

HEK293T-expressed GPR56 WT and mutant:

HEK293T cells were transfected with GPR56 constructs (wild-type or mutant) and co-stained with monobody β3+neutravidin-488 tetramers and TG2+neutravidin-650 tetramers. Tetramers were independently prepared in the excess of free biotin and free neutravidn before mixing together to avoid the possibility of forming β3+NAV650 and TG2+NAV488 tetramers. To normalize for differential expression of GPR56 mutants, TG2 binding signal was normalized to a particular bin of β3 binding signal. This way, the TG2 binding signal of high- and low-expressing mutants could be directly compared because, in essence, only equivalently-expressing cells were included in the TG2 binding analysis. Data were collected on an AccuriC6 flow cytometer and processed in FlowJo.

HEK293T-expressed GPR56; TG2 and monobody competition:

HEK293T cells were transfected with GPR56 constructs (wild-type or mutant), incubated with 500-fold molar excess of unlabeled tetramerized monobody competitor (β7 or β12) and then and co-stained with monobody β3+neutravidin-488 tetramers and TG2+neutravidin-650 tetramers. (Monobody competitor was added at 500nM and TG2 was added at 1nM, both tetramerized.) Binding signal was normalized to GPR56 expression as described above. Data were collected on an AccuriC6 flow cytometer and processed in FlowJo.

TG2-coated M280 beads and affinity measurement:

M280 bead-binding assay was carried out as previously described (93). In short, M280 beads were coated with TG2, following which, purified and biotinylated GPR56 fragments were incubated with beads at various concentrations. Neutravidin-650 was then added to detect free the GPR56 fragments. Binding signal was plotted versus GPR56 fragment concentration to calculated the dissociation constant, K_D. Data were collected on an Intellicyt flow cytometer, initially processed in FlowJo, and standard 1-to-1 binding curve-fitting was done in Prism.

Methods used in Chapter VII[3]

Vector and Cloning

For crystallization and protein purification, the LRR repeats of human FLRT3 (residues K29-D357; Uniprot ID: Q9NZU0) was cloned into the XmaI and Not I sites of the pAcGP67a. The olfactomedin domain of human LPHN3 (residues V132-G392; Uniprot ID: Q9HAR2) was cloned into the BamHI and XbaI sites of pAcGP67a. Similarly, the Ig-like domain of mouse Unc5D (residues G49-Q161; Uniprot ID: Q6UXZ4) was cloned into BamHI and NotI sites of pAcGP67a. A His8 tag or biotin tag was added at the C terminus for affinity purification. For mammalian expression and functional analysis, full length human FLRT3 (residues S30-S649), human LPHN3 (residues F20-L1447), and human Unc5B (G27-E934; Uniprot ID: Q8IZJ1), and mouse Unc5D (S46-L884) constructs with preprotrypsin leader sequence containing N-terminal myc, FLAG, His and His-tags, respectively, were cloned into pCMV5 plasmid using Gibson Assembly (NEB). Site directed mutagenesis for FLRT3 and LPHN3 mutants was performed using the Qiagen QuikChange Mutagenesis (QIAGEN).

---

[3] The text of this section is copied verbatim (with minor changes) from:
Lu YC, Nazarko OV, Sando R III, Salzman GS, Südhof TC, and Araç D. (2015) Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion. Structure 23(9):1678–1691. Y.C.L. designed and performed the crystallography, structure determination, mutagenesis, and bio-layer interferometry binding experiments. O.N. designed and performed the HEK cell expression and flow cytometry binding as- says. R.S. designed and performed the cell aggregation assays. G.S. designed and performed the structure determination, differential scanning fluorimetry assay, and designed and assisted with the flow cytometry binding assays. T.C.S. designed the cell-aggregation assays and contributed to the preparation of the manuscript. D.A. designed all the experiments and wrote the paper. (68)

Protein Expression and Purification

Spodoptera frugiperda (Sf9) cells (Life Technologies) were transfected with pAcGP67a carrying the gene and linearized baculovirus DNA (BestBac 2.0, v-cath/chiA deleted; Expression Systems) using Cellfectin (Life Technologies). Baculovirus was amplified in Sf9cells in 10%(v/v) fetal bovine serum containing SF900-II SFM medium (Life Technologies).

Large-scale protein expression was performed by infection of *Tri choplusiani* (Hi5) cells in Insect-Xpress medium (Lonza) at a cell density of 2×10^6cells/ml with an infection course of 72hr. The secreted, glycosylated recombinant proteins were purified using nickel-nitrilotriacetic agarose resin (QIAGEN) and size exclusion chromatography (Superdex 200 10/300 GL; GE) in HBS buffer (10mM HEPES (pH7.2), 150mM NaCl). FLRT was concentrated to 15mg/ml in 10kDa Centricon (Millipore) at 12°C and used for crystallization trials. It was observed that FLRT at high concentration tends to precipitate at 4**°** C, displaying a white cloudy color. However, precipitation was reversible when concentrated FLRT was warmed up to room temperature.

For FLRT/LPHN3 complex crystallization, purified FLRT and LPHN3 were mixed at 1:1.2 ratio. The complex was purified by size-exclusion chromatography in buffer containing 10mM HEPES (pH7.2), 150 mM NaCl and concentrated to 20 mg/ml in 10 kDa Centricon (Millipore) at 12°C.

For BLITZ and some flow cytometry experiments FLRT3 LRR and LPHN3 Olf were cloned into a pACGP67a vector that carries a C-terminal Avi-tag (sequence GLNDIFEAQKIEWHE) followed by a 6XHis tag. The purified Avi-tagged proteins were

biotinylated *in vitro* with purified BirA enzyme that recognizes the Avi-tag and biotinylates it.

Crystallization, Data Collection, and Processing

Prior to crystallization, purified proteins were incubated with carboxypeptidase A (1:100 enzyme to protein ratio; Sigma-Aldrich) and carboxypeptidase B (1:100 enzyme to protein ratio; EMD Millipore) to cleave off the C-terminal residues such as the His8 tag. Initial screens for crystallization of FLRT and FLRT-LPHN3 complex were carried out using 96-well format kits (Qiagen JCSG Core Suites 1-IV; Rigaku Wizard Crystallography Screens) on a Mosquito Crystal robot (TTP Labtech) at room temperature. Crystals of FLRT3 grew in D10 well of JCSG Core III suite (0.1M Tris pH7, 50% (v/v) PEG200) within 10days. Crystals were cryoprotected by transferring the crystals into mother liquor with 35 % glycerol. The best crystals diffracted to d_{min}=2.6 Å.

Similarly, crystallization trials for the FLRT/LPHN3 complex were obtained with 96-well screens. Initial crystal hit was observed in E7 well of Wizard Cubic LCP Block (Rigaku) containing 10% (w/v) PEG 3000, 100 mM MES pH6.0, 200 mM Lithium sulfate. To further optimize, 20 mg/ml protein complex in 10 mM HEPES pH7.2, 150 mM NaCl was mixed with equal volume of mother liquor, and equilibrated against 500 ul mother liquor in a 24-well format. Crystals grew to full size within 4 days and were cryoprotected with 20% glycerol. Many crystals displayed an intergrowth of two separate crystals in a variety of specific configurations, resulting in crystal twinning. The best crystals diffracted to d_{min}=2.6 Å.

Structure Determination and Refinement of FLRT3 and FLRT3/LPHN3 complex

Diffraction data was collected at Advanced Photon Source of the Argonne National Laboratories beamline 23-IBD and 19-BM. Data sets were processed using HKL2000. FLRT3 structure was solved by molecular replacement with Phaser-MR (Phenix) using a model of FLRT structure (PDB ID 4V2E). Similarly, FLRT3/LPHN3 structure was determined by molecular replacement using a homology model of the LPHN3 Olfactomedin domain based on the myocilin olfactomedin domain crystal structure (PDB ID 4WXQ). For both structures, refinement was performed in phenix.refine (Phenix) with nonscrystallographic symmetry (NCS) restraints. Olf structure in the complex was further refined using LPHN3 olf structure (PDB 5AFB). Since FLRT3/LPHN3 complex crystal displayed twinning, twin law of h, -k, -l was employed throughout the entire refinement process. Minor adjustments of the model were performed manually using COOT, followed by another round of refinement in Phenix. The final FLRT3/LPHN3 model contains four N-linked NAG carbohydrate moieties attached to each chain of the FLRT molecules in the FLRT3/LPHN3 complex structure. PISA program in the CCP4 package was used to analyze the structure such as calculation of the interface area and the hydrogen bonding interactions.

Flow cytometry

Cell culture:

HEK293 cell line (ATCC; a generous gift from S. Koide lab) was cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% FBS (Gemini Bio-Products) at 37°C in 5% CO_2 humidified incubator. To seed for transfection cells were

washed with PBS buffer and trypsinized (0.05% Trypsin-EDTA; Invitrogen). Cells were transiently transfected with Fugene 6 (PRE2693, Promega) at 60-70% confluence.

Sample preparation:

HEK293 cells were transfected with 2 µg of DNA /well in 6-well plates using Fugene6. After 48 hrs of incubation, cells were detached with citric saline solution and washed with PBS and PBS+0.1% BSA (Bovine serum albumin, A3803, Sigma). The pellet was then stained with primary antibodies (in PBS+0.1% BSA) for 30 min (with rotation at room temperature), washed twice with PBS+0.1% BSA and incubated with secondary antibodies for another 30 min and washed twice again. Pellets were resuspended in PBS+0.1% BSA. Flow cytometry data were collected on Guava Easycyte flow cytometer (10000 events measured).

Cell labeling:

To test LPHN3 WT and mutants expression, cells were stained with mouse anti-FLAG M2 antibodies, 1:1000 (F3165, “Sigma”). Fluorescence was determined by incubating with anti-mouse FITC, 1:100 (F0257, “Sigma”). For binding assays purified monomeric His-FLRT LRR was added to primary antibodies to final concentration 10 µM ; 100 nM precomplex of BTtrisNTA - NeutrAvidinDyLight 650 (NAV650) (84607, “Thermo Scientific”) was used for fluorescent labeling. Otherwise cells were incubated with tetramerized His-FLRT LRR as a component of 100 nM precomplex HIS-FLRT3 LRR – NAV650 – BTtrisNTA. BTtrisNTA (biotinylated tris-nitrilotriacetic acid) – a reagent that

effectively noncovalently biotinylates any poly-His-tagged protein - was generously provided by S. Koide lab.

To test FLRT3 WT and mutant expression, cells were stained with mouse anti c-Myc antibodies (9E10, “DSHB”) 1:20 and anti-mouse FITC, 1:100. 100 nM pretetramerized biotin-LPHN3 olf - NAV650 was used to test binding. Pretetramerization was done by mixing and incubating the reagents for 30 minutes.

For Unc5 FLRT3 binding assays, cells transfected with Unc5B and Unc5D constructs as well as untransfected cells were stained with 100nM pretetramerized NAV650 – biotin-LPHN3 olf - His-FLRT3 LRR or NAV650 – biotin-LPHN3 olf - HIS-FLRT3-UF LRR.

Flow cytometry analysis:

Sample analysis was performed using FlowJo Single Cell Analysis software and presented as histograms (for protein expression data) and dot plots (an overlay of two fluorescent signals which correspond for protein expression and binding). Also protein expression was presented as diagrams of mean fluorescence intensity for each sample. Gates were set to define a population of cells to show binding.

Bio-layer Interferometry Binding Measurement

Binding between FLRT3 and LPHN3 was performed using the BLITZ system (forteBio). FLRT3 or LPHN3 was then immobilized onto streptavidin sensors, and the unbound protein was washed off by HBS buffer (10mM HEPES pH7.2, 150mM NaCl). The sensor was then immersed into solutions containing immobilized protein’s binding

partner at various concentrations for 200s. Dissociation was carried out by immersion of sensor into HBS for 250s. Since LPHN3 failed to dissociate completely from the sensor, 3M $MgCl_2$ in HBS was used to regenerate the streptavidin sensor for 250s. The regeneration solution (3M MgCl2 in HBS) was tested multiple rounds prior to the binding experiment to ensure it did not cause any undesired effects on the ligand or sensor chip.

Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) assays were performed on a *CFX384*™ (Bio-Rad). Prior to DSF assays, proteins were purified using Gel Filtration chromatography in a buffer containing 10mM HEPES pH 7.2 and 150mM NaCl. After gel filtration, samples were equilibrated to 25˚C. A 1:1000 dilution of SYPRO Orange (Invitrogen Molecular Probes) was used as a reporter dye to monitor the denaturing process of the proteins. Samples were assayed on a 384-well plate with final protein concentrations of 6 μM. The temperature was raised with a step of 0.5 °C per minute from 25 °C to 95 °C, collecting fluorescence readings at the end of each interval. Each sample was run in quadruplicate. Collected data was analyzed by *CFX-Manager*™ from Bio-Rad. The negative temperature derivative of the fluorescence intensity was plotted as a function of temperature. The temperature corresponding to the minimum value of each curve reflects the midpoint of a two-state transition from folded to unfolded protein. This temperature was thus reported as the melting temperature of the protein (T_m).

Cell aggregation assays

FreeStyle HEK293 cells (Life Technologies) grown to a density of $1x10^6$ cells/mL in a 30mL volume were co-transfected with 30µg of either pCMV-Emerald or pCMV-dsRed and 30µg of the indicated construct using FreeStyle Max reagent (Life Technologies). Cells were grown at 37°C/8%CO_2 with shaking at 125rpms. All cDNAs were driven by the CMV promoter. Transfected cells were mixed in a 1:1 ratio two days post-transfection and incubated for an additional 2 days. Live cells were imaged by dropping 100µL of cell suspension onto a glass slide (FisherBrand). Aggregation index was calculated as shown previously (Boucard et al. 2013. J. Cell Biol.).

CHAPTER III

Geometric preferences of engineered protein binding sites[1]

The strong tendency of synthetic proteins binding to a functional site was first observed for the VHH/Nanobodies, and it was rationalized based on the geometric matching between the generally concave surfaces of protein functional sites and the compact prolate shape of the target-recognition surface presented by the VHH scaffold (132). It seems that this mechanism of action explains a number of cases for Monobodies that are structurally similar to VHH/Nanobody and appear to bind to a concave cleft. However, as discussed below, it appears that some Monobodies may also preferentially bind to a functional surface that is not strongly concave. In order to more quantitatively describe the role of geometric matching in determining the binding sites of synthetic proteins, we established a robust and practical tool for quantification of the concavity of protein-protein interaction sites. With this tool, we set out to compare the preferred binding site concavities of several commonly-used synthetic protein scaffolds.

---

[1] Text in this chapter was taken verbatim (with minor changes) from:
Sha F, Salzman G, Gupta A, and Koide S. (2017) Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci* 26(5):910–924.
My contributions to this review article included writing and editing the text and performing the monobody concavity analysis shown in Figure 3 of the review and discussed in this chapter. (233)

Method

Starting from a crystal structure of a protein-protein complex (e.g. monobody + antigen), the epitope (monobody-contacting atoms on the antigen) and paratope (antigen-contacting atoms on the monobody) are identified. The epitope is defined as any atoms on the antigen within 4.5Å of any atom on the monobody. The paratope is defined as any atoms on the monobody within 4.5Å of any atom on the antigen. A spherical shell was fit to the paratope, with four free parameters: the (x,y,z) coordinates of the center of the sphere (Cs) and the radius of the sphere (Rs). The inverse of the radius of the sphere is defined as the curvature.

Next, the direction of the interface (e.g. concave vs. convex paratope) was determined. The center point of the paratope (Cp, defined as the average position of all the points in the paratope) and the center point of the interface (Ci, defined as the average position of all the points in the epitope or paratope) were calculated. The vector from Ci to Cp was projected onto the vector from Cs to Ci. The sign of this projection corresponds to the direction of the interface: a positive or negative projection would correspond to a convex or concave paratope, respectively. The curvature was correspondingly assigned a negative sign to represent a concave paratope.

Results

A Monobody, YSX1, derived from a loop library bound to a concave surface around the sugar-binding cleft of maltose-binding protein (Figure I.5, Figure III.1A) (133). Similarly, Monobody HA4 bound to the peptide-binding cleft of the Abl SH2 domain (115). Although this epitope is convex, the observed binding mode can be

rationalized by the fact that the Monobody mimics the natural peptide ligand. In contrast, Monobody AS25 derived from the side library bound to a convex surface on the opposite side of the SH2 domain that is used for intramolecular interaction with the kinase domain of Abl (Figure III.1C) (116). This surface does not have a cleft for peptide binding, and the Monobody does not mimic the binding mode of the kinase domain. In another example, Monobody NS1, also generated in an unbiased manner from the side library, was bound to a nearly flat surface of H-RAS that is involved in dimerization (Figure III.1B) (134). These cases clearly show that the preference toward a functional site is not only due to the geometric matching between a functional cleft and a small globular binding protein. However, we note that the geometric matching is an important factor in the ability of these Monobodies presenting a concave binding surface to bind to a convex or flat surface of their target.

Conclusions

Then, what is the molecular basis of the strong preference of these small binding proteins toward a functional surface? Although the paucity of binding proteins directed to a clearly nonfunctional surface makes it impossible to elucidate the basis, we speculate that the key is the surface characteristic inherent to natural proteins. It is well established that functional surfaces of natural proteins are enriched with amino acids that are conducive to forming interactions such as Tyr, Trp and Arg, whereas nonfunctional surfaces contain higher fractions of amino acids that tend to break interactions such as Glu and Lys (98, 135). Because synthetic binding proteins are generated in a short period under strong selection pressure for high affinity, it is not

difficult to imagine that this approach should enrich clones that bind to surfaces that are more conducive to forming high-affinity interactions, as opposed to other surfaces that have not evolved to interact with other molecules. Although Nanobodies are not fully synthetic, they are also generated in a short period under strong selection pressure of animal immunization and phage-display selection. Thus, the Nanobody generation processes should also enrich those clones that bind to target surfaces conducive to forming interactions. Therefore, although shape complementarity is an important factor in epitope selection, the dominant determinant appears to be the surface chemical properties of natural proteins. The strong preference of Nanobodies toward concave surfaces may be due to the fact that the natural immune repertoires of Nanobodies produce mostly convex antigen-binding site (Figure III.1D). This notion in turn suggests the exciting possibility of controlling virtually all types of protein functions by utilizing synthetic binding proteins capable of presenting a target-binding site with diverse topography.

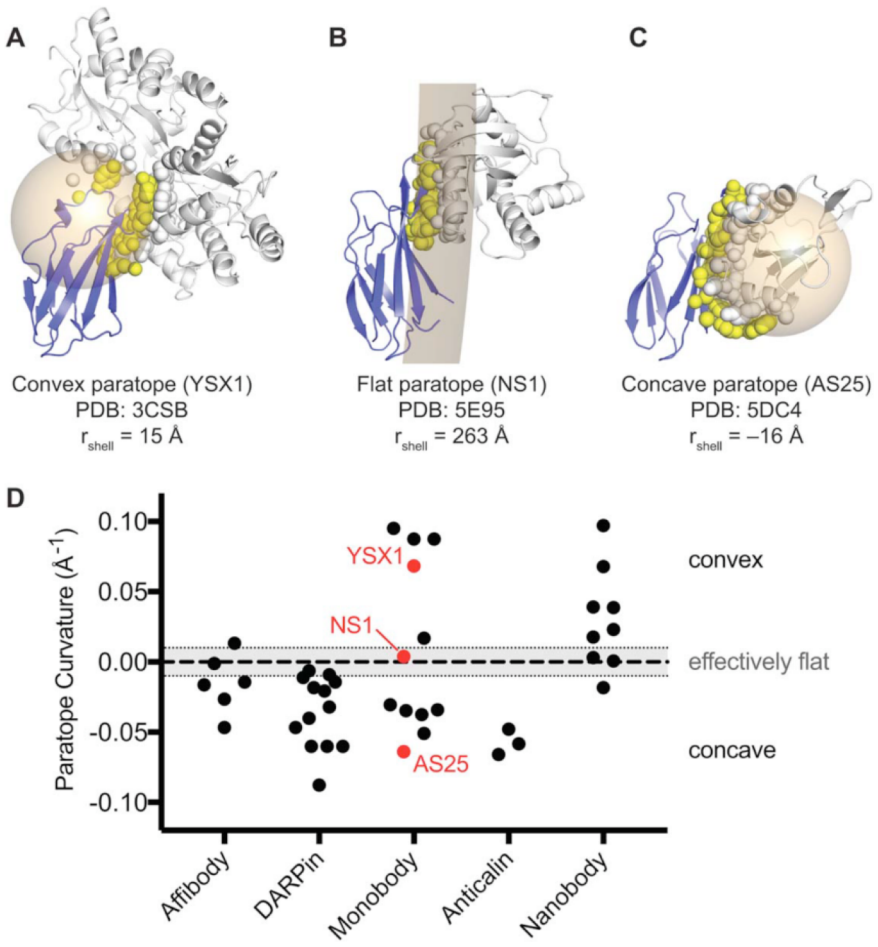


Figure III.1 Concavity analysis of binding protein-target interfaces.

(A–C) Three representative structures of Monobody-target complexes with different levels of concavity. For each crystal structure, a spherical shell (tan) was fit to all the atoms that compose the target-contacting residues on the Monobody (i.e., the paratope; yellow spheres). A spherical shell with a large radius approximates a flat interaction. Spherical shells with smaller radii, centered within the Monobody or target represent convex or concave paratopes, respectively. To distinguish between the two orientations, the radii of shells corresponding to concave paratopes were assigned negative values. Monobody and target structures are shown as blue and gray cartoons, respectively. Atoms composing the Monobody-contacting residues on the target (i.e., the epitope) are shown as gray spheres. (D) Concavity analysis on 34 synthetic binding protein-target complex structures from the PDB. Nanobody complexes are also included for comparison. Curvature is defined as the inverse of the radius of the spherical shell as described above. An arbitrary threshold of $|r_{shell}| \geq 100 A^{\circ}$ ($|curvature| \leq 0.01\ A^{\circ -1}$) was defined as an effectively flat interface.

CHAPTER IV

Structural basis for regulation of GPR56/ADGRG1 by its alternatively spliced extracellular domains[1]

Summary

Adhesion G-protein-coupled receptors (aGPCRs) play critical roles in diverse neurobiological processes including brain development, synaptogenesis, and myelination. aGPCRs have large alternatively spliced extracellular regions (ECRs) that likely mediate intercellular signaling; however, the precise roles of ECRs remain unclear. The aGPCR GPR56/ADGRG1 regulates both oligodendrocyte and cortical development. Accordingly, human *GPR56* mutations cause myelination defects and brain malformations. Here, we determined the crystal structure of the GPR56 ECR, the first structure of any complete aGPCR ECR, in complex with an inverse-agonist monobody, revealing a GPCR-Autoproteolysis-Inducing domain and a previously unidentified domain that we term Pentraxin/Laminin/neurexin/sex-hormone-binding-

[1] The text of this section is copied verbatim (with minor changes) from: Salzman GS, Ackerman SD, Ding C, Koide A, Leon K, Luo R, Stoveken HM, Fernandez CG, Tall GG, Piao X, Monk KR, Koide S, and Araç D. (2016) Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. Neuron 91(6):1292–1304. G.S.S., S.K., and D.A. designed all experiments, interpreted results, and wrote the manuscript. G.S.S. performed all experiments involving aGPCR ECR cloning, purification, MALLS, and crystallography (with assistance from C.D.); monobody engineering, purification, and characterization (with guidance from A.K.); mammalian cell expression; SRE signaling (with assistance from C.G.F.); bioinformatics; and mass spectrometry. A.K. and S.K. provided the monobody libraries; S.D.A. and K.R.M. designed and performed zebrafish experiments. R.L. and X.P. designed and performed RT-PCR experiments. H.M.S. and G.G.T. designed and performed direct G protein coupling experiments (with assistance from K.L.). All authors discussed the results and provided comments and revisions on the manuscript. (93)

globulin-Like (PLL). Strikingly, PLL domain deletion caused increased signaling and characterizes a GPR56 splice variant. Finally, we show that an evolutionarily conserved residue in the PLL domain is critical for oligodendrocyte development *in vivo*. Thus, our results suggest that the GPR56 ECR has unique and multifaceted regulatory functions, providing novel insights into aGPCR roles in neurobiology.

Introduction

Brain development requires precise coordination of numerous key processes that are individually complex. For instance, to ensure rapid action potential propagation in the vertebrate nervous system, many axons must be insulated by myelin, a multilamellar lipid-rich membrane (136). In the central nervous system (CNS), oligodendrocytes govern myelination by extending and iteratively wrapping their plasma membranes around axon segments (137). Loss of myelin leads to severe neurological disorders such as multiple sclerosis (138, 139). The interplay between the control of myelination and cortical development is poorly understood.

Genetic studies have revealed that GPR56/ADGRG1, a cell-surface G protein-coupled receptor (GPCR), plays important roles in both oligodendrocyte and cortex development, potentially providing a molecular link between these processes (70, 73, 74, 76–78). Mutations in *GPR56* cause a human brain malformation called bilateral frontoparietal polymicrogyria (BFPP) that is characterized by disorganized cortex lamination and patterning, especially in the frontal cortex, a region responsible for many human-specific functions (70). In addition to cortex malformation, the brains of BFPP patients exhibit myelination abnormalities, such as a reduced white matter volume,

indicative of myelinated axon defects (70, 140). Furthermore, recent studies have revealed that GPR56 has a critical role in the regulation of oligodendrocyte development in both zebrafish (78) and mouse (77). Altogether, these studies have established GPR56 as a key molecule with multiple functions in CNS development.

GPR56 belongs to the adhesion G protein-coupled receptor (aGPCR) family, a large family of chimeric proteins that have both adhesion and signaling functions (45, 141). aGPCRs are cell-surface molecules that are believed to mediate intercellular communication via cell-cell and cell-matrix interactions. Many aGPCRs have critical roles in nervous system function including peripheral nervous system myelination by Schwann cells (142, 143), CNS angiogenesis (144, 145), and excitatory synapse formation (61). As in the canonical GPCR families, aGPCRs have a seven-pass transmembrane helix bundle (7TM) that, for many aGPCRs, can be activated to initiate a signaling cascade via interactions with cytosolic G proteins. Unlike the canonical GPCR families, aGPCRs also have large and diverse extracellular regions (ECRs), mainly composed of domains generally involved in adhesion-related functions (45). Although this architecture suggests ECRs have functional importance, their biological roles are incompletely understood.

aGPCRs are characterized by the presence of an extracellular GPCR-Autoproteolysis-INducing (GAIN) domain located immediately N-terminal to the 7TM (55). During aGPCR maturation, autoproteolysis occurs within the GAIN domain (146), cleaving the receptor into two fragments: (1) an N-terminal fragment (NTF) comprising various extracellular adhesion domains and the majority of the GAIN domain; (2) a membrane-bound C-terminal fragment (CTF) comprising the C-terminal β-strand of the

GAIN domain, termed the '*Stachel* peptide' (147) (also called 'tethered agonist' (35), or 'stalk' (91)), the 7TM, and the intracellular region (Figure IV.1A). After autoproteolysis, the NTF and CTF remain associated to form the mature, plasma membrane-localized receptor (55, 90). To date, two non-mutually exclusive models have been proposed for ECR-regulated aGPCR activation. According to the 'shedding' model, ligand binding to the adhesion domains in the ECR may induce dissociation of the NTF from the membrane-anchored CTF, termed 'shedding'. After shedding, the *Stachel* peptide on the CTF is freed from the GAIN domain and functions as a tethered agonist to activate the 7TM (35, 130, 141, 147–150). An alternative model suggests that aGPCR extracellular domains govern receptor activity by directly interacting with the 7TM, perhaps in a ligand-dependent fashion (90, 91). However, without detailed characterization of aGPCR ECRs, the validities of these models have been difficult to verify.

Human and mouse GPR56 are 693 and 687 residues long, respectively, each including a ~377-residue ECR. The localization of six BFPP mutations to the GPR56 ECR suggests a critical role for the ECR in GPR56 function (151, 152). The ECR comprises an N-terminal domain with no previously defined motifs and a GAIN domain (Figure IV.1A) (55). Truncation-based studies suggest that the N-terminal domain mediates the interaction of GPR56 with two known natural extracellular ligands, collagen III and tissue transglutaminase (TG2), and that these interactions may regulate GPR56 function (75, 82). However, sequence-based bioinformatics analyses have

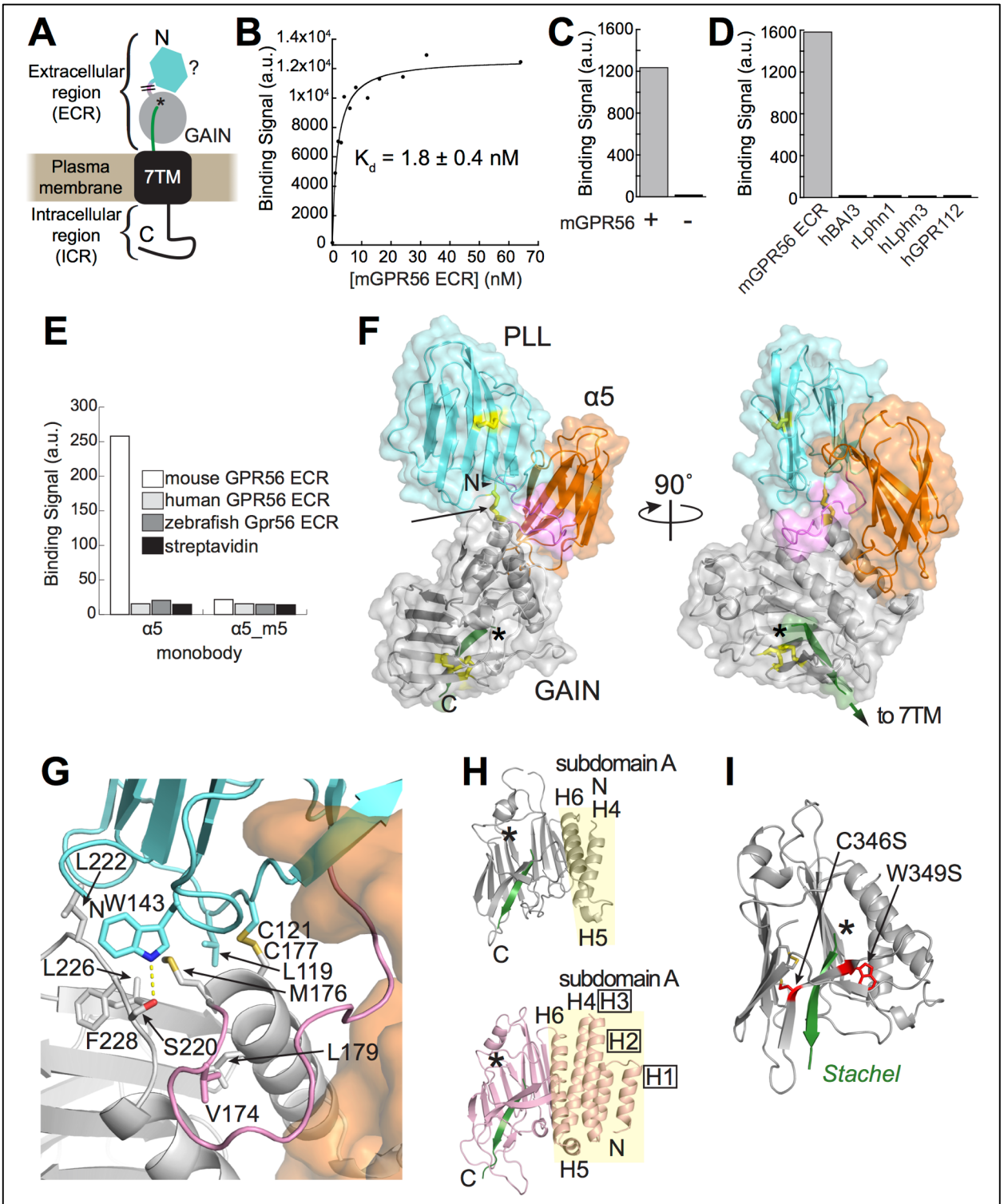


Figure IV.1 Crystal structure of the GPR56 extracellular region in complex with high-affinity and specific monobody.

(legend on the following page)

Figure IV.1, continued.

(A) Schematic of predicted GPR56 domain structure including ECR composed of unidentified N-terminal domain (cyan), linker (pink), and cleaved GAIN domain (NTF, gray; *Stachel*, green; autoproteolysis site, *). Though the terms 'Extracellular Domain' or 'Ectodomain' (both abbreviated ECD) are conventional, we have chosen to refer to the extracellular part of GPR56 as 'extracellular region' (ECR) to avoid confusion given that the ECR is composed to two protein domains (236). \\ represents unclear domain/linker boundary. (B) Binding titration of purified mouse GPR56 ECR to yeast-displayed monobody α5. Bound GPR56 was quantified using flow cytometry. (C) Binding signal of purified α5 (25 nM) to HEK293T cells overexpressing full-length mouse GPR56 (+) and control cells (-) detected by flow cytometry. (D) Binding signals of different purified aGPCR extracellular fragments at 250 nM (BAI3, ADGRB3; Lphn1, ADGRL1; Lphn3, ADRGL3; GPR112, ADGRG4; m, mouse; h, human; r, rat) to yeast-displayed α5. (E) Binding signal of α5 and α5_m5 to purified GPR56 ECR-coated M280 beads (see Figure S2B). (F) The crystal structure of GPR56 ECR in complex with α5 (orange). Cys residues involved in a disulfide bond are colored yellow, with the interdomain disulfide bond (C121-C177) indicated by the arrow. The linker and *Stachel* are colored pink and green, respectively and the asterisk indicates the autoproteolysis site. (G) Close-up view of the binding interface between the PLL domain, PLL-GAIN linker, and GAIN domain. Residues at the binding interface are shown as sticks. The PLL domain, PLL-GAIN linker, and GAIN domain are colored cyan, pink, and gray, respectively. α5 is shown as a transparent orange surface. Polar contacts are indicated by yellow dashes. (H) Crystal structures of autoproteolyzed GAIN domains of GPR56 (top) and Lphn1 (PDB: 4DLQ; bottom) in identical orientations. The α-helices in subdomain A (yellow background) are labeled, and the boxed labels indicate α-helices present in Lphn1 but not GPR56. (I) Human disease-causing GPR56 mutations (red) mapped to the GAIN domain.

failed to determine the identity of the N-terminal domain as well as the domain boundaries of the N-terminal domain and GAIN domain.

Alternative splicing (AS) has been observed in the coding and non-coding regions of *GPR56* transcripts (76, 153). AS in non-coding upstream elements of *GPR56* regulates human-specific cerebral cortical patterning, leading to the suggestion that rapid evolution of *GPR56* AS might have influenced cortex evolution of gyrencephalic brains, such as the human brain (76). Intriguingly, the coding region of *GPR56* also

undergoes AS to generate four variants in humans, two of which include large deletions in the ECR. (153). Considering the established pathophysiological importance of AS in the coding regions of many proteins, especially in the brain (154–156), it is likely that these four *GPR56* splice variants have distinct and important roles. However, due in large part to the absence of three-dimensional structural information, the effect of AS on ECR architecture is unknown and isoform-specific roles of GPR56 remain elusive.

In this study, we set out to determine the 3D structure of the entire ECR of GPR56 at atomic resolution. To this end, we engineered monobodies that recognize the ECR of GPR56. Monobodies are synthetic binding proteins based on the human fibronectin type-III (FN3) scaffold (108), which have recently emerged as powerful tools to facilitate structure determination as 'crystallization chaperones.' Monobodies can also act as agonists or antagonists, further underscoring their utility in probing the function of a given protein (115, 121, 123). In this study, we determined the crystal structure of the ECR of GPR56 in complex with a monobody; excitingly, this represents the first crystal structure of a full ECR for any aGPCR. The structure revealed the identity and boundaries of two extracellular domains: a previously unidentified N-terminal domain with low homology to all known folds, and a short but functional GAIN domain. Notably, we discovered that the entire newly defined N-terminal domain was deleted in GPR56 splice variant 4 (S4), but not the other variants, and deletion of this domain increased basal activity of the receptor. Finally, we identified a highly conserved, surface-exposed patch on the N-terminal domain, mutation of which abolished GPR56 function *in vivo*. Together these results elucidate the multifaceted manner by which the ECR regulates

GPR56 function and broadens our understanding of aGPCR biology and oligodendrocyte development.

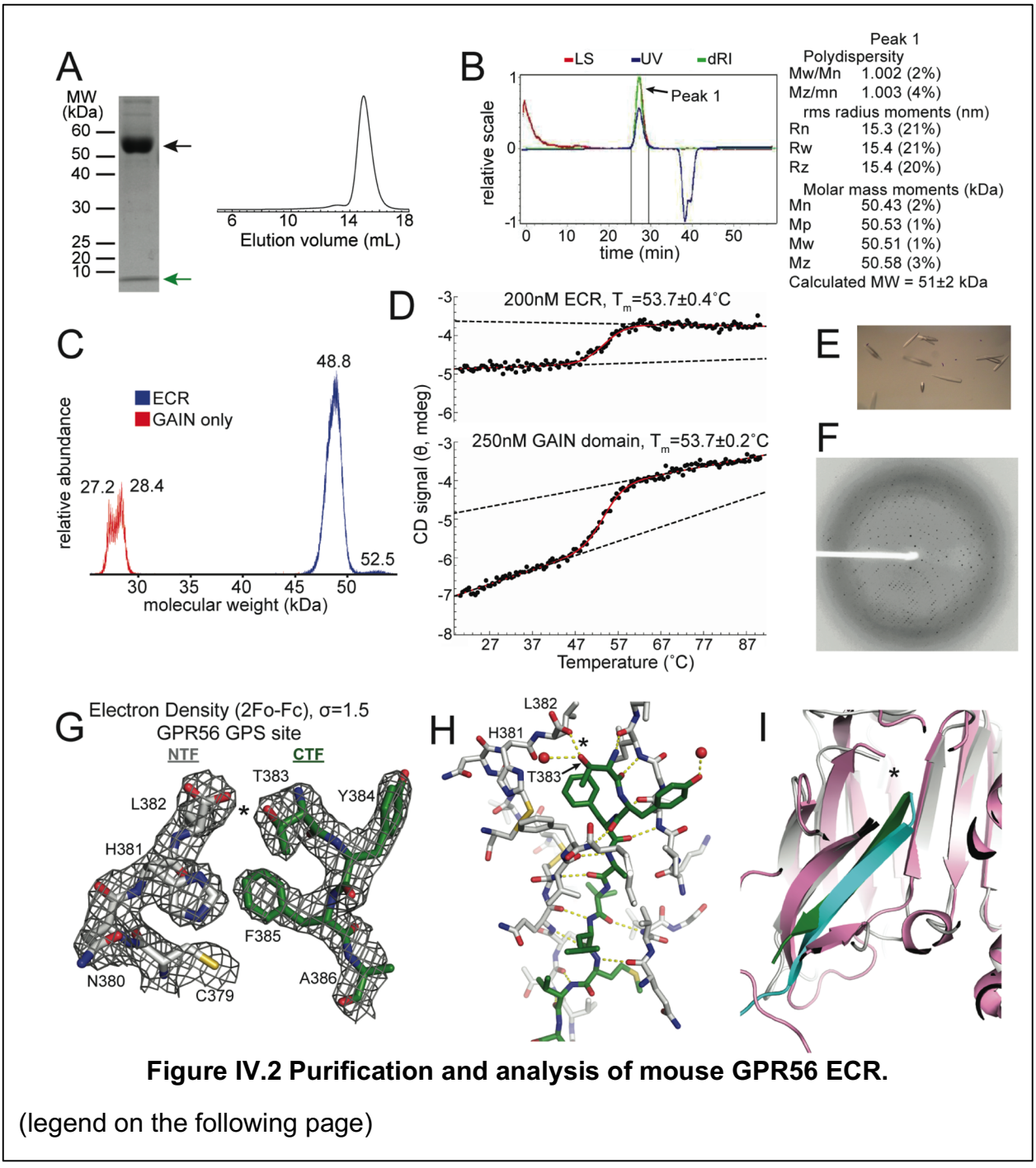


Figure IV.2 Purification and analysis of mouse GPR56 ECR.

(legend on the following page)

Results

A specific and high-affinity monobody directed to the extracellular region of GPR56

Using a baculovirus expression system as previously described (55), recombinant mouse GPR56 ECR was purified from High Five insect cells. The protein was folded, monomeric, and properly underwent autoproteolysis within the GAIN domain (Figure IV.2A-D). From combinatorial phage-display libraries (114, 115), monobody clones that bound to mouse GPR56 ECR were enriched. After gene shuffling and additional library sorting using yeast surface display (114), a total of 13 monobodies with different degrees of affinity for the GPR56 ECR were identified (Figure IV.3A). The clone with the highest affinity, termed Mb(mGPR56_α5), was chosen for further analyses (Figure IV.3A). This clone will be abbreviated as 'α5' hereafter.

Figure IV.2, continued

(A) (Left) SDS-PAGE analysis of mouse GPR56 ECR purified from insect cells. N-terminal fragment (NTF, black arrow) and *Stachel* (green arrow) are observed. (Right) Gel filtration chromatography of cleaved GPR56 ECR. (B) Multi-Angle Laser Light Scattering (MALLS) analysis of purified GPR56 ECR and calculated molecular weight, which is consistent with purely monomeric glycosylated GPR56 ECR and inconsistent with any higher-order oligomer. (C) MALDI-TOF mass spectrometry of purified GPR56 ECR (blue) and GAIN domain only (red). Calculated molecular weights are displayed above peaks. Calculated molecular weights are consistent with cleaved NTFs (with inhomogeneous glycosylation). (D) Circular Dichroism (CD) spectroscopy temperature melts of mouse GPR56 ECR (top) and GAIN domain only (bottom). (E) Representative image of mouse GPR56 ECR-monobody α5 complex crystals used for x-ray data collection. (F) Representative diffraction pattern for native data set. (G) Representative electron density of GPR56 ECR crystal structure showing lack of density at GPS cleavage site, indicating complete autoproteolysis. (H) Interactions between the GPR56 *Stachel* and the GAIN domain NTF. Water molecules shown as red spheres. (F) Orientation of cleaved *Stachel* in GAIN domain of GPR56 (NTF, gray; *Stachel*, green) and Lphn1 (NTF, pink; *Stachel*, cyan). * indicates autoproteolysis site.

α5 bound mouse GPR56 ECR purified from insect cells with an apparent dissociation constant (K_D)=1.8±0.4 nM in the yeast surface display format, as expected from the design of monobody selection (Figure IV.1B, Figure IV.3B). Importantly, the apparent K_D values of monobodies determined in this manner are consistent with those from more conventional biophysical measurements of purified monobodies such as surface plasmon resonance (114, 123). Purified α5 also bound to full-length mouse GPR56 expressed on the surface of HEK293 cells with apparent K_D=17±2 nM, indicating that this monobody recognizes the ECR in the context of full-length GPR56 (Figure IV.1, Figure IV.3B-C). To determine the α5 residues responsible for interacting with GPR56, mutations were made in several regions of α5. Variants harboring mutations in the so-called CD or FG variable loop of α5 (termed α5_m2, α5_m4 or α5_m5; Figure IV.3D) independently decreased affinity by >100-fold, suggesting that both of these loops interact with GPR56. To assess the specificity of α5, we tested its ability to bind to other aGPCR extracellular fragments that contain GAIN domains. No binding was detected between α5 and any of these fragments including GPR112/ADGRG4, an aGPCR in the same subfamily as GPR56 (Figure IV.1D). Furthermore, no binding was detected between α5 and human or zebrafish GPR56 ECR (Figure IV.1E). Finally, differential scanning fluorimetry showed α5 increased the thermostability of the GPR56 ECR (Figure IV.3E). Together, these results show that α5 has high affinity and specificity for mouse GPR56 ECR.

A

Monobody name	Sequence	Subjective Affinity From Single Point Measurement	Preliminary Kd (Mb on yeast)
α1	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDHYVITYGETGASYWSVQEFEVPGSKSTATISGLKPGVDYTITVYAYSWWV------PSSPISINYRT	Medium	10±4nM
α3	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDYYVITYGETGGTA-GGQEFTVPGSKSTATISGLKPGVDYTITVYAYSDDDGVWWPGYYSPISINYRT	Medium	Not determined
α4	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDYYVITYGETGGYA-GGQEFTVPGSKSTATISGLKPGVDYTITVYAYSDDDGVWWPGYYSPISINYRT	Low	Not determined
α5	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDHYVITYGETGGSPWSWQEFEVPGSKSTATISGLKPGVDYTITVYASSFDWTIFPNYYSSPISINYRT	High	1±0.2nM
α6	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDYYVITYGETGGNS-PVQEFTVPGSKSTATISGLKPGVDYTITVYAYTSYYGVWSPAYYSPISINYRT	Medium	Not determined
α7	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDHYVITYGETGASYWSVQEFEVPGSKSTATISGLKPGVDYTITVYADYNPWFSQYP-YYSPISINYRT	Medium	Not determined
α10	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDHYLITYGETGGNS-PVQEFTVPGSKSTATISGLKPGVDYTITVYAYYYSWGYYYP--YSPISINYRT	Low	Not determined
α11	VSSVPTKLEVVAATPTSLLISWDASEAFYVSYYRITYGETGGNS-PVQEFTVPGSYSTATISGLKPGVDYTITVYAYDTW---WMEEDWSPISINYRT	Medium	Not determined
α12	VSSVPTKLEVVAATPTSLLISWDA-YYWYVSYYRITYGETGGNS-PVQEFTVPGSYSTATISGLKPGVDYTITVYAYEQWADYYYE-ASSPISINYRT	Medium	Not determined
α13	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDYYLITYGETGGNS-PVQEFTVPGSKSTATISGLKPGVDYTITVYAESEYTNFWYGYAYSPISINYRT	Medium	Not determined
α14	VSSVPTKLEVVAATPTSLLISWDASEAFYVSYYRITYGETGGNS-PVQEFTVPGSYSTATISGLKPGVDYTITVYAYEQWADYYYE-ASSPISINYRT	Low	Not determined
α15	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDYYFITYGETGGNS-PVQEFEVPGSKSTATISGLKPGVDYTITVYAYS----WGFMSSSPISINYRT	Medium	40±7nM
α16	VSSVPTKLEVVAATPTSLLISWDA-PAVTVDYYVITYGETGGNS-PVQEFTVPGSKSTATISGLKPGVDYTITVYAYYYSWGYYYP--YSPISINYRT	Low	Not determined

Figure IV.3 Monobody generation and characterization.

(continued on the following page)

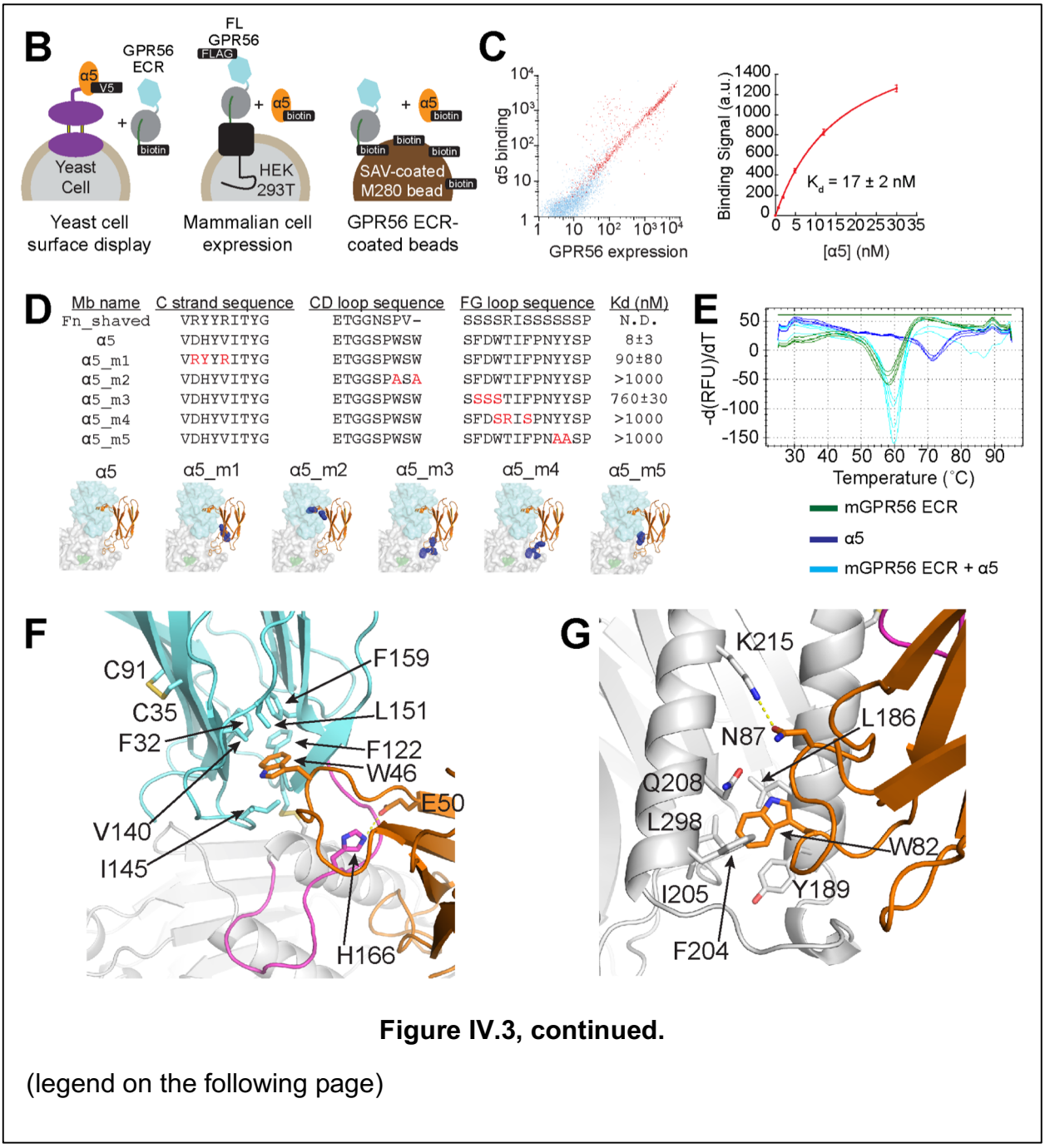


Figure IV.3, continued.

(legend on the following page)

Figure IV.3, continued.

(A) Sequences of 13 monobodies selected with affinity for mouse GPR56 ECR. Highest affinity clone, α5 (yellow), used for subsequent experiments. Affinities evaluated by yeast surface display. (B) Schematic of monobody yeast surface display, mammalian cell expression of full-length GPR56, and purified GPR56 ECR-coated M280 beads used for binding affinity measurements. Note: after coating with GPR56 ECR and before α5 binding, M280 beads were blocked with excess biotin. (C) Flow cytometry of HEK293 cells stained with α5. Untransfected and full-length-FLAG-GPR56-transfected cells shown in blue and red, respectively. (Left) Dot plot showing correlation between GPR56 expression (anti-FLAG staining) and α5 binding. (Right) Concentration titration of α5 to measure binding affinity. (D) (Top) α5 point mutations result in decreased affinity for GPR56 ECR by M280 bead assay. (Bottom) Location of α5 point mutaitons m1-m5 on the crystal structure. Mutated residues shown as blue spheres. (E) Differential scanning fluorimetry temperature melts of mGPR56 ECR (T_m = 58.13 ± 0.13˚C), α5 (T_m = 71.38 ± 0.13˚C), and mGPR56 ECR-α5 complex (T_m = 59.75 ± 0.14˚C). Quadriplucate measurements for each sample are plotted as negative temperature derivative of RFU vs temperature. Peak corresponds to T_m. (F-G) Close-up views of the binding interfaces between α5 and (E) the PLL domain and PLL-GAIN linker and (F) the GAIN domain. Residues at the binding interface are shown as sticks. The PLL domain, PLL-GAIN linker, GAIN domain, and α5 are colored cyan, magenta, gray, and orange, respectively. Polar contacts are indicated by yellow dashes.

The structure of the α5-GPR56 ECR complex reveals two domains with an interdomain disulfide bond

We first attempted to crystallize the GPR56 ECR alone using standard techniques but obtained only crystals that diffracted poorly (>8Å resolution; data not shown). Thus, we instead used α5 as a crystallization chaperone, which yielded high-quality crystals (Figure IV.2E-F). We determined the structure of the GPR56 ECR-α5 complex at 2.5 Å resolution with experimental phases obtained from iodine single wavelength anomalous diffraction data (Figure IV.1F and Table IV.1).

Protein	GPR56 ECR-α5 Native Dataset	GPR56 ECR-α5 Iodide SAD
Data Collection		
Integration Package	HKL2000	HKL2000
Wavelength	1.033 Å	1.771 Å
Space group	P 6_5	P 6_5
Cell dimensions		
a, *b*, *c*	120.34 Å, 120.34 Å, 72.85 Å	121.23 Å, 121.23 Å, 72.67 Å
α, β, γ	90°, 120°, 120°	90°, 120°, 120°
Resolution	46.39–2.45 Å (2.49–2.45 Å)	46.58–3.00 Å (3.05–3.00 Å)
R_{sym} or R_{merge}	0.035 (0.421)	0.057 (0.639)
$CC_{1/2}$	0.923 (0.663)	0.934(0.691)
$I/\sigma I$	19.75 (1.5)	13.9 (1.4)
Completeness	99.9% (99.5%)	89.2% (80.3%)
Redundancy	5.7 (5.7)	11.2 (10.9)
Number measured reflections	631,603	451,228
Number unique reflections	22,286	12,201
Refinement Statistics		
R_{work}/R_{free}	0.215/0.263	
Number of atoms		
Protein	3576	
Water	24	
Other	98	
Average B-factors		
Protein	79.5 $Å^2$	
Water	60.0 $Å^2$	
Other	101.8 $Å^2$	
Rmsds		
Bond lengths	0.003 Å	
Bond angles	0.611°	
Ramachandran plot statistics		
Most favorable	96.0%	
Allowed	4.0%	
Disallowed	0.0%	

Table IV.1 Data collection and refinement statistics.

This structure revealed two domains in the GPR56 ECR with an overall dimension of 84Å x 54Å x 42Å: a previously unidentified domain with a β-sandwich architecture at the N-terminus (P28-S160) and, as predicted from the sequence, a GAIN domain at the C-terminus (M176-S391). The 15-residue linker between the two GPR56 domains is ordered in the crystal, despite its lack of defined secondary structure. We found an interdomain disulfide bond linking the two domains formed by cysteine residues C121 and C177, both of which are highly conserved among GPR56 orthologs

(Table IV.2). This disulfide bond may restrict the movements of the two domains with respect to each other as observed in other proteins with an interdomain disulfide bond (157). We observe an interface between the N-terminal domain and GAIN composed of mostly conserved and hydrophobic residues (L119, W143, M176, and F228). At this interface, we observe a buried surface area of ~680Å^2 (~300Å^2 between the two domains and an additional ~380Å^2 contributed by the linker; Figure IV.1G). It is likely that the interdomain disulfide bond and the hydrophobic residues are sufficient to stabilize this conformation of the ECR.

α5 interacts with the N-terminal domain and GAIN domain simultaneously via its CD and FG variable loops, respectively, which are located at opposite ends of the scaffold (Figure IV.3F-G). This is consistent with the α5 mutagenesis data (Figure IV.3D). At the GPR56 ECR-α5 interface, we observe a buried surface area of ~1620Å^2 (N-terminal domain-α5, linker-α5, and GAIN-α5 contributing ~260Å^2, ~300Å^2, and ~1060Å^2, respectively). The fact that we were able to readily generate a monobody with high affinity suggests that α5 binds to a highly populated conformational species, rather than a high-energy, rare species that would require an extraordinarily high-affinity monobody to capture (117). Thus, it is likely that the conformation of GPR56 ECR captured in our crystal structure represents a dominant conformational state.

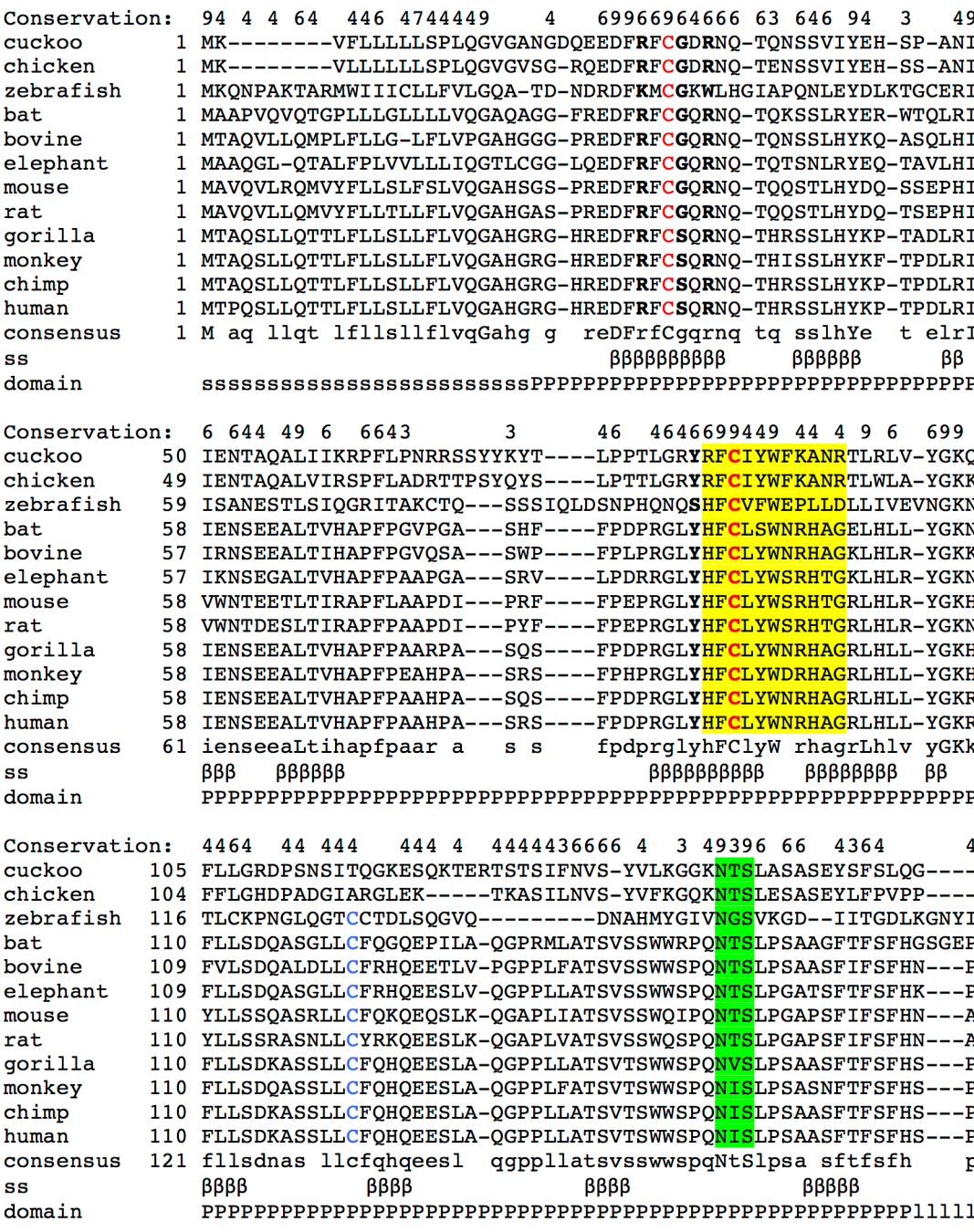


Table IV.2 Multiple sequence alignment of full-length GPR56 from 14 species.

(page 1 of 5; continued on the following pages)

```
Conservation:          66        64443 4  4   4  6   34           4  67  43
cuckoo     159 ----NYFSEIKESMTVWEQDVEKHLNTLDSLIAQPLPPAAGAKEQQMLRRELGKLENILA
chicken    153 -----------EMMPVWKQDVEDELAALDGLLARVP--SAGSTAQRSLRHKLRALERTLA
zebrafish  165 DGAHINCK-------EKFCDEARLKPRGANMIEEVVMRFNAKGRVDLPCAQGTVIE-MDE
bat        169 -KVPEEGESVIQGALREICG-KGAVRVEG-AKSHPEAA--------GKAQVKQRLPPERP
bovine     165 -HKASHNVSV------DVCELKRDLQWLSQFLKHPRK-TSRRPPFTSIGQQLQSLESKLA
elephant   165 -QKASHNASV------DICELKRDLELLSKLLKHPRK-SSRRPSSTPARQHLQSLESKLT
mouse      166 -HKVSHNASV------DMCDLKKELQQLSRYLQHPQK-AAKRPTAAFISQQLQSLESKLT
rat        166 -HKVSHNASV------NMCDLKKELQLLSKFLQHPHK-ASKRPSAAFISQQLQNLESKLT
gorilla    166 -HTAAHNASV------DMCELKRDLQLLSQFLKHPQK-ASRRPSAAPARQQLQSLESKLT
monkey     166 -HTAAHNASV------DMCELKRDLQLLSQFLKHPQK-ASRRPSATPASQQLQSLESKLT
chimp      166 -HTAAHNASV------DMCELKRDLQLLSQFLKHPQK-ASRRPSAAPASQQLQSLESKLT
human      166 -HTAAHNASV------DMCELKRDLQLLSQFLKHPQK-ASRRPSAAPASQQLQSLESKLT
consensus  181  h a hnasv      dmcelkrdlqllsnflkhp k asrrpsa    qqlqslesklt
ss                              aaaaaaaaaaaaaaaaa aaa       aaaaaaaaaaaaaaaaa
domain         lllllllllllllllllGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

Conservation:   4 3  4433 44   4 74664 364 4 4 49 446   664 466  4674696366
cuckoo     215 KVEF-EGQNQTFGKATVHATVLRVQPTQAPQ-HLAFASQ-R-EKGREVHGFAVDLPSSLF
chicken    200 QVEL-EEQNQTFGKAALRATVLSISPTQAPQ-HLAFASP-K-EENREVQGFAVDLPSSLF
zebrafish  217 EFTGHNF-TVPAPRF-VDAN-------------------------T--IPSVYIPSSLR
bat        218 CVRWCLGWAWQHPGKCVRGRGWRGRRGRGVCLWLTLRTLCLQEGPGEILEYSVLLPRALF
bovine     217 SVNF-TGDAVSFEEERINATVWKLQPAFSSQ-DLHIHSR-QEEEQSEILEYSVLLPSVLF
elephant   217 SMRF-AGDTVSFEEDRINATVWKLQPTAGLQ-DLHLHSR-QEEEQSEVLEYSVLLPRVLF
mouse      218 SVSF-LGDTLSFEEDRVNATVWKLPPTAGLE-DLHIHSQ-KEEEQSEVQAYSLLLPRAVF
rat        218 SVSF-LGDTLSFEENRVNATVWKLPPTAGLE-DLQIHSQ-QEEEQSEVQAYSVLLPRAVF
gorilla    218 SVRF-TGDTVSFEEDRINATVWKLQPTAGLQ-DLHIHSR-QEEEQSEILEYSVLLPRTLF
monkey     218 SVRF-MGDTVSFEEDRVNATVWKLQPTAGLQ-DLHIHSR-QEQEQSEILEYSVLLPRTLF
chimp      218 SVRF-MGDTVSFEEDRINATVWKLQPTAGLQ-DLHIHSR-QEEEQSEILEYSVLLPRTLF
human      218 SVRF-MGDMVSFEEDRINATVWKLQPTAGLQ-DLHIHSR-QEEEQSEIMEYSVLLPRTLF
consensus  241 svrf lgdtvsfeedrvnatvwklqptaglq dlhihsr qeeeqseileysvllPrslf
ss             α       βββββββ  βββββββ    αααα   ββββ          βββββββ αααα
domain         GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

Conservation:    46464 4  4 6644664466669664666449666669646 6446 9946996669
cuckoo     271 MMAKEKEEV-LEHRVLLVDINSQTMFQDENRSHVLGDKVVGISLVDMVVANLSDPVVLTF
chicken    256 VMAKRKEEEVVEHRVLLVDINSQAMFQDENSSHILGDKVVGISLVDTVVANLSEPVVLTF
zebrafish  247 SVSRR------KSKVVCTYYKNKTLFERGPSKSALLDDIVGLSVENETIRNLIEPVKIRF
bat        278 QKTKGRRGE-AAQRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVPNTKVANLSEPVVLTF
bovine     274 QKAKGRRQE-AEKRLFLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLSEPVVLTF
elephant   274 QRAKGRRGE-AEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIIVQNTKVVNLSEPVVLTF
mouse      275 QQTRGRRRD-DAKRLLVVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVTNLSDPVVLTF
rat        275 QQTRGRRRD-AAKRLLVVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVTNLSDPVVLTF
gorilla    275 QRTKGRRGE-AEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTF
monkey     275 QRTKGRRGE-AEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTF
chimp      275 QRTKGRRGE-AEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVDNLTEPVVLTF
human      275 QRTKGRSGE-AEKRLLLVDFSSQALFQDKNSSQVLGEKVLGIVVQNTKVANLTEPVVLTF
consensus  301 qktkgrrge aekrlllvdfssqalFqdknssqvLgekvlGivvqntkvaNLsePVvltF
ss                          βββββββ                  βββββββ           βββββ
domain         GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
```

Table IV.2, continued.

(page 2 of 5; continued on the following pages)

```
Conservation:  4943464666449969497 4  4 63946 99 94  4636949 97999999696646
cuckoo     330 FHNQLPRNVTPLCVFWQEDTTD-SSGSWDSYGCATVT-RDSQTDCRCNHLTYFAVLMVSS
chicken    316 FHNRLPRNVTPLCVFWREDGTA-SSGNWDSYGCTTVE-GDTRTDCRCNHLTYFAVLMVSS
zebrafish  301 HHRPFAPDSSGRCVSW--DTKQDNEVNWKDDGCDTVKINEEQTECHCNHLTYFAILVQVE
bat        337 QHQPQPKNVTLQCVFWVEGSALTGPGSWSSAGCKTIS-REMQTSCHCDHLTYFAVLMVAS
bovine     333 QHQPQPKNVTLQCVFWVEDMTLSSPGSWSDAGCETIR-RETQTSCRCNHLTYFAVLMVAS
elephant   333 QHQPQPKNVTLQCVFWVEDPTLNSPGSWSDAGCETLR-RETQTSCLCNHLTYFAVLMVSS
mouse      334 QHQPQPKNVTLQCVFWVEDPASSSTGSWSSAGCETVS-RDTQTSCLCNHLTYFAVLMVSS
rat        334 QHQPQPKNVTLQCVFWVEDPASSSTGSWSSEGCETVS-RDTQTSCLCNHLTYFAVLMVSS
gorilla    334 QHQLQPKNVTLQCVFWVEDPTLSSPGHWSSAGCETVR-RETQTSCFCNHLTYFAVLMVSS
monkey     334 QHQPQPKNVTLQCVFWVEDPTLSNPGRWSSAGCETVR-RETQTSCFCNHLTYFAVLMVSS
chimp      334 QHQLQPKNVTLQCVFWVEDPTLSSPGHWSSAGCETVR-RETQTSCLCNHLTYFAVLMVSS
human      334 QHQLQPKNVTLQCVFWVEDPTLSSPGHWSSAGCETVR-RETQTSCFCNHLTYFAVLMVSS
consensus  361 qHqpqpknvtlqCVfWvedptlsspgsWssaGCeTvr retqTsC CnHLTYFAvLmvss
ss             ββ       βββββββββ          βββ   βββββββ  ββββββ   βββββββββ
domain         GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

Conservation:  0443439444964666669946966664 6946494       966 66   46  69669
cuckoo     388 PDITYVHRDYLSIITYIGCLISALASICTIL-FLYF------RSKQRDQIMSM-QIHMNL
chicken    374 PEISYLHRDSLSIITYIGCLISALASISTIF-FLYF------RSKQRDQITSM-HIHMNL
zebrafish  359 QKSTVRHLKALTFITAVGCAVSLVSCLVL-FYWLCK------RRRGKK--NQISLVHRGL
bat        396 VEVDAVHKHYLTLLSYVGCVISALACVLTIAAYLFS------RRKSRD--YTI-KVHMNL
bovine     392 LEVDAVHKHYLSLLSYVGCVISALACVLTIAAYLCS------RRKSRD--YTI-KVHMNL
elephant   392 VEVDAVHKHYLTFLSYVGCVISALACVCTIAAYLCS------RRKPRD--YTI-KVHMNL
mouse      393 TEVEATHKHYLTLLSYVGCVISALACVFTIAAYLCS------RRKSRD--YTI-KVHMNL
rat        393 MEVEATHKHYLTLLSYVGCVISALACVFTIAAYLCT------RRKSRD--YTI-KVHMNL
gorilla    393 VEVDAVHKHYLSLLSYVGCVVSALACIVTIAAYLCS------RRKPRD--YTI-KVHMNL
monkey     393 VEVDAVHKHYLSLLSYVGCVVSALACVVTIAAYLCS------RRKPRD--YTI-KVHMNL
chimp      393 VEVDAVHKHYLSLLSYVGCVVSALACVVTIAAYLCS------RRKPRD--YTI-KVHMNL
human      393 VEVDAVHKHYLSLLSYVGCVVSALACLVTIAAYLCSRVPLPCRRKPRD--YTI-KVHMNL
consensus  421 vevdavHkhyLsllsyvGCviSalacvvtiaayLcs      Rrk rd  yti kvHmnL
domain         lllllll77777777777777777777777777777777777777777777777777777

Conservation:  6 969996 496466469434664494446 69966994  9646964666696666666
cuckoo     440 LGAIFLLDITFLISEHLASSSSEAVCRAGGLFLHFSLLS--CLTWMGIEGYNLYRLVIEV
chicken    426 LGAIFLLDFTFLVSEHLASSSSQAACRAGGLFLHFSLLS--CLTWMGIEGYNLYRLVIEV
zebrafish  410 VVAIFLLCLFFILTGILANVANETVCQLTGSLLHYGLLSTLCWMAMEVFHTFLLVRKVFN
bat        447 LLAVFLLDVSFLLSEPVALTGSEAGCRASAIFLHFSLLA--CLSWMGLEGYNLYRLVVEV
bovine     443 LVAVFLLDVSFLLSEPVALSGSEAACRASAIFLHFSLLA--CLSWMGLEGYNLYRLVVEV
elephant   443 LLAIFLLDVSFLLSEPVALTGSEAGCHASAMLLHFSLLA--CLTWMGLEGYNLYRLVVEV
mouse      444 LSAVFLLDVSFLLSEPVALTGSEAACRTSAMFLHFSLLA--CLSWMGLEGYNLYRLVVEV
rat        444 LLAVFLLDVSFLLSEPVALMGSEAACRTSAMFLHFSLLA--CLSWMGLEGYNLYRLVVEV
gorilla    444 LLAVFLLDTSFLLSEPVALTGSEAGCRASAIFLHFSLLA--CLSWMGLEGYNLYRLVVEV
monkey     444 LLAVFLLDVSFLLSEPVALTGSQSGCRASAIFLHFSLLA--CLSWMGLEGYNLYRLVVEV
chimp      444 LLAVFLLDTSFLLSEPVALTGSEAGCRASAIFLHFSLLA--CLSWMGLEGYNLYRLVVEV
human      450 LLAVFLLDTSFLLSEPVALTGSEAGCRASAIFLHFSLLT--CLSWMGLEGYNLYRLVVEV
consensus  481 llAvFLLdvsFllsepvAltgseagCrasaifLHfsLLa  ClswMglegynLyrlvvev
domain         777777777777777777777777777777777777777777777777777777777777
```

Table IV.2, continued.

(page 3 of 5; continued on the following pages)

```
Conservation:  6446444649994 496646 446494 9444 6663646463444    4464469944
cuckoo     498 FNAYHDHFLLKLCLVGWGIPFFCVTLIFLASWTNYGPFSIPIHESVGGRSTN-ATICWIT
chicken    484 FNAYHDHFLLKLCLVGWGLPFFCVMLILLANWTNYGPFYIPVYESIGGRSTN-ATICWLT
zebrafish  470 SPLPIWIFY--LMGFGFPFLLVSILLSV-GDIYGERKI------KPSDDVNNPYRMCWMT
bat        505 FGAYVPGYLLKLSVVGWGFPAFLVTLVALVDINNYGSIILSVQRTPES--VIYPSMCWIR
bovine     501 FGTYVPGYLLKLSIVGWGFPASLVMLVALVDVNNYGRIILAVHKTPES--VIYPSMCWIQ
elephant   501 FGTYVPGYMLKLSLMGWGFPIFLVTLVALVDVNNYGPIILAVHRTPER--VIYPSMCWIR
mouse      502 FGTYVPGYLLKLSIVGWGFPVFLVTLVALVDVNNYGPIILAVRRTPER--VTYPSMCWIR
rat        502 FGTYVPGYLLKLSTVGWGFPVFLVTLVALVDVNNYGPIILAVRRTPDH--VIYPSMCWIR
gorilla    502 FGTYVPGYLLKLSAMGWGFPIFLVTLVALVDVDNYGPIILAVHRTPEG--VIYPSMCWIR
monkey     502 FGTYIPGYLLKLSAMGWGFPIFLVTLVALVDVDNYGPIILAVHRTPES--VIYPSMCWIR
chimp      502 FGTYVPGYLLKLSAMGWGFPIFLVTLVALVDVDNYGPIILAVHRTPEG--VIYPSMCWIR
human      508 FGTYVPGYLLKLSAMGWGFPIFLVTLVALVDVDNYGPIILAVHRTPEG--VIYPSMCWIR
consensus  541 fgtyvpgyllkLslvGwgfplflvtLvalvdv nygpiilavhrtpe   viypsmCWir
domain         7777777777777777777777777777777777777777777777777777777777

Conservation:      64746464969664696666 6974496646494364 474  6694649964669
cuckoo     557 ----SPLIHNVVNLGFFSLVFLFNSVMLGAMVREIVRQNKKGHKLK-HVLALLGLSILLG
chicken    543 ----SPLIHNVVNLGFFSVVFLFNLVMLGAMIREILRQNKKGHKLK-HVLALFGLSILLG
zebrafish  521 EGDKSQLAHYIINIGLLAVVVSSGLVMLFL-VVREIRNRPDWKKIHVAFLSIWGLTCLYG
bat        563 ----DSLVSHITNLGLFSLVFLFNAAMMATMVVQILRLHPHTQKWP-HVLTLLGLSLVLG
bovine     559 ----DSLVSHVTNLGLFSLVFLFNTAMLGTMVVQILRLRPHAQKWP-HVLTLLGLSLVLG
elephant   559 ----DSLVSHITNLGLFGLVFLFNMAMLGTMVVQILRLQPHAQKWP-HVLTLLGLSLVLG
mouse      560 ----DSLVSYVTNLGLFSLVFLFNLAMLATMVVQILRLRPHSQNWP-HVLTLLGLSLVLG
rat        560 ----DSVVSYVTNLGLFSLVFLFNMAMLATMVVQILRLRPHSQKWP-HVLTLLGLSLVLG
gorilla    560 ----DSLVSYITNLGLFSLVFLFNMAMLATMVVQILRLRPHTQKWS-HVLTLLGLSLVLG
monkey     560 ----DSLVSYITNLGLFSLVFLFNMAMLGTMVVQILRLRPHTQKWS-HVLTLLGLSLVLG
chimp      560 ----DSLVSYITNLGLFSLVFLFNMAMLATMVVQILRLRPHTQKWS-HVLTLLGLSLVLG
human      566 ----DSLVSYITNLGLFSLVFLFNMAMLATMVVQILRLRPHTQKWS-HVLTLLGLSLVLG
consensus  601     dslvsyitNlGlfslVflfnmaMlatmvvqilRlrphtqkw  hvLtllGLslvlG
domain         7777777777777777777777777777777777777777777777777777777777

Conservation:  4699949669699 9469 764949469699766676964644 664 4 694  46969
cuckoo     612 IPWALVFFSFTSGIFRLVSLYIFTIINSLQGFLIFLWYWTMVL-QARKS-P-DMQSSSDS
chicken    598 IPWALVFFSFTSGVFRLVSLYIFTIVNSLQGFLIFLWYWTMVL-QAKKS-S-DFHGSSDS
zebrafish  580 TTWALGFLDF--GPFSEVTLFLFCIINSLQGFFLMLRYYALERMKKKDV-SSSDGSSSGS
bat        618 LPWALVFFSFASGTFQLVVLYFFSIMTSFQGFLIFLWYWSMRL-QAQGG-PSPLKSSSDS
bovine     614 LPWALVFFSFASGTFQLVILYFFSIITSFQGFLIFLWYWSMRL-QAQVG-PSPLKSNSDS
elephant   614 LPWALVFFSFASGTFQLVVLYLFSIITSFQGLLIFLWYWSMRL-QARGGGSSSLKSSSDS
mouse      615 LPWALVFFSFASGTFQLVILYLFSIITSFQGFLIFLWYWSMRF-QAQGG-PSPLKNNSDS
rat        615 LPWALVFFSFASGTFQLVIIYLFSIMTSFQGFLIFLWYWSMRF-QAQGG-PSPLKNNSDS
gorilla    615 LPWALIFFSFASGTFQLVVLYLFSIITSFQGFLIFIWYWSMRL-QARGG-PSPLKSNSDS
monkey     615 LPWALIFFSFASGTFQLVVLYLFSIITSFQGFLIFLWYWSMRL-QARGG-PSPLKSNSDS
chimp      615 LPWALIFFSFASGTFQLVILYLFSIITSFQGFLIFIWYWSMRL-QARGG-PSPLKSNSDS
human      621 LPWALIFFSFASGTFQLVVLYLFSIITSFQGFLIFIWYWSMRL-QARGG-PSPLKSNSDS
consensus  661 lpWALvFfsFasGtFqlVvlylFsIitSfQGfliflwYwsmrl qargg psplksnSdS
domain         7777777777777777777777777777777777777IIIIIIIIIIIIIIIIIIIII
```

Table IV.2, continued.

(page 4 of 5; continued on the following page)

```
Conservation:   64444666  6  4
cuckoo      669 VKLQP--NSRESHLG
chicken     655 VKLQP--NSSQSHPG
zebrafish   637 SKQHM---LQTNEKS
bat         676 TRLPITSGSTSSSCI
bovine      672 ARLPISSGSTSSSRI
elephant    673 ARIPINSGSTSSGRI
mouse       673 AKLPISSGSTSSSRI
rat         673 AKLPISSGSTSSSRI
gorilla     673 ARLPISSGSTSSSRI
monkey      673 ARLPISTGSTSSSRI
chimp       673 ARLPISSGSTSSSRI
human       679 ARLPISSGSTSSSRI
consensus   721 arlpissgstsssri
domain          IIIIIIIIIIIIIII
```

protein:	**Uniprot ID:**
cuckoo: cuckoo GPR56	(A0A091FQY3)
chicken: chicken GPR56	(E1C0Q2)
zebrafish: zebrafish GPR56	(F1QZM9)
bat: *Myotis lucifugus* GPR56	(G1PD76)
bovine: bovine GPR56	(A4IF70)
elephant: elephant GPR56	(G3T7E2)
mouse: mouse GPR56	(Q8K209)
rat: rat GPR56	(Q8K3V3)
gorilla: gorilla GPR56	(Q50DM6)
monkey: Rhesus macaque GPR56	(Q50DM8)
chimp: chimpanzee GPR56	(Q50DM7)
human: human GPR56	(Q9Y653)

Table IV.2, continued.

Conservation score is calculated for positions with at least some detectible conservation (9 is most conserved). ECR domain structure is annotated based on the crystal structure of the ECR, while 7TM and intracellular region (ICR) are annotated based on sequence-based predictions: s, signal peptide; P, PLL; l, linker; G, GAIN; 7, 7TM; I, ICR. Secondary structure is also annotated for the ECR: α, α-helix; β, β-strand based on the crystal structure. Cysteine residues involved in the intra-PLL domain disulfide bond are colored red. Cysteine residues involved in the interdomain (PLL-GAIN) disulfide bond are colored blue. The conserved pentraxin motif on the PLL domain, beginning with mouse H89, is highlighted yellow. The conserved N-linked glycosylation site (glycan is adjacent to the conserved patch on the PLL domain) is highlighted green. Autoproteolysis occurs between the two residues highlighted in black. Bolded residues have been found mutated in human diseases.

(page 5 of 5)

Superimposition of the GPR56 GAIN domain structure with the previously determined crystal structures of the GAIN domains of aGPCRs Latrophilin 1 (Lphn1/ADGRL1) and brain angiogenesis inhibitor 3 (BAI3/ADGRB3) shows that, intriguingly, subdomain A of GPR56 contains only three helices and is much smaller than subdomains A of Lphn1 and BAI3 that each have six helices (Figure IV.1H) (55). On the other hand, the conformation of subdomain B, which contains the autoproteolysis site and the *Stachel*, is highly conserved among the three aGPCRs, particularly around the *Stachel* (55) (Figure IV.2G-I). Analysis of the 2Fo-Fc electron density map confirmed that the GPR56 ECR underwent autoproteolysis (Figure IV.2G). Mass spectrometry showed that the GPR56 GAIN domain alone was sufficient to mediate autoproteolysis (Figure IV.2C), consistent with previous observations for Lphn1 and BAI3(55). Two BFPP mutations now unambiguously mapped to the GAIN domain: C346S, which eliminates a conserved disulfide bond, and W349S, which mutates a conserved hydrophobic core residue, are likely to cause global folding problems of the GAIN domain, consistent with previous findings that these mutants undergo little to no autoproteolysis (158, 159) (Figure IV.1I). Together, these results show that even though the GAIN domain of GPR56 is unexpectedly smaller than other known GAIN domains (55), it retains autoproteolytic activity.

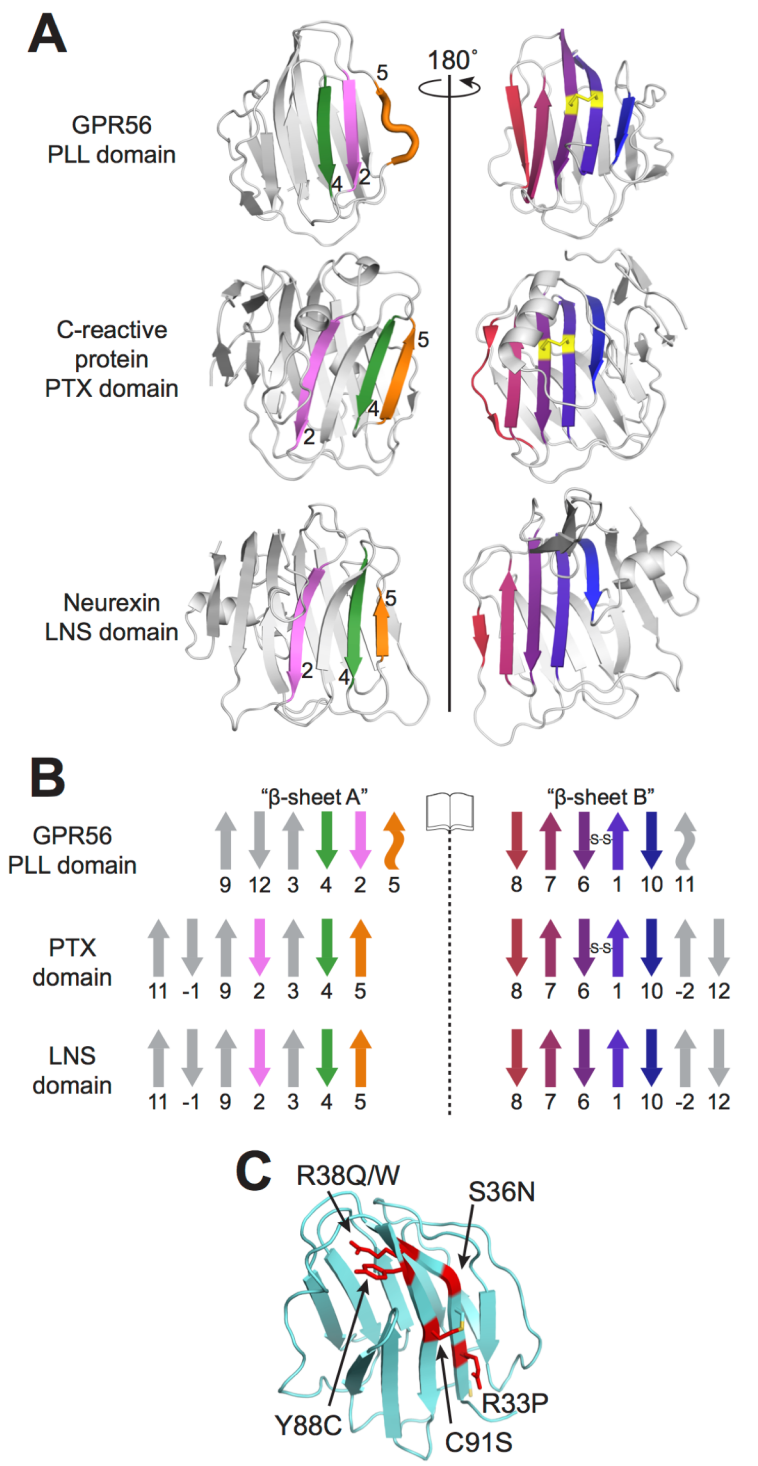


Figure IV.4 The PLL domain of GPR56 is a previously unidentified fold that likely diverged from the pentraxin and LNS folds.

(legend on the following page)

The N-terminal domain of GPR56 has a previously unidentified fold

The crystal structure of the GPR56 ECR reveals a 133 amino acid, 12-stranded β-sandwich domain at the N-terminus of mature GPR56 (residues P28-S160) (Figure IV.1F). We denote the two β-sheets as β-sheet A (strands 2-5, 9 and 12) and β-sheet B (strands 1, 6-8, 10 and 11). Using the DALI (160) and HorA (161) servers, we found that this domain has weak homology to the pentraxin (PTX) and laminin/neurexin/sex hormone-binding globulin (LNS) domain families, but no strong homology to any known fold (top DALI hit: LNS, Z-score=6.5; top HorA hit: PTX, combined score=4.48). Superimposition of the GPR56 N-terminal domain with LNS or PTX domains yields a high backbone rmsd (5.7 Å and 4.7 Å, respectively), whereas superimposition of LNS domain with PTX domain yields a lower backbone rmsd (~3.2 Å), suggesting that the GPR56 N-terminal domain has diverged more from both PTX and LNS domains than the PTX and LNS domains have from each other (Figure IV.4A). Though the N-terminal domain of GPR56 has low sequence identity with PTX and LNS domains (18% and 19%, respectively for the family member with the highest identity), we found that it has a conserved motif (HΦC^{91}xxWxxxxG) that we identified among canonical PTX domains

Figure IV.4, continued

(A) The PLL domain of GPR56, the PTX domain of C-reactive protein (PDB ID: 3PVN), and the LNS domain of Neurexin-1 beta (PDB ID: 3QCW) in similar orientation. β-Strands are numbered from N to C terminus, and equivalent β-strands are colored in the same manner. Cys residues involved in a disulfide bond are colored yellow. (B) Schematic of β-strand connectivity comprising the two β-sheets of each domain. Wavy arrows represent loops with geometry similar to a β-strand. (C) Human disease mutations (red) mapped to the PLL domain.

(Table IV.3 and Table IV.4). Thus we termed this GPR56 domain the PTX/LNS-Like (PLL) domain.

The connectivity of the β-strands in the PLL domain of GPR56 is substantially different from the completely conserved connectivity within the PTX and LNS families (Figure IV.4A-B). Interestingly, the majority of the changes in β-strand connectivity map to β-sheet A. All PTX and LNS domains have completely antiparallel β-sheets, whereas β-sheet A of the PLL domain of GPR56 is a mixed β-sheet with β2 and β4 strands parallel to each other (Figure IV.4A-B). In contrast, the other β-sheet, β-sheet B, is antiparallel as in the PTX and LNS domains, and contains all six PLL domain-localized BFPP mutations (Figure IV.4A-C). Additionally, the locations of β2, β11, and β12 strands with respect to the other β-strands sets the PLL domain apart from the known PTX and LNS folds (Figure IV.4B). Thus, the PLL domain of GPR56 has a unique fold that likely diverged from the PTX and LNS domain folds.

```
hCRP        47 PL-KAFTVCLHFYTE-LS-S----TRGYSIFS-YATKRQDNEI-LIF-WS---KDIGYS-
hSAP        48 PL-QNFTLCFRAYSD-LS-------RAYSLFS-YNTQGRDNEL-LVY-KE---RVGEYS-
mGPR56      28 P-REDFRFCGQRN---Q-------TQQSTLHY---DQSSEPHI-FV--WN---TEETLT-
hGPR126    178 PELSAFTLCFEATKV-GH-E----DSDWTAFS-YSNASF-TQL-LSF-GK---AKSGYF-
hGPR112     50 PELSRFTACIDLVFM-DD-N----SRYWMAFS-YITNN----ALLGR-ED---IDLGLAG
hGPR144    138 PELAALTACTHVQWD-CA-S----PDPAALFS-VAAPALPNAL-QLRAFAEPGGVVRAA-
hGPR133     97 PEQCGPEGVTFSFFWKTQ-G----EQSRPIPSAYGGQVISNGF-KV--CS-SGGRGSVE-
hGPR98    1346 PSR--NNTIANFTFS-AWVMPNANTNGFIIAKDDGNGSIYYGV-KI--QT---NESHVT-
hSHBG       65 KI---TKTSSSFEVR-TW-----DPEGVIFYG-DTN-PKDDWFMLGL-RD---GRPEIQ-
bNRX1a     316 PI-QSSSDEITLSFK-TL-----QRNGLMLH--TGK-S-ADYVNLAL-KN---GAVSLV-
mNRX3b     103 PS--TRSDRLAVGFS-TT-----VKDGILVRI-DSAPGLGDFLQLHI-EQ---GKIGVV-
mLAMa2    2953 K--VGLDLLVEFEFR-TT-R----PTGVLLG--ISSQK-MDGMGIEM-ID---EKLMFH-


hCRP        93 --FT--VG--GSEIL--FEVP-EV-TV---------APVHICTSWESASGIVEFWVDGKP
hSAP        92 --LY--IG--RHKVT--SKVI-EK-FP---------APVHICVSWESSSGIAEFWINGTP
mGPR56      67 --I----------R----APFLAAPDIPRFF--PEPRGLYHFCLYWSRHTGRLHLRYGKHD
hGPR126    224 --LS--IS--DSKCL--LNNALPVKEKEDI--FA-ESFEQLCLVWNNSLGSIGVNFKRNY
hGPR112     95 DHQQ--LI--L--YR-LGKTF--SIRH--HLAS--FQWHTICLIWDGVKGKLELFLNKER
hGPR144    189 --LV--VR--GQHAP--FLAA-FR-AD--------GRWHHVCATWEQRGGRWALFSDGRR
hGPR133    147 --LY--TR--DNSMT--WEAS-FS-PPG-------PYWTHVLFTWKSKEG-LKVYVNGTL
hGPR98    1396 --LSLHYKTLGSN----ATYIAKTTVMKYL--EE-SVWLHLLII--LEDGIIEFYLDGNA
hSHBG      109 --LH--NH--WAQLT--VG---AGPRL--D--D--GRWHQVEVK--MEGDSVLLEVDGEE
bNRX1a     360 --IN--LG--SGAFE--ALVEPVNGKF--N--D--NAWHDVKVT--RNLRQVTISVDGIL
mNRX3b     150 --FN--IG--TVDIS--IKE--ERTPV--N--D--GKYHVVRFT--RNGANATLQVDNWP
mLAMa2    2997 --VD--NG--AGRFTAIYDAE-IPGHM--C--N--GQWHKVTAK--KIKNRLELVVDGNQ


hCRP       134 -RV-R--KSL-K-KG--YTVGAEASIILG--Q-EQ-----------SQSLV-GDIGNVNM
hSAP       133 -LV-K--KGL-R-QG--YFVEAQPKIVLG--Q-EQ-DSYGG-KFDRSQSFV-GEIGDLYM
mGPR56     110 -YL-LSSQAS-R-LL--CFQKQEQSLKQGAPLIAT-SVSSW-QIPQNTSLP-GAPSFI--
hGPR126     96 -ET-VPCD-S-T-IS--KVIPGNGKLLLG--S----------NQNEIVSLK-GDIYNFRL
hGPR112    142 -IL-E---VTD--QP--HNLTPHGTLFLG--H-FLKNESSE-VKSMMRSFP-GSLYYFQL
hGPR144     94 -RAGA--RGL-G-AG--HPVPSGGILVLG--Q-DQ-DSLGG-GFSVRHALS-GNLTDFHL
hGPR133    189 STSDP--SGK-V-SR--DYGESNVNLVIG--S-EQ-DQA--------KCYENGAFDEFII
hGPR98    1445 -MP-RGIKSL-K-GE--AITDGPGILRIG--A----------GINGNDRFT-GLMQDVRS
hSHBG      150 -VL-R--LRQVS-GPLTSKRHPIMRIALG--G-LLFPASNL-RLPLVPALD-GCLRRDSW
bNRX1a     404 -TT-T--GYTQE-DY--TMLGSDDFFYVG--G-SPSTADLP-GSPVSNNFM-GCLKEVVY
mNRX3b     192 -VN-E--HYPTGRQL--TIFNTQAQIAIG--G-KDKG-RLFQGQLSGLYYD-GLKVLNMA
mLAMa2    3042 -VD-A--QSPNS-AS--TSADTNDPVFVG--G-FPGGLNQF-GLTTNIRFR-GCIRSLKL


hCRP       180 WDFVLSPDEINTIYLGGPFSP-NVL-NWRALKYEVQGE--VFTKPQLWP
hSAP       179 WDSVLPPENILSAYQGTPLPA-NIL-DWQALNYEIRGY--VIIKPLVWV
mGPR56     159 FSFHNAPHKVSHNASV-------------------------------D
hGPR126    313 WNFTMNAKIL-SNLSC-NVKGNVV--DWQNDFWNIPNL--ALKAESNLS
hGPR112    188 WDHILENEEFMKCLD-----G-NIV-SWEEDVWLVNKI----------I
hGPR144    278 WARALSPAQLHRARACAPPSE-GLLFRWDPGALD--------------V
hGPR133    231 WERALTPDE-----------------------------------------I
hGPR98    1486 YERKLTLEE-----------------------------------------I
hSHBG      200 LDKQAEIS---ASAP-----------------TSLRS-----------C
bNRX1a     137 KNNDVRLELSRLAKQGDP-------------KMKIHG-----------V
mNRX3b     241 AENNPN--------------------------IKINGS-------VRLV
mLAMa2    3090 TKGTGKPLEVNFAKA-----------------LELRGVQPVS---CPTT
```

Table IV.3 Multiple sequence alignment of PTX and LNS domains.

(page 1 of 2; continued on the following page)

hCRP: human C-reactive_protein (P02741)
hSAPC: human Serum amyloid p-component (P02743)
mGPR56: mouse GPR56 (Q8K209)
hGPR126: human GPR126 (Q86SQ4)
hGPR112: human GPR112(Q8IZF6)
hGPR144: human GPR144(Q7Z7M1)
hGPR133: human GPR133 (Q6QNK2)
hGPR98: human GPR98 (Q9WXG9)
hSHBG: human sex hormone-binding globulin (P04278)
bNRX1a: bovine neurexin-1-alpha, isoform 9a (Q28146-9)
mNRX3b: mouse neurexin-3-beta, without splice insert 4 (Q8C985)
mLAMa2: mouse laminin subunit alpha-2 (Q60675)

Table IV.3, continued.

Canonical Pentraxin (PTX) and Laminin/Neurexin/Sex hormone-binding globulin (LNS) domains were identified including all PTX and LNS domains annotated to be present in all aGPCR ECRs and aligned with the PLL domain from GPR56/ADGRG1. Bolded sequences have well-defined PTX domain features (including an intradomain disulfide bond, cysteine residues colored red, and the conserved PTX motif, highlighted yellow). GPR133/ADGRD1 and GPR98/ADGRV1 are annotated to have a PTX domain, though they lack the intradomain disulfide bond and therefore do not have a complete PTX motif, highlighted cyan.

(page 2 of 2)

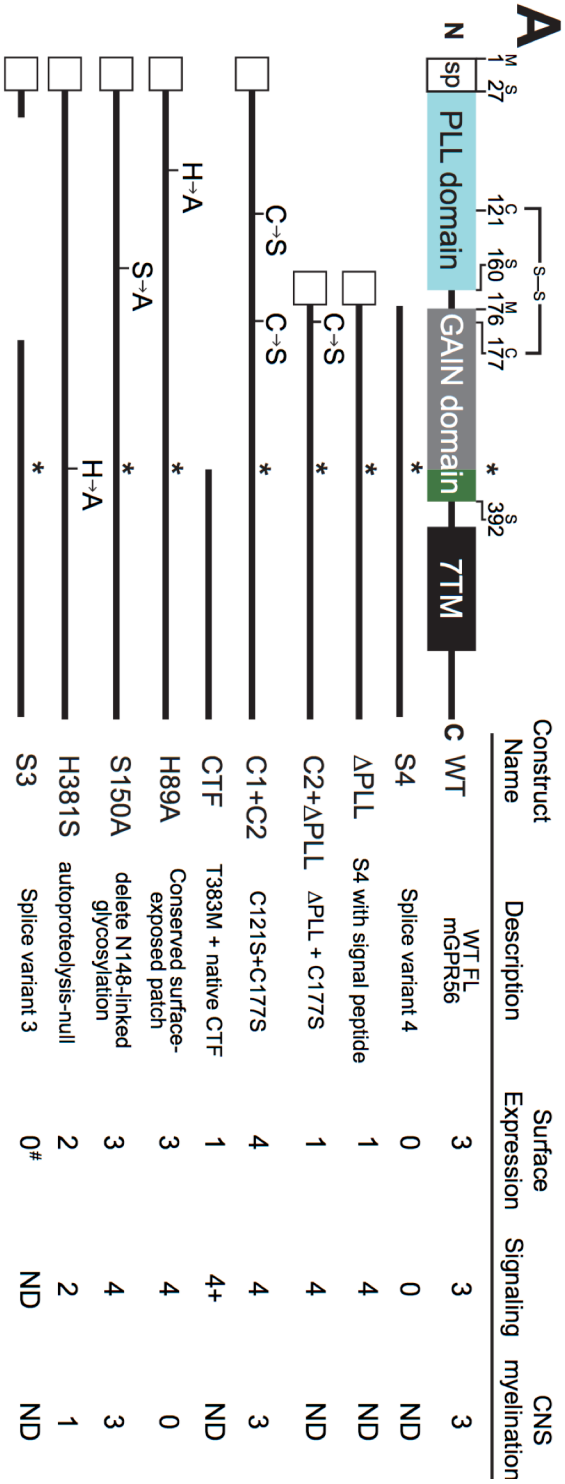
```
Conservation: 9 7  77 9        4      777     44 4           4
hCRP         47 P-LKAFTVCLHFYTELSSTRGYSIFSYATKRQDNEILI--FWSKDIGYSFTV--G-GSE-
hSAPC        48 P-LQNFTLCFRAYSDLS--RAYSLFSYNTQGRDNELLV--YKERVGEYSLYI--G-RHK-
mGPR56       28 P-REDFRFCGQR-----NQTQQSTLHYDQSSEPHIF----VWNTEETLTIRAPFLAAPD-
hGPR126     178 PELSAFTLCFEATKVGHEDSDWTAFSYSNASFTQLL-S--FGKAKSGYFLSI---SDSK-
hGPR112      50 PELSRFTACIDLVFMDDNSRYWMAFSYITN---NAL----LGREDIDLGLAG--D-H-Q-
hGPR144     138 PELAALTACTHVQWDCASPDPAALFSVAAPALPNALQLRAFAEPGGVVRAALVVR-GQHA
consensus     1 Pel aftvC h   e    r ysifsy t    n l i  fg  d  ysl v    g k

Conservation:    4        4               4 9  9    9   4   4  3 4   4   4
hCRP        100 IL-F-EVP----EV--------TV-APVHICTSWESASGIVEFWVDGK-PRVRKSLKKGY
hSAPC        99 VT-S-KVI----EK--------FP-APVHICVSWESSSGIAEFWINGT-PLVKKGLRQGY
mGPR56       77 IP----------RFFPE-----P-RGLYHFCLYWSRHTGRLHLRYGKH----------DY
hGPR126     231 CL-L-NNALPVKEKEDI-----FAESFEQLCLVWNNSLGSIGVNFKRNYETVPCDSTISK
hGPR112      98 QLILYRLG----KTFSIRHHLASF-QWHTICLIWDGVKGKLELFLNKE--RILEVTDQPH
hGPR144     197 PF-L-AAF----RA--------DG-RWHHVCATWEQRGGRWALFSDGRRRAGARGLGAGH
consensus    61 il l  v     e            af hiClsWe  sGrvelwv gk   v k l  gy

Conservation:  4   4 4 97  4    4     6   44 7 43 4 49   4 4  4     4 6
hCRP        144 TVGAEASIILGQEQ----------SQSLVGDIGNVNMWDFVLSPDEINTIYLGGPFSP
hSAPC       143 FVEAQPKIVLGQEQ---DSYGGKFDRSQSFVGEIGDLYMWDSVLPPENILSAYQGTPLPA
mGPR56      111 LLSSQASRLLCFQKQ--EQ-----SLKQGAPLIATSVSSWQIPQN--------TSLPGAP
hGPR126     284 VIPGNGKLLLGSNQ---NE-------IVSLKGDIYNFRLWNFTMNAKILSN--LSCNVKG
hGPR112     151 NLTPHGTLFLGHFLKNESSEVK--SMMRSFPGSLYYFQLWDHILENEEFMK----C-LDG
hGPR144     242 PVPSGGILVLGQDQ---DSLGGGFSVRHALSGNLTDFHLWARALSPAQLHRARACAPPSE
consensus 121  v aqgsliLgqeq   ds     sl qsl gdi  f lWd vl pd i       pl a

Conservation: 4 4  7      6   6
hCRP        201 NVL-NWRALKYEVQGEVFTKPQLWP
hSAPC       200 NIL-DWQALNYEIRGYVIIKPLVWV
mGPR56      156 SFIFSFHNAPHKVSHNASV-----D
hGPR126     331 NVV-DWQNDFWNIPNLALKAESNLS
hGPR112     204 NIV-SWEEDVWLVNKI--------I
hGPR144     299 GLLFRWDPGALD------------V
consensus 181 nvl  w    yev           v


hCRP: human C-reactive_protein (P02741)
hSAPC: human Serum amyloid p-component (P02743)
mGPR56: mouse GPR56 (Q8K209)
hGPR126: human GPR126 (Q86SQ4)
hGPR112: human GPR112(Q8IZF6)
hGPR144: human GPR144(Q7Z7M1)
```

Table IV.4 Multiple sequence alignment of PTX domains.

Canonical pentraxin (PTX) domains were identified including all the well-defined PTX domains present in all aGPCR ECRs and aligned to the PLL domain from GPR56/ADGRG1. Conservation score is calculated for positions with at least some detectible conservation (9 is most conserved). PTX domain features are illustrated (including an intradomain disulfide bond, cysteine residues colored red, and the conserved PTX motif, highlighted yellow). The conserved PTX motif is defined as HΦC*xxWxxxxG, where Φ is a hydrophobic residue and the C* participates in the intradomain disulfide bond.

The PLL domain is deleted in GPR56 splice variant 4

Although AS of upstream regulatory elements of *GPR56* controls regional cerebral cortical patterning (76), the role of AS of *GPR56*'s coding sequence is unknown. AS in the coding region of genes is an important mechanism to greatly expand the functional and regulatory capacity of metazoan genomes and its regulatory role in brain function has been repeatedly demonstrated (156, 162, 163). For instance, recent studies suggest a specific expression pattern for hundreds of alternatively spliced isoforms of neurexins, key proteins that organize synapse architecture and encode cellular identity and diversity (164). The coding sequence of human *GPR56* consists of 13 exons and the ECR is encoded by exons 2-9 (Figure IV.5A-B). AS occurs in exons 2, 3, and 10, resulting in a total of five variants (WT and S1-S4), of which only S3 and S4 result in substantial changes to the ECR (153). S3 has a large deletion encompassing the 3' end of exon 2 and all of exon 3 (ΔR38-Q207) including the C-terminal portion of the PLL domain and a small N-terminal portion of the GAIN domain, likely resulting in a hybrid domain with unknown structure (Figure IV.5). In S4, 43 nucleotides at the 5' end of exon 2 are deleted, resulting in a frameshift and therefore a new translation start site at M176 (Figure IV.5A-C). AS in the ECR-coding region of human and mouse GPR56 is identical, specifically the formation of S3 and S4. Strikingly, the crystal structure of the GPR56 ECR revealed that M176 corresponds exactly to the first residue of the GAIN domain. Therefore, S4 lacks the N-terminal 175 residues including the signal peptide sequence (M1-G26), PLL domain (S27-S160), and PLL-GAIN linker (F161-D175). These observations suggest that by regulating AS, a cell may generate GPR56 with or without a PLL domain in the ECR, which could diversify functionality.



Construct Name	Description	Surface Expression	Signaling	CNS myelination
WT	WT FL mGPR56	3	3	3
S4	Splice variant 4	0	0	ND
ΔPLL	S4 with signal peptide	1	4	ND
C2+ΔPLL	ΔPLL + C177S	1	4	ND
C1+C2	C121S+C177S	4	4	3
CTF	T383M + native CTF	1	4+	ND
H89A	Conserved surface-exposed patch	3	4	0
S150A	delete N148-linked glycosylation	3	4	3
H381S	autoproteolysis-null	2	2	1
S3	Splice variant 3	0#	ND	ND

Figure IV.5 Precise deletion of the PLL domain, as in splice variant 4, leads to increased basal activity.

(continued on the following page)

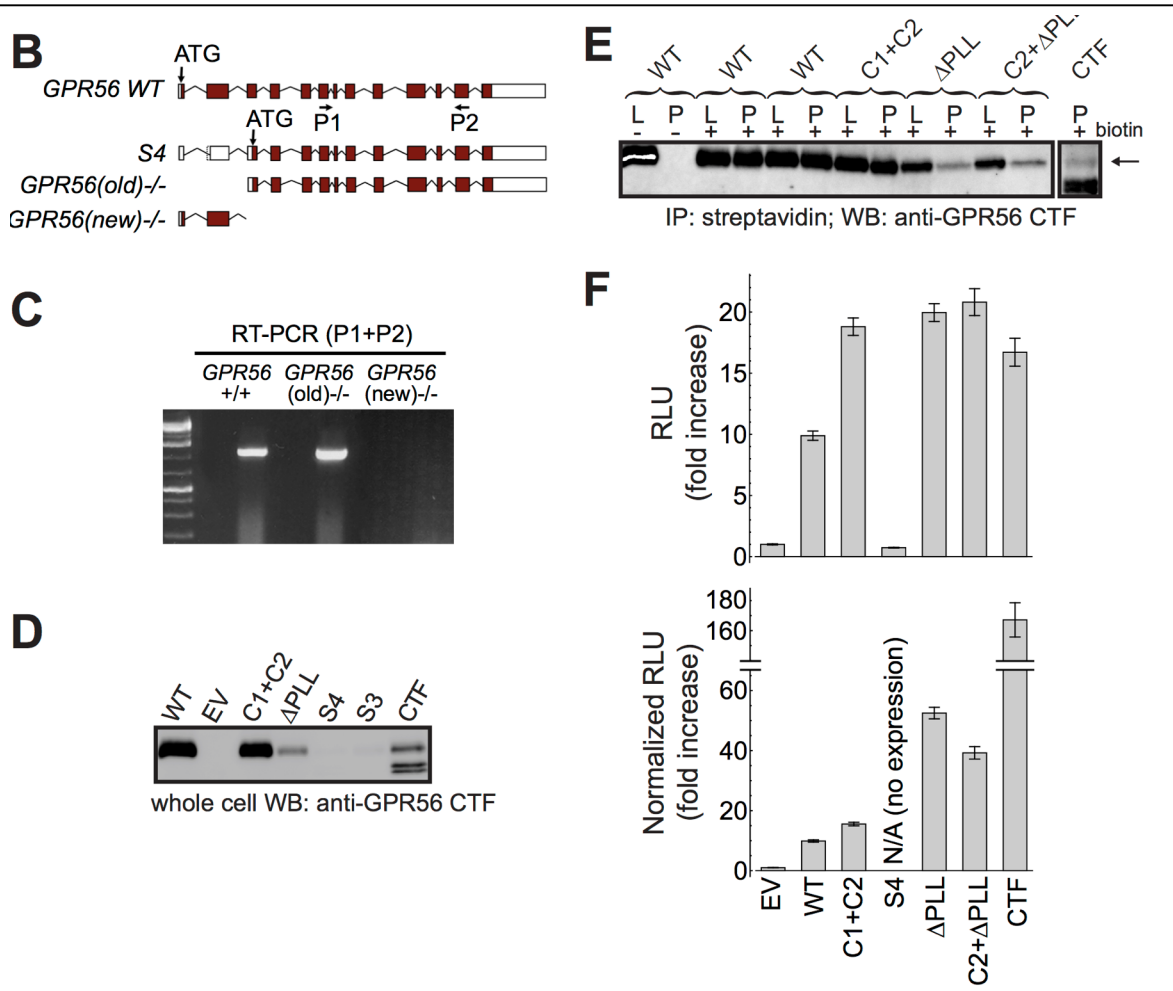


Figure IV.5, continued.

(legend on the following page)

PLL domain deletion increases GPR56 basal activity

In order to test the role of the PLL domain in GPR56 G protein signaling, we generated various constructs and assayed their cell surface expression and signaling capability. These constructs include splice variant 4 (termed 'S4', deleting residues M1-D175 including the signal peptide), and splice variant 4 with the signal peptide (effectively a deletion of the PLL domain and PLL-GAIN linker, termed 'ΔPLL', deleting residues G26-D175). Due to the interdomain disulfide bond, PLL domain deletion in both constructs generates a free cysteine that may mediate nonspecific interactions. Therefore, we also generated a construct with the C177S mutation on ΔPLL (termed 'C2+ΔPLL'). Additionally, we generated a construct corresponding to full-length GPR56

Figure IV.5, continued.

(A) Domain architecture schematics and function-metrics of important GPR56 constructs (data compiled from D-F, Figure 5 and Table S1). Residue numbers for domain boundaries based on the crystal structure and the interdomain disulfide bond are shown. Function metrics used are: (0) none, (1) very little, (2) less than WT, (3) comparable to WT, (4) more than WT, (ND) not determined. #see also Figure IV.6. (B) Expected transcripts in two knockout mouse alleles. The starting ATG for WT *GPR56* is in exon 2. The S4 variant has its starting ATG in exon 4. The targeting strategy for *GPR56(old)*$^{-/-}$ mice was to delete exons 2 and 3, which preserved the S4 variant, whereas the *GPR56(new)*$^{-/-}$ allele deletes exon 4-6, causing a frameshift that leads to a deletion of all splicing variants of *GPR56*. (C) RT-PCR showing the presence of the S4 transcript in *GPR56(old)*$^{-/-}$ but absence in *GPR56(new)*$^{-/-}$ mouse brains. (D-E) In order to quantify cell-surface expression of GPR56 mutant constructs with decreased affinity for α5, IP-western blot was performed. (D) Western blot of whole cell lysates of cells expressing different GPR56 constructs. (E) Western blot of lysate (L) and lysate subject to streptavidin pull-down (P) of HEK293T cells transfected with WT and mutant *GPR56* constructs. (F) Basal activity of mutant GPR56 constructs as measured by the SRE-luciferase reporter assay. Top: Basal activity of mutant constructs. Bottom: Basal activity of mutant constructs normalized for cell-surface expression using band densities from E. Data are presented as mean ± S.E.M.; *n* = 3. sp, signal peptide.

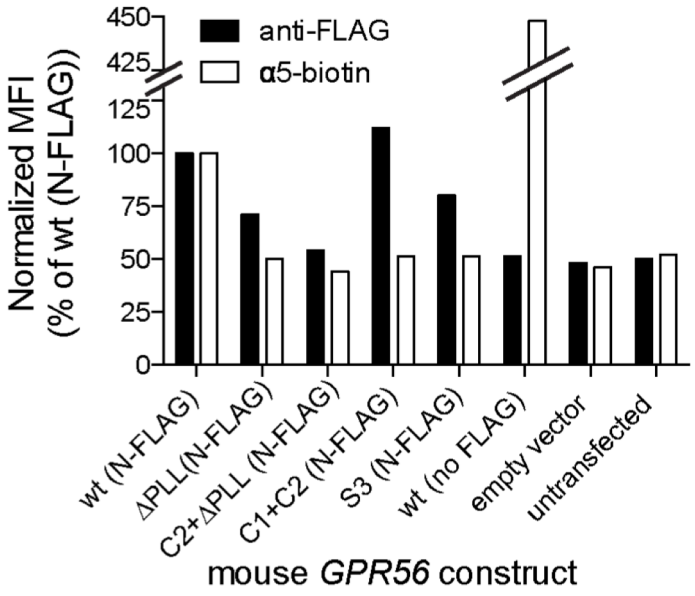


Figure IV.6 Cell-surface expression of mouse GPR56.

Flow cytometry of GPR56 constructs stained with anti-FLAG antibody and α5.

without the interdomain disulfide bond (termed 'C1+C2' corresponding to C121S+C177S) and one that corresponds to the cleaved C-terminal fragment (CTF, including the *Stachel* and 7TM) that has been reported to have dramatically increased basal activity (Figure IV.5A and Table IV.5) (35). Proper cell-surface expression and trafficking of these constructs in HEK293T cells was quantified using a cell-surface biotinylation assay followed by streptavidin pull-down and western blot using an antibody directed to the C-terminus of GPR56 (35, 90). We note that we did not attach an N-terminal epitope tag such as FLAG to these constructs, because modifications near the N-terminus led to dramatically reduced cell-surface expression and signaling (Figure IV.6).

Construct	Description of mutation	Surface expression (HEK293T)	SRE assay signaling (not normalized for expression)	function in zebrafish
wild-type	wt FL mGPR56	3	3	0
wild-type (N-FLAG)	wt FL mGPR56 with N-FLAG	1	0	ND
R33E	R33N is a human disease mutation	2	ND	ND
G36D+Q37A	S36P is a human disease mutaiton	0	ND	ND
R38Q	human disease mutation	0	ND	ND
N39A	delete N39-linked glycosylation	2	ND	ND
T41A	delete N39-linked glycosylation	2	ND	ND
Y88S	Y88C is a human disease mutation	1	ND	ND
Y88C	human disease mutation	0	ND	ND
H89A	conserved, surface-exposed patch	3	4	3
C91S	human disease mutation	0	ND	ND
G106R	conserved, surface-exposed patch	1	ND	ND
C121S+C177S	delete interdomain disulfide	4	4	0
A137R+S139R	conserved, surface-exposed patch	2	ND	ND
N148A	delete N148-linked glycosylation	3	ND	ND
S150A	delete N148-linked glycosylation	3	4	0
H381S	autoproteolysis-null	2	2	1
T383G	autoproteolysis-null	4	2	ND
S3	splice variant 3	0	ND	ND
S4	splice variant 4 (Δsignal peptide + ΔPLL)	0	0	ND
ΔPLL	S4 with signal peptide	1	4	ND
CTF	T383M,Y384,etc. through native C-terminus	1	4	ND
7TM	H401M,Y402,etc. through native C-terminus	0	0	ND

Construct	Description of mutation	Yield in insect cells (purified ECR)
wild-type	wt FL mGPR56	3
R38Q	human disease mutation	0
R38W	human disease mutation	0
Y88C	human disease mutation	1
Y88S	Similar to Y88C	1
C121S+C177S	delete interdomain disulfide	4

Scale:	
none	0
little bit	1
less than wt	2
comparable to wt	3
more than wt	4
not determined	ND

Table IV.5 Expression, signaling, and myelination phenotypes of all mouse GPR56 mutants tested.

Human and mouse GPR56 are both reported to activate $G\alpha_{13}$, which is upstream of RhoA and serum response element (SRE) (35, 75) (Figure IV.7A). Interestingly, we found mouse GPR56 weakly coupled to an additional G protein, $G\alpha_q$, but not $G\alpha_{i/o}$, which can couple to human GPR56 (Figure IV.7B-D) (35), likely illustrating different roles for GPR56 across species. We used an SRE-luciferase assay to measure $G\alpha_{13}$ G protein signaling of HEK293T cells overexpressing WT or mutant GPR56 constructs. Overexpression of full-length GPR56 in HEK293T cells resulted in higher luciferase activity when compared to cells transfected with an empty vector, showing that the

basal activity of overexpressed GPR56 can be detected in this assay. We detected much lower basal activity for S4 expressed in HEK293T cells relative to cells expressing WT GPR56, as reported (153), but we found that S4 had undetectable expression in this system (Figure IV.5D), rendering it difficult to evaluate the functional state of S4. In contrast, ΔPLL, which differs from S4 only by the presence of the N-terminal secretion signal sequence, showed detectable surface expression in HEK293T cells (Figure

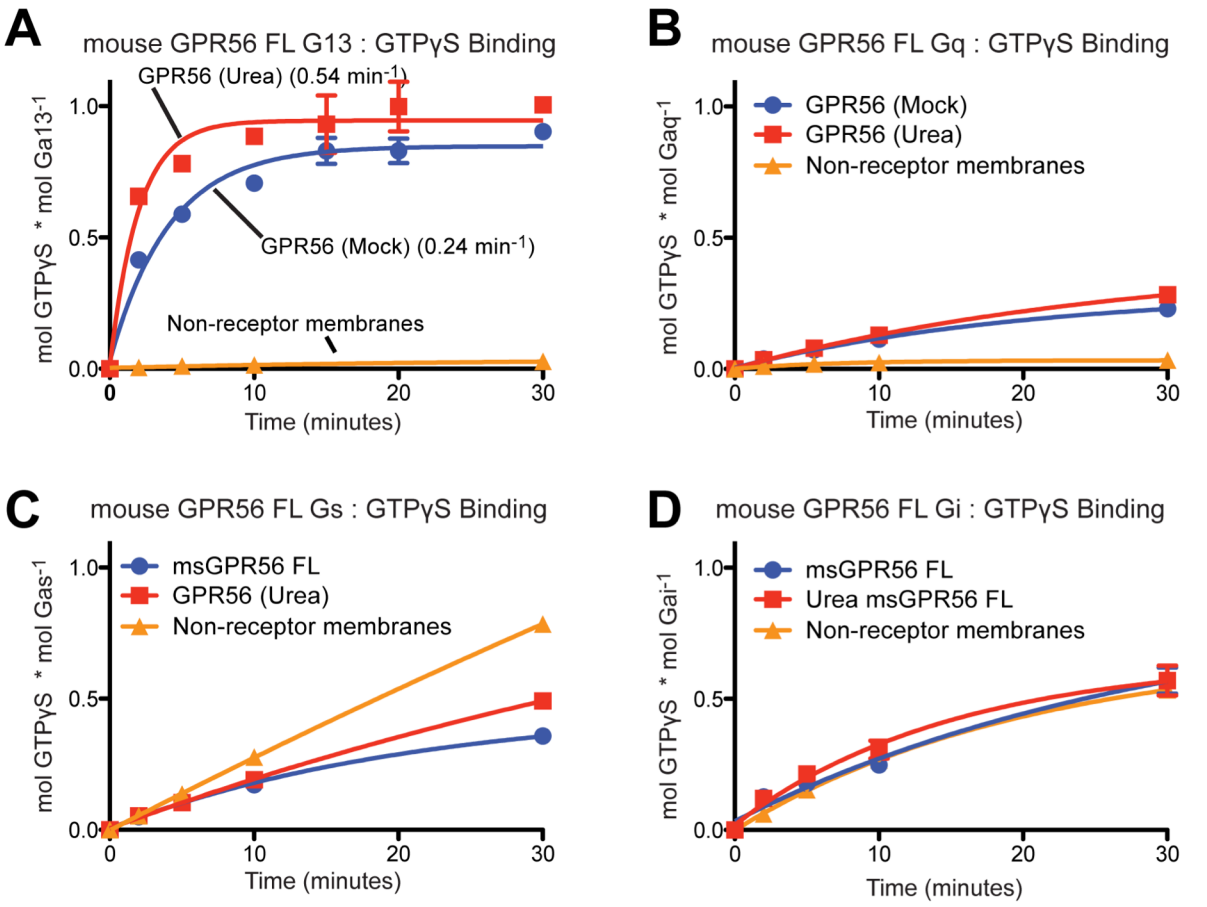
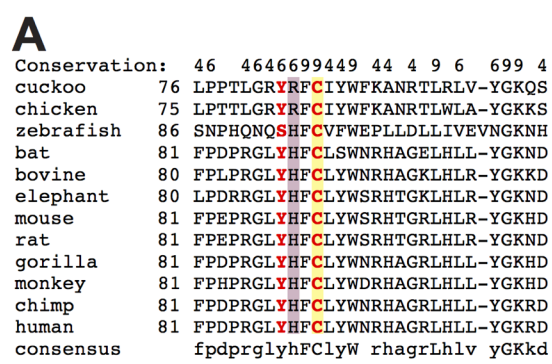


Figure IV.7 G protein signaling of mouse GPR56.

High-Five membranes with full-length mouse GPR56 were subject to mock or urea treatment to induce NTF-shedding and compared to membranes with no receptor. Membranes were reconstituted with G proteins (α and βγ). The receptor-mediated G protein activation kinetics were measured using the [35S]-GTPγS binding assay. (A) GPR56 and $G\alpha_{13}$. (B) GPR56 and $G\alpha_{q}$. (C) GPR56 and $G\alpha_{s}$. (D) GPR56 and $G\alpha_{i}$.

IV.5D-E); thus, we used this construct to represent the function of S4 in HEK293T-based signaling assays. Importantly, the majority of GPCRs do not contain an N-terminal secretion signal peptide, yet are still properly trafficked to the plasma membrane (165). This suggests that despite the requirement of a signal peptide in HEK293T cell culture (as in ΔPLL), GPR56 S4 may be properly trafficked *in vivo*, consistent with its detectable expression in mouse brain (Figure IV.5C).

Strikingly, the ΔPLL construct had higher basal activity than WT GPR56 (~2 fold without normalization for surface expression and ~5 fold with normalization). This suggests that deletion of the PLL domain in S4 would result in increased basal G protein signaling as long as S4 is properly trafficked to the plasma membrane. C2+ΔPLL, which removes the unpaired Cys residue in the GAIN domain after the deletion of the PLL domain, produced similar results (~2 fold and ~4 fold increase in basal activity without and with normalization for surface expression, respectively), indicating that the elevated activity is not caused by potentially anomalous conjugation involving this unpaired sulfhydryl group (Figure IV.5F). Moreover, eliminating the disulfide bond between the PLL and GAIN domains in full-length GPR56 also increased GPR56 basal activity (Figure IV.5F), suggesting the importance of restricting the flexibility within the ECR in keeping the receptor in the basal state. We generated several additional GPR56 ECR mutants and assayed their basal activity using the same SRE-luciferase reporter (Table IV.5). Altogether, these results suggest that the ECR regulates GPR56 signaling via complex mechanisms and that AS of the coding region modulates GPR56 function.



B

Gpr56 fragment	% identity to human			
	mouse	zebrafish	hedgehog	gorilla
full length	79%	26%	76%	97%
ECR	73%	22%	72%	97%
PLL	65%	21%	58%	96%
PLL-β-sheet B	77%	23%	66%	97%
PLL-β-sheet A	53%	24%	51%	95%
GAIN	79%	29%	79%	98%
7TM	88%	35%	86%	97%
Mean over proteome*	86%			
Median over proteome*	89%			
Range over proteome*	41-100%			

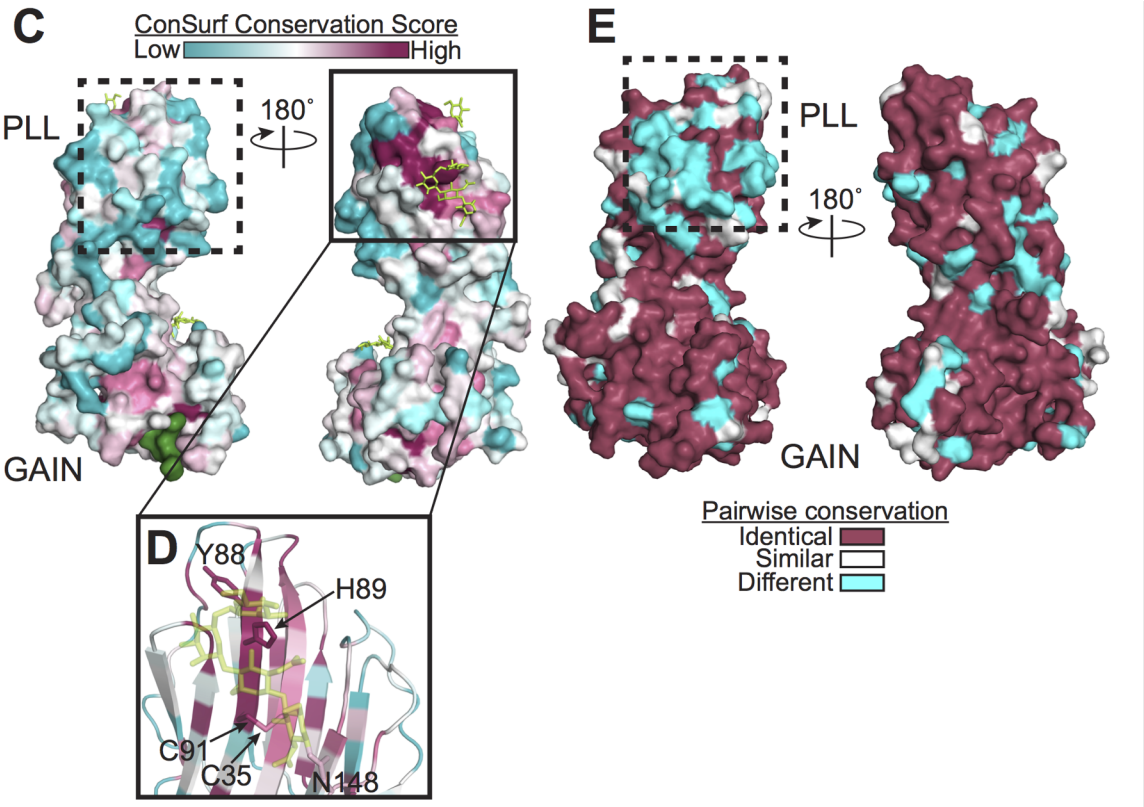
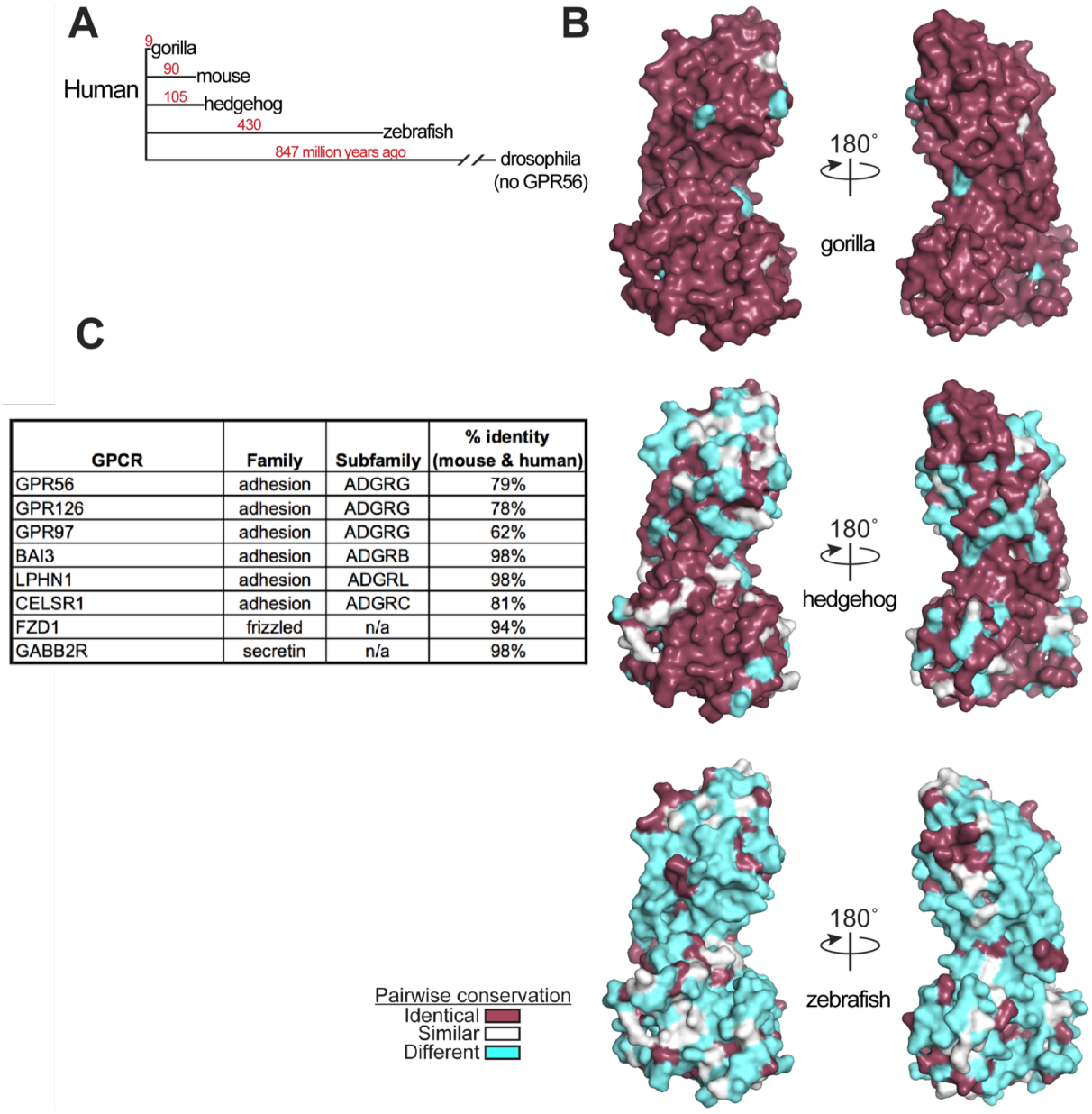


Figure IV.8 The two β-sheets of the PLL domain experienced different evolutionary pressures.

(A) Sequence alignment of a segment of the PLL domain from 12 species of GPR56. Conservation scores greater than 3 are shown at the top, with 9 representing the highest conservation. Residues found mutated in BFPP patients (Y88 and C91) are in red. C91, participating in the intra-PLL domain disulfide bond, is highlighted in yellow. H89 is highlighted in maroon. (B) Sequence identities for different fragments of GPR56 between human and the indicated species. *Taken from (237). (C) Conservation score of each residue is mapped on the GPR56 ECR structure. *Stachel* and N-linked glycans are shown as green surface and yellow sticks, respectively. Note that the conserved patch on β-sheet B (solid box, panel D) and the non-conserved patch on β-sheet A (dashed boxes in C and E) are the most and least



C

GPCR	Family	Subfamily	% identity (mouse & human)
GPR56	adhesion	ADGRG	79%
GPR126	adhesion	ADGRG	78%
GPR97	adhesion	ADGRG	62%
BAI3	adhesion	ADGRB	98%
LPHN1	adhesion	ADGRL	98%
CELSR1	adhesion	ADGRC	81%
FZD1	frizzled	n/a	94%
GABB2R	secretin	n/a	98%

Figure IV.9 Pairwise surface conservation analysis of the PLL domain.

(A) Cartoon illustrating approximate time since last common ancestor with human. (B) Pairwise surface conservation analysis of human GPR56 ECR vs gorilla, hedgehog, and zebrafish mapped to the structure of mouse GPR56 ECR. (C) Percent identity between full length human and mouse protein sequences of GPCRs from several families.

The two β-sheets of the PLL domain experienced different evolutionary pressures

In order to characterize the evolution of the PLL domain, GPR56 protein sequences from various species were aligned (Figure IV.8A). The PLL domain of GPR56 has lower sequence identity (65% between mouse and human) when compared with those for the GAIN (79%) and 7TM domains (88%), suggesting the individual domains of GPR56 have evolved to different degrees across species (Figure IV.8B). Furthermore, these analyses revealed that even the two β-sheets of the PLL domain have strikingly different sequence conservation across species: β-sheet B had 77% sequence identity between mouse and human, while β-sheet A had only 53%. This distinctly lower conservation suggests that β-sheet A has evolved rapidly since the divergence of mouse, human, and hedgehog from their common ancestor about 105 million years ago (Figure IV.9A, Table IV.6).

Functional sites (i.e. natural ligand-binding sites) are often highly conserved patches on protein surfaces. Thus, we performed surface conservation analysis of 102 GPR56 protein sequences from diverse organisms using the ConSurf server (166) to identify any putative functional sites on the GPR56 ECR. When the conservation score for each residue was mapped onto the GPR56 ECR structure (Figure IV.8C), the largest and most obvious conserved patch on the entire ECR was formed by a group of residues (G86, Y88, H89, and G106) located on β-sheet B of the PLL domain (Figure IV.8D). The residues involved in the conserved intra-PLL domain disulfide bond between strands β1 and β6 (C35 and C91) also contribute to this conserved patch. Finally, a glycan with a conserved N-linked glycosylation motif (N148-X-S150) sits adjacent to this patch. Notably, the side chain of H89 points out of β-sheet B into the

solvent. Pairwise surface conservation analysis between human GPR56 and protein sequences from various organisms (gorilla, mouse, hedgehog, and zebrafish) shows that this patch is indeed highly conserved, even in zebrafish, the earliest known organism with *GPR56* in its genome (Figure IV.8E and Figure IV.9A-B). Thus, we speculated that this conserved patch, particularly H89, has an essential role in GPR56 function (see below for *in vitro* and *in vivo* analysis).

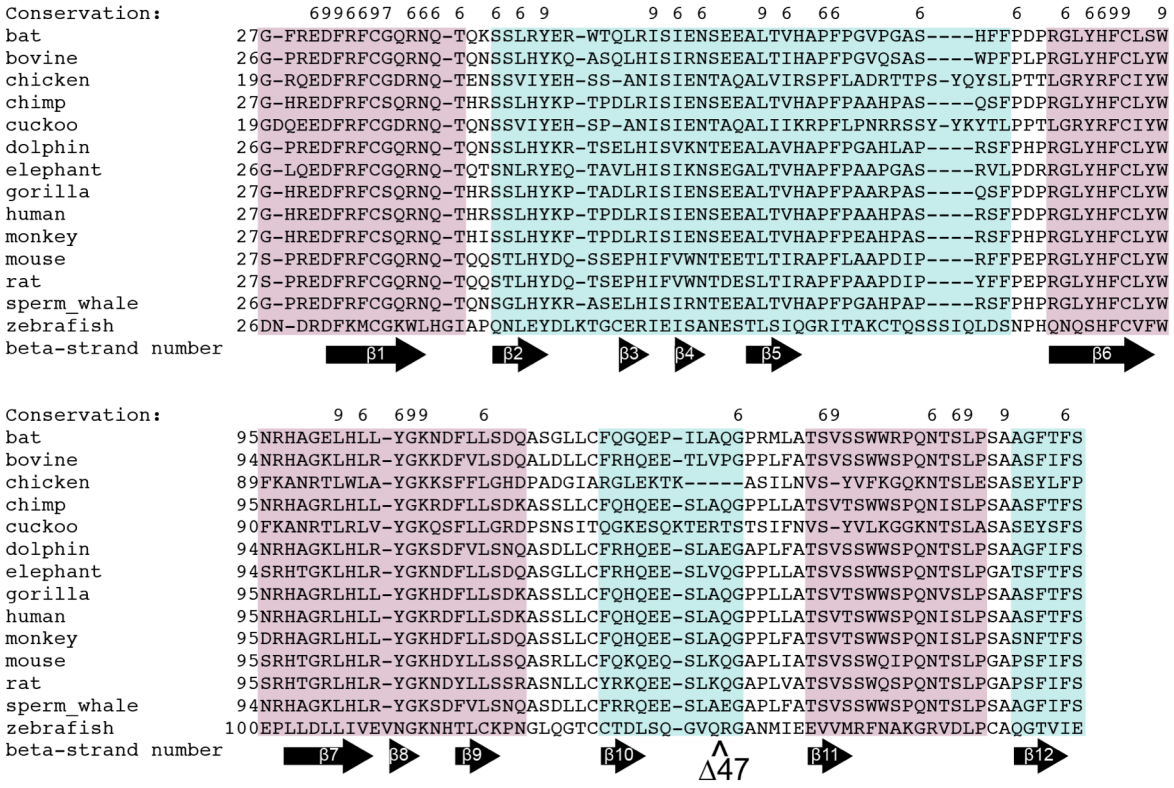


Table IV.6 Multiple sequence alignment of the GPR56 PLL domain from 14 species.

Conservation score is calculated for positions with at least some detectible conservation (9 is most conserved). The β-strands were identified and numbered, and the β-sheets were colored based on the crystal structure (β-sheet A, the more divergent from PTX and LNS domains is colored cyan; β-sheet B, the more conserved with PTX and LNS domains is colored maroon). A 47-residue insertion in zebrafish Gpr56 was removed from this alignment for clarity (Δ47).

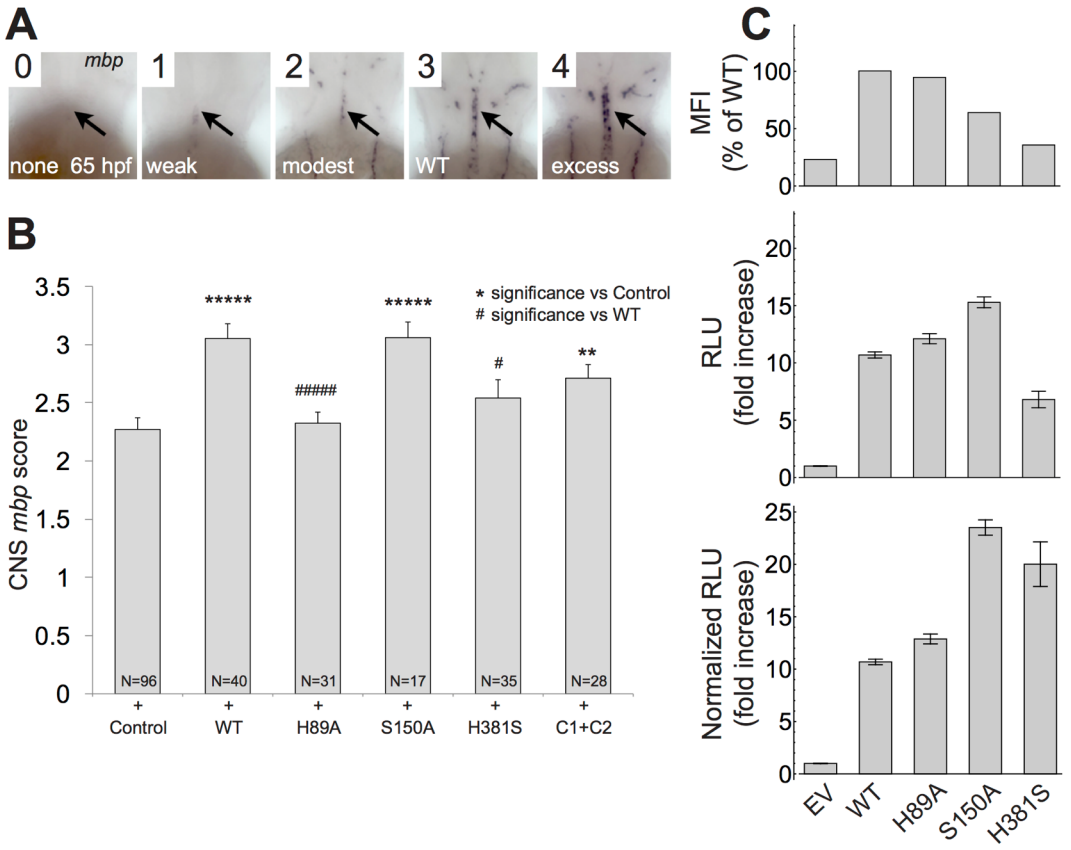


Figure IV.10 The conserved patch of the PLL domain is required for GPR56 function *in vivo*.

(A) Injection of WT and mutant mouse *GPR56* mRNAs generate embryos with varied levels of CNS *myelin basic protein* (*mbp*) expression (black arrow, hindbrain) at 65 hours post-fertilization (hpf). Embryos were given the following scores to signify (0) none, (1) weak, (2) modest, (3) WT, and (4) excess CNS *mbp* expression. (B) Average CNS *mbp* score (± S.E.M.) for phenol red (+ Control), WT, H89A, H381S, S150A, and C121S+C177S (C1+C2) injected embryos (from left to right). Injection of WT *GPR56* causes an increase in CNS *mbp* score compared to control-injected embryos ($p<7.56x10^{-6}$). H381S and H89A abolish the effect of GPR56 overexpression on CNS *mbp* expression (no significant difference from control injected, significantly less than WT injected: H381S, $p<.02$; H89A, $p<2.68x10^{-5}$). S150A and C1+C2 do not affect GPR56-induced CNS *mbp* overexpression (versus control injected: S150A, $p<3.78x10^{-5}$; C1+C2, $p<.005$). (C) Cell surface expression and basal activity of mutant GPR56 constructs as measured by the SRE-luciferase reporter assay. Top: Cell-surface expression of untagged WT and mutant GPR56 constructs with affinity for α5 comparable to WT measured by flow cytometry. Middle: Basal activity of mutant constructs. Bottom: Basal activity of mutant constructs normalized for cell-surface expression using MFI from Top. Data are presented as mean ± S.E.M.; $n = 3$.

A residue in the conserved patch of the PLL domain is critical for oligodendrocyte development in zebrafish

We hypothesized that the residues comprising the aforementioned conserved patch of the PLL domain are involved in an evolutionarily conserved function of GPR56. To dissect this *in vivo*, we tested GPR56 point mutants using zebrafish. Briefly, in zebrafish, Gpr56 promotes oligodendrocyte proliferation in the CNS, such that loss of this aGPCR results in reduced numbers of mature oligodendrocytes and myelinated axons (78). Gpr56 activity in zebrafish can be readily measured by assessing the expression of *myelin basic protein* (*mbp*), which encodes a structural component of the myelin sheath. Importantly, transient expression of mouse *GPR56* mRNA increases *mbp* expression above WT levels as reported previously (78) and confirmed in the present study (Figure IV.10A-B).

We tested the following mouse GPR56 point mutations in this assay: H89A, S150A, H381S, and C121S+C177S (C1+C2), described in Figure IV.5A. Strikingly, injection of mRNA encoding the mouse GPR56 H89A mutant failed to enhance *mbp* expression, suggesting an essential role of this evolutionarily conserved residue in CNS myelination (Figure IV.10B). The H381S mutant also failed to enhance *mbp* expression, suggesting a possible role for receptor autoproteolysis in GPR56-dependant oligodendrocyte development, consistent with previous studies that implicate autoproteolysis in GPR56 function (86, 91). On the other hand, S150A and C1+C2 resulted in a significant increase in *mbp* expression, similar to injection of WT *GPR56* (Figure IV.10B). To ensure the *in vivo* effects of these mutants were not simply due to mutation-dependent cell-surface expression, we quantified surface expression and SRE

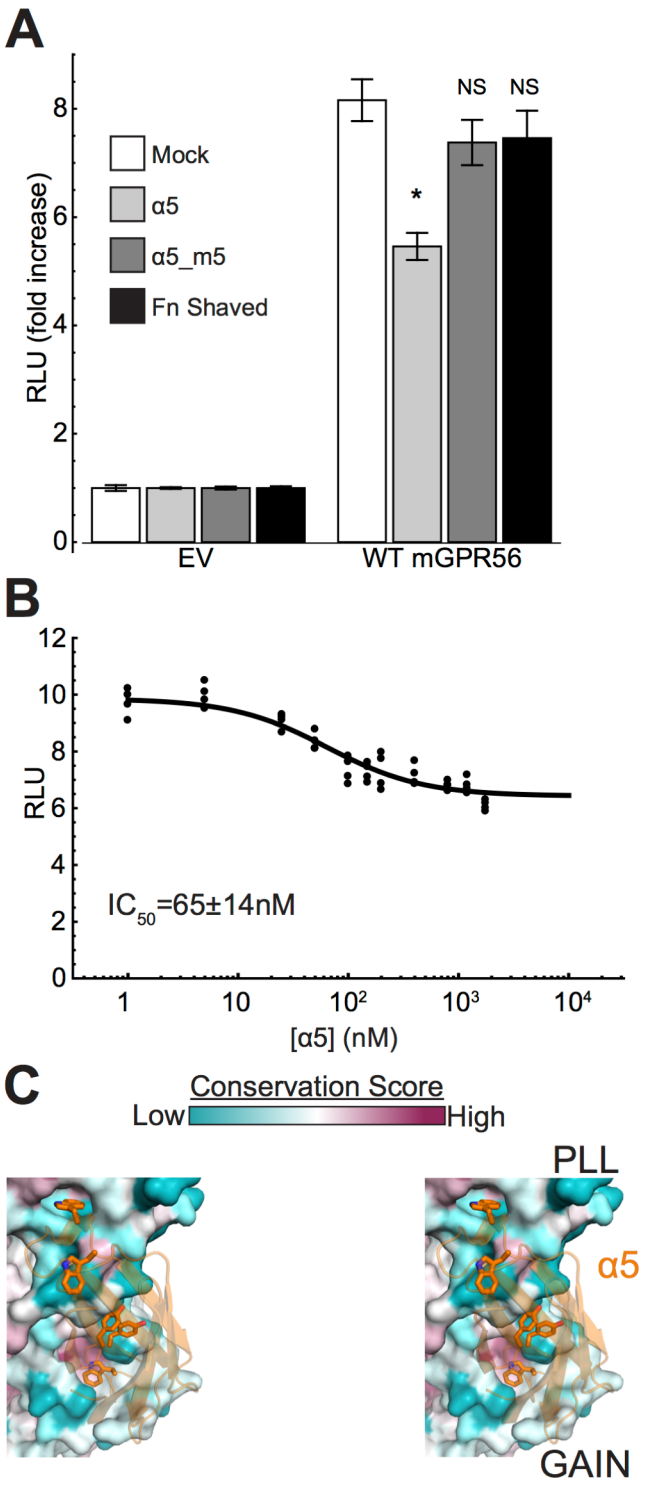


Figure IV.11 Monobody α5 is an allosteric inverse-agonist for GPR56.

(legend on the following page)

Figure IV.11, continued.

A) Effect of 1 μM monobody on GPR56 activation as measured by SRE-luciferase assay in HEK293T cells. Data are presented as mean ± S.E.M.; *n* = 3; *, p<0.01 compared to Mock by two-tailed Student's *t*-test; NS, not significant compared to Mock. (B) SRE-luciferase activity in HEK293T cells is plotted as a function of α5 concentration. Line represents the best fit of the 1:1 binding model for calculation of IC_{50}. (C) Stereo image of the interface between α5 and GPR56 ECR. ECR is colored by conservation score, as in Figure 4C. α5 residues important for GPR56 binding (Figure IV.3D) shown as sticks.

signaling in HEK293T cells. We found H89A had no effect on surface expression or basal activity as compared to WT GPR56, suggesting any differences between WT and H89A phenotypes *in vivo* are not due to cell-surface expression or basal activity Figure IV.10C). Altogether, these results reveal that the conserved patch of the PLL domain mediates an essential function in CNS myelination.

α5 monobody is an allosteric inverse-agonist for GPR56

We hypothesized that since α5 interacts with the GAIN and PLL domains, both shown to regulate signaling, α5 may itself modulate GPR56 basal activity. We found that addition of α5 to the SRE-luciferase signaling assay causes a ~25% decrease in GPR56 basal activity with an IC50=65 ± 14 nM (Figure IV.11A-B), the same order of magnitude as the measured affinity of purified α5 to GPR56-expressing HEK293T cells (KD=17±2nM, Figure S2C). The addition of an unrelated, non-binding monobody, or α5_m5 (double Tyr to Ala mutant in the FG loop with >100-fold decreased affinity for GPR56 [Figure IV.1E and Figure IV.3D]), produced no significant effect on GPR56 basal activity (Figure 6A). In addition, α5 had no effect on the background luminescence

of HEK293T cells transfected with an empty vector (Figure IV.11A), demonstrating the specificity of α5 in this assay. The binding site of α5 is poorly conserved among mouse, human, and zebrafish GPR56 (Figure IV.11C), consistent with the observation that α5 does not detectibly interact with human or zebrafish GPR56 (Figure IV.1E). Furthermore, as expected from the crystal structure, α5 binds to full-length mouse GPR56 and purified soluble mouse GPR56 ECR, but it does not detectibly interact with N-terminally truncated mouse GPR56 constructs expressed in HEK293T cells including ΔPLL (Figure IV.6). Thus, the entire ECR is necessary for α5 binding, indicating that any effect on GPR56 activity mediated by α5 is due to its interaction with the ECR and not the 7TM, the canonical site for GPCR ligand interaction. Therefore, α5 represents an 'allosteric inverse-agonist' for GPR56 (34).

Discussion

CNS myelination likely requires GPR56 activation by a PLL-binding ligand

Our study yielded the first crystal structure of the full ECR of an aGPCR and provides a functional framework to understand the molecular mechanisms by which aGPCR ECRs govern receptor function. The crystal structure revealed a previously unidentified PTX/LNS-Like (PLL) domain at the N-terminus of the GPR56 ECR (Figure IV.4). Both PTX and LNS domains predominantly occur in secreted proteins and in ECRs of cell-surface proteins. LNS domains in particular are mostly known for their adhesion properties, especially in the brain, and exist in adhesion molecules such as agrins, laminins, and neurexins to mediate cell-cell and cell-extracellular matrix interactions (167–169). Thus, the remote similarity of the PLL domain to PTX and LNS

domains supports an adhesion-related role for the PLL domain in GPR56 as was previously suggested (74) (Figure IV.4A-B, Table IV.3, and Table IV.4). Future studies focusing on biochemical and structural characterization of ECR-ligand interactions will provide more insight into the pathophysiological role of GPR56.

The suggested 'shedding' mechanism for aGPCR activation involves engagement of an aGPCR ECR by an extracellular ligand, which results in separation of the NTF from the CTF and initiation of G protein signaling (35, 147). On the other hand, recent studies suggest that the non-shed NTF of GPR56 may negatively regulate signaling by interacting with the 7TM directly (91). Importantly, ligand binding may activate GPR56 through each of these mechanisms individually or perhaps in concert (Figure IV.13). It is widely accepted that GPR56 signals through SRE via RhoA and $G\alpha_{13}$ (35, 57, 86) and it has been shown that transient expression of constitutively active RhoA can suppress myelination defects in *gpr56* mutant zebrafish (78). We discovered a surface-exposed conserved patch on the PLL domain that is necessary to promote CNS myelination *in vivo*, but that does not affect basal activity *in vitro* (Figure IV.8 and Figure IV.10). Thus, as is common for highly conserved patches, we speculate that this patch on the PLL domain directly engages a GPR56 ligand such as collagen III, TG2, or an as yet unidentified ligand. Indeed, truncation-based analyses suggest that the regions of GPR56 responsible for binding TG2 and collagen III are within residues 108-177 and 27-160, respectively, both of which map to the PLL domain (170, 171). Together, these observations are consistent with the hypothesis that CNS myelination is dependent on GPR56 activation induced by a PLL domain-binding ligand (Figure IV.13).

Alternative splicing restricts PLL domain expression

The crystal structure of the GPR56 ECR defined the boundaries of the PLL and GAIN domains, and, remarkably, showed that S4, a *GPR56* splice variant present in both human and mouse, encodes an isoform that lacks the entire PLL domain, precisely

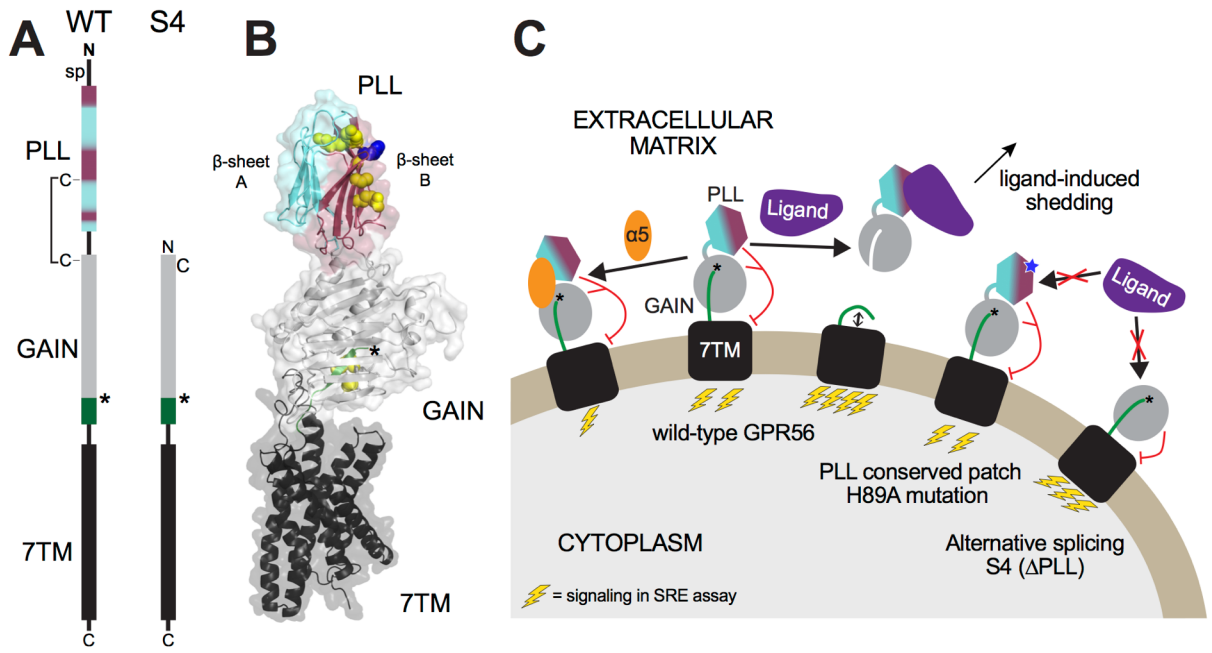


Figure IV.12 Working model of mechanisms underlying GPR56 function.

(A) Schematic of GPR56 domain structure comparing WT and S4. In all panels, the PLL domain is colored cyan and maroon, corresponding to β-sheets A and B, respectively. (B) Scale model of full-length GPR56 based on the crystal structure of the ECR and a model of the 7TM (generated based on GCGR structure, PDB ID: 46LR). An arbitrary orientation of the ECR with respect to the 7TM is chosen. Residues mutated in BFPP are shown as yellow spheres. H89 is shown as blue spheres. (C) A working model of aGPCR signaling involves ligand-induced activation. In this model, full-length GPR56 is activated when a natural ligand binds to the conserved patch on the PLL domain including H89, causing conformational changes, perhaps including shedding. Introducing the H89A mutation (blue star) to the conserved patch of the PLL domain or deleting the PLL domain completely (as in S4) would result in abrogation of ligand binding and therefore no ligand-induced activation. Binding of α5 likely stabilizes the ECR, causing decreased signaling. sp, signal peptide. See Figure S6 for further possibilities.

starting with the first residue of the GAIN domain (Figure IV.5A-B and Figure IV.12A). Intriguingly, in the originally described *GPR56* knock-out mouse, which displays phenotypes of cortical neuronal ectopia and impaired oligodendrocyte development, the expression of the full-length GPR56, but not the S4 transcript, is disrupted (Figure IV.5B) (73, 77). This demonstrates that GPR56 S4 expression alone is not sufficient for normal CNS development and suggests that the PLL domain has a critical role in both cortical development and oligodendrocyte development. We speculate that observations of a recently published null knock-out mouse with all GPR56 isoforms deleted will reveal distinct, and perhaps more severe phenotypes (77).

Taken together, our results support a model in which domains in the ECR directly or indirectly regulate the distinct but interrelated functions of GPR56: a possible adhesion function mediated by the PLL domain and a G protein signaling function mediated by the 7TM domain (Figure IV.12 and Figure IV.13). With regard to mechanism, we show that ECR modification leads to altered basal activity by allosterically altering NTF shedding propensity and/or altering the conformational states sampled by the 7TM (*e.g.*, absence of the PLL domain leads to increased basal activity) (Figure IV.13). Moreover, by fine-tuning the expression levels of different GPR56 isoforms, cells may regulate their response to diverse extracellular ligands as well as their basal level of G protein signaling (Figure IV.12).

aGPCR inhibition by an ECR-directed synthetic allosteric inverse-agonist

The α5 monobody represents, to our knowledge, the first synthetic allosteric inverse-agonist that interacts with the GPR56 ECR with high affinity and specificity (Figure

IV.1B-F, Figure IV.11A-C, Figure IV.3, and Figure IV.6). Agonistic antibodies directed to GPR56 have been reported but their mechanisms of action are not fully understood (92, 172). The lack of well-characterized agonists and antagonists has hampered mechanistic studies of GPR56 and other aGPCRs. Our success in generating a modulator of GPR56 suggests that like the canonical GPCRs, aGPCRs, including GPR56, are realistically 'druggable'. Such synthetic modulators will help advance mechanistic analyses of aGPCRs. Our finding that α5 alters basal activity by binding to the ECR is an encouraging proof of concept for developing highly selective modulators of aGPCRs. A major challenge in GPCR-targeted drug design is the high conservation of 7TM, which demands high specificity of drugs so as to minimize undesirable side-effects (173). As aGPCR ECRs are much more diverse than 7TMs, the pursuit of aGPCR ECR-targeted (*i.e.*, allosteric) synthetic ligands, such as monobodies or antibodies, will likely result in highly specific reagents. Furthermore, as the therapeutic potential of allosteric GPCR modulators that exhibit moderate effects has been demonstrated (30, 32, 34, 174), this work validates the aGPCR ECR as an effective drug target.

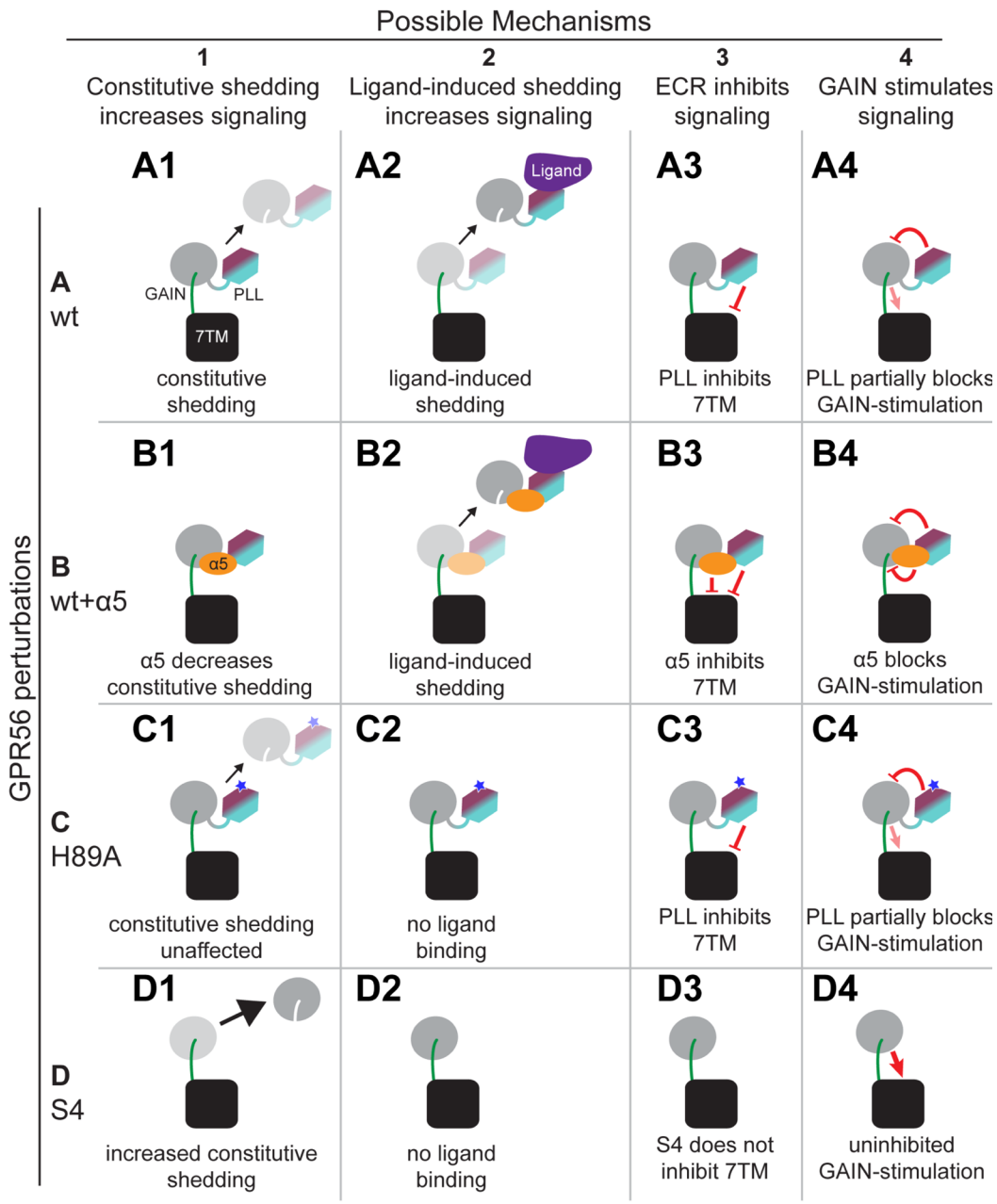


Figure IV.13 Detailed possible mechanisms leading to experimental observations of GPR56 activity.

(page 1 of 3; legend on the following pages)

Figure IV.13, continued.

Four possible mechanisms are proposed (numbered 1-4) that explain the observed phenomena. We do not rule out the possibility of additional mechanisms nor do we deny that more than one proposed mechanism may occur. In our experiments, four systems were tested, each corresponding to a GPR56 perturbation (labeled A-D). Black arrows represent NTF shedding. Pointed (stimulating G-protein signaling) and flat-ended (inhibiting G-protein signaling) red arrows may represent conventional binding interactions, weak and transient interactions, or random collisions.

Proposed mechanism 1 relies on the hypothesis that, even in the absence of ligand, there exists some small but significant constitutive level of shedding in GPR56. By this mechanism, the higher the level of constitutive shedding, the more G protein signaling occurs, due to the observation that the CTF of the receptor is highly active. (A1) wt GPR56 undergoes some level of constitutive shedding leading to the observed basal activity of the receptor. (B1) Upon addition of α5, the constitutive shedding decreases, leading to the observed decrease in G protein signaling. (C1) The H89A mutation on the conserved face of PLL domain does not alter constitutive shedding, resulting in the observation that H89A has similar basal activity to wt. (D1) The ECR of S4, corresponding to only the GAIN domain, has an increased level of constitutive shedding, leading to the observed high basal activity of the ΔPLL construct.

(page 2 of 3; continued on the following page)

Figure IV.13, continued.

Proposed mechanism 2 relies on the hypothesis that a ligand may bind the conserved face of the PLL domain, exert mechanical force, and thereby cause NTF shedding. Similar to mechanism 1, by this mechanism, with more ligand-induced shedding, more G protein signaling occurs, due to the observation that the CTF of the receptor is highly active. (A2) wt GPR56 undergoes some level of ligand-induced shedding. (B2) Upon addition of α5, the ligand-induced shedding is not affected. (C2) The H89A mutation on the conserved face of PLL domain abolishes ligand binding, and therefore blocks ligand-induced shedding. (D2) The ECR of S4 lacks the ligand-binding site on the PLL domain, and therefore does not undergo ligand-induced shedding.

Proposed mechanism 3 relies on the hypothesis that domains in the ECR directly or indirectly inhibit signaling by interacting in some way with the 7TM. (A3) The PLL domain in wt GPR56 partially inhibits 7TM. (B3) Together, the PLL domain and α5 strongly inhibit the 7TM. (C3) The H89A mutation on the conserved face of PLL domain does not affect PLL-mediated inhibition of the 7TM. (D3) The ECR of S4 lacks the inhibitory PLL domain, and therefore has higher basal activity than wt.

Proposed mechanism 4 relies on the hypothesis that the GAIN domain stimulates signaling by directly or indirectly interacting with the 7TM. (A4) The PLL domain in wt GPR56 partially inhibits the stimulatory GAIN domain. (B4) Together, the PLL domain and α5 strongly inhibit the stimulatory GAIN domain. (C4) The H89A mutation on the conserved face of PLL domain does not affect PLL-mediated inhibition of the GAIN domain. (D4) The ECR of S4 lacks the inhibitory PLL domain, and therefore has higher basal activity than wt.

(page 3 of 3)

CHAPTER V

Stachel-independent activation and inhibition of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region[1]

Summary

Adhesion G protein-coupled receptors (aGPCRs) play critical roles in diverse biological processes including neurodevelopment and cancer progression. aGPCRs are characterized by large and diverse extracellular regions (ECRs) that are autoproteolytically cleaved from their membrane-embedded signaling domains. Although ECRs regulate receptor function, it is not clear whether ECRs play a direct regulatory role in G protein signaling or simply serve as a protective cap for the activating "*Stachel*" sequence. Here, we present a mechanistic analysis of ECR-mediated regulation of GPR56/ADGRG1, an aGPCR with two domains (PLL and GAIN) in its ECR. We generated a panel of high-affinity monobodies directed to each of these domains from which we identified activators and inhibitors of GPR56-mediated signaling. Surprisingly, these synthetic ligands modulated signaling of a GPR56 mutant

[1]The text of this section is copied verbatim (with minor changes) from:
Salzman GS, Zhang S, Gupta A, Koide A, Koide S, and Araç D. (2017) Stachel-independent activation and inhibition of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. Proceedings Nat Acad Sci USA: in revision. G.S.S, A.K., S.K., and D.A. designed experiments. G.S.S. cloned GPR56 constructs, purified GPR56 fragments, engineered monobodies, and performed signaling assays. G.S.S and S.Z. purified monobodies. G.S.S. and A.G. performed binding assays. A.K. generated the monobody library. G.S.S., S.K., and D.A. analyzed data and wrote the manuscript.

defective in autoproteolysis and hence in *Stachel* peptide exposure. These results provide compelling support for a ligand-induced and ECR-mediated mechanism that regulates aGPCR signaling in a transient and reversible manner, which occurs in addition to the *Stachel*-mediated activation.

Introduction

The G protein-coupled receptor (GPCR) superfamily exhibits great diversity with regard to the length and complexity of the extracellular region (ECR). Landmark mechanistic and functional studies of GPCRs to date have almost exclusively focused on receptors without prominent extracellular domains, particularly those from the rhodopsin family (24–26). In contrast, receptors from the adhesion, secretin and frizzled/taste2 families, and even some from the rhodopsin family, have one or more extracellular domains (22). Members of the relatively poorly characterized adhesion GPCR (aGPCR) family are characterized by particularly diverse and large ECRs: hundreds to thousands of amino acid residues compose multiple protein domains (45, 46). Although their spatial proximity to the seven-pass transmembrane helices (7TM) region suggests potentially important roles for these complex ECRs in GPCR signaling (40), their functions are incompletely understood.

aGPCRs are expressed in many tissues and have been linked to myriad biological and physiological processes ranging from the establishment of ovarian cell polarity (Celsr1/ADGRC1) (175) to synapse formation (Lphn3/ADGRG3) (61) to regulation of lung surfactant production (GPR116/ARGRF5) (176). For example, the aGPCR GPR56/ADGRG1 is involved in cortex development, oligodendrocyte

development, muscle cell development, innate immunity, and cancer progression (70, 77, 78, 81, 82, 84, 148). Recent studies have highlighted the role of GPR56 in promoting progression of acute myeloid leukemia (AML) (49) and progastrin-dependant colon cancer (48) and suggested that a GPR56 inhibitor would be clinically desirable. A mechanistic understanding of the biology mediated by aGPCRs, and their ECRs in particular, will be a critical milestone on the path to treating aGPCR-mediated pathologies.

The aGPCR ECRs are characterized by the presence of a conserved juxtamembrane GPCR Autoproteolysis INducing (GAIN) domain (55) and various

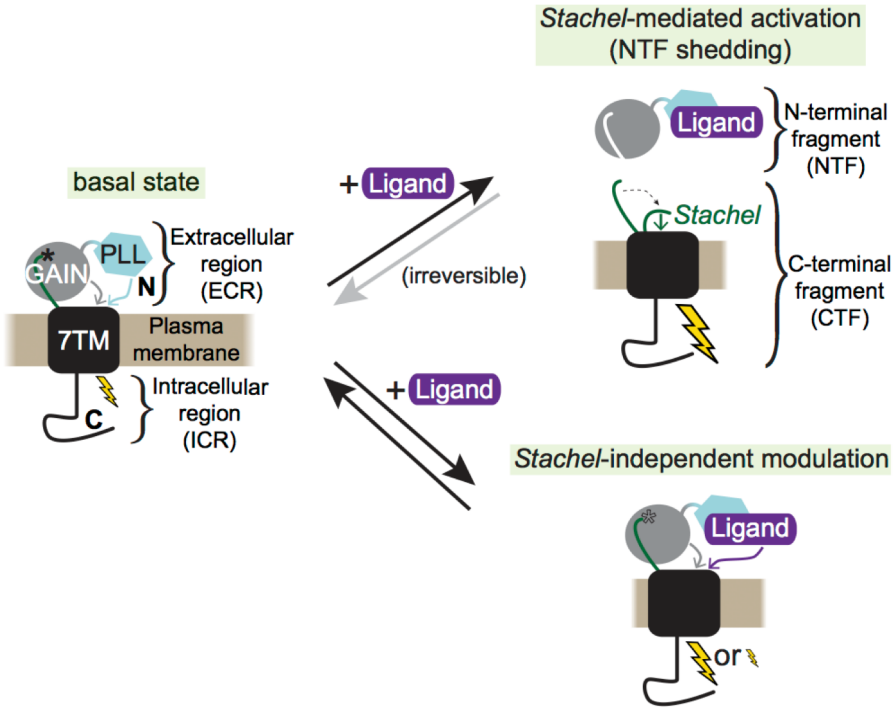


Figure V.1 Models for ligand-induced GPR56 G protein signaling.

Autoproteolysis site indicated by *. Unneeded autoproteolysis indicated by outlined *. Lightning bolt size represents signaling intensity. Colored arrows represent proposed regulation of 7TM signaling by the ECR.

adhesion-type domains (located N-terminal to the GAIN domain), which allow aGPCRs to bind protein ligands (45, 68). An autoproteolytic event occurs within the GAIN domain during aGPCR maturation, cleaving the receptor into two fragments: an N-terminal fragment (NTF; composed of the N-terminal adhesion domains and the majority of the GAIN domain) and a C-terminal fragment (CTF; composed of the C-terminal β-strand of the GAIN domain, termed "*Stachel"* or "stalk", the 7TM, and the intracellular region) (Figure 1). After cleavage, the NTF and CTF remain noncovalently but tightly associated throughout trafficking and localization to the plasma membrane (55, 90). The conservation of the GAIN domain suggests it plays a role in aGPCR function.

Preliminary studies have proposed that GPCR ECRs regulate receptor functions, likely including G protein signaling, upon binding to extracellular ligands (42–44, 61, 86, 88, 130, 177). Two complementary models for ligand-induced aGPCR activation have been proposed (Figure V.1). In the *Stachel*-mediated model, the NTF serves as a protective cap for the *Stachel* and has no direct role in modulating 7TM function. Upon ligand binding to an N-terminal adhesion domain, the NTF dissociates from the CTF, termed "shedding", exposing the *Stachel* to function as a "tethered agonist" (35, 91, 147, 149). Key to this model is GAIN domain autoproteolysis, a necessary reaction to precede shedding and *Stachel* exposure. Though it has been proposed that natural ligands may induce shedding upon binding to N-terminal adhesion domains and thereby activate the receptor, direct proof of ligand-induced shedding remains elusive. Several recent observations, including that some aGPCRs do not undergo autoproteolysis and therefore cannot undergo shedding (55, 56), have necessitated the introduction of a *Stachel*-independent model, in which the ECR has a direct role in modulating the 7TM

signaling (90–93). In this model, the ECR directly communicates with the 7TM (i.e. via transient interactions), such that ligand binding events or conformational changes in the ECR may directly result in altered signaling. Direct proof of this model has also remained elusive. A major bottleneck in discriminating these mechanisms is a lack of high-affinity, water-soluble ligands that can perturb aGPCR function in a well-controlled manner. Although natural ligands have been identified for several aGPCRs, most of them are not suitable for quantitative assays.

GPR56 is among the better-characterized members of the aGPCR family. It has a 377-residue ECR composed of two domains: an N-terminal Pentraxin and Laminin/neurexin/sex hormone-binding globulin-Like (PLL) domain and a GAIN domain (93). Previously, we have shown that deletion of the PLL domain increases basal activity of the receptor (93). Additionally, we engineered a binding protein, termed monobody, that targets the ECR of mouse GPR56, bridges the PLL and GAIN domains, and functions as an allosteric inverse agonist of G protein signaling. Though both of these findings support ECR-mediated regulation of signaling, mechanistic detail was lacking.

In the present study, we set out to elucidate the regulatory mechanism of aGPCR signaling by ligands to the ECR. To this end, we developed a panel of new monobodies that target specific extracellular domains of human and mouse GPR56, and identified an activator and an inhibitor of human GPR56 among these monobodies. Based on the activity of these synthetic ligands on an autoproteolysis-defective and thus shedding-defective receptor, our results provide support for *Stachel*-independent regulation of GPR56 signaling mediated by the ECR.

Results

Monobodies targeted to GAIN and PLL domains of mouse and human GPR56

Guided by the new knowledge of the structure of the ECR and the autoinhibitory role of the PLL domain gained from our previous work (93), we hypothesized that ligands that engaged different regions within the ECR would differentially affect GPR56 signaling. In addition, we were interested in developing a detection reagent for splice variant 4 (S4) whose ECR is composed of only the GAIN domain as a consequence of alternative splicing leading to deletion of the N-terminal 175 residues (including the signal peptide, PLL domain, and PLL-GAIN linker; Figure V.2A). As we have shown that PLL domain deletion results in increased basal activity (93), we were particularly interested in facilitating the study of S4 expression and function. Taken together, we set out to engineer a diverse panel of monobodies for the ECRs of human and mouse GPR56.

Because the ECRs of mouse and human GPR56 are not highly conserved (73% amino acid sequence identity), we anticipated that most monobodies would not cross-react with both human and mouse GPR56. Indeed, this was the case for the α5 monobody that we generated previously (93). Consequently, we decided to perform monobody selection separately for human and mouse GPR56 samples. We prepared the purified samples of the full-length ECRs as well as individual PLL and GAIN domains of both mouse and human GPR56. Although the single domain constructs have exposed hydrophobic surfaces that would be sequestered in the interface between

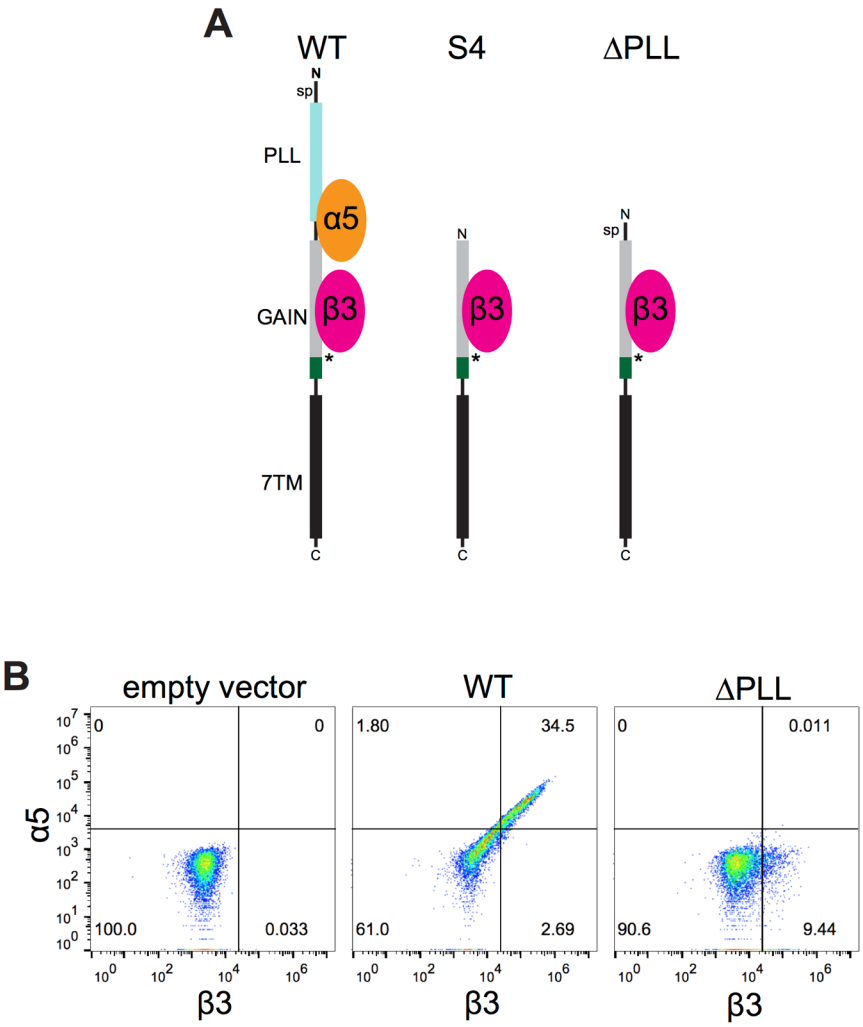


Figure V.2 Proposed experimental setup to identify the endogenous protein product of GPR56 S4.

(A) Domain architecture schematics illustrating that S4 and ΔPLL only differ by the presence of an N-terminal signal peptide on ΔPLL. Thus, the mature protein products of these two transcripts would be identical. The binding sites of α5 and β3 are illustrated. *, autoproteolysis site; sp, signal peptide. (B) Flow cytometry experiment of HEK293T cells transfected with EV or mouse GPR56 (WT or ΔPLL) constructs. Co-staining with labeled α5 and β3 distinguishes WT GPR56 from ΔPLL. A similar experiment may distinguish WT GPR56 from endogenous S4.

the PLL and GAIN domains, they were highly soluble and predominantly monomeric. Using a total of six samples as antigens we carried out phage- and yeast-display selection of monobodies as previously described and identified 19 new clones (93, 114)

Full clone name	Abbr. clone name	Sequence
Mb(hGPR56_β1)	β1	VSSVPTKLEVVAATPTSLLISWDAPAVTVDFYVITYGETGGNWYAAQEFTVPGSKSTATISGLKPGVDYTITVYAYPDHHYQGRSPISINYRT
Mb(hGPR56_β2)	β2	VSSVPTKLEVVAATPTSLLISWDAPAVTVDFYIITYGETGGSWYSSQEFAVPGSKSTATISGLKPGVDYTITVYASMPGSWYYSPISINYRT
Mb(hGPR56_β3)	β3	VSSVPTKLEVVAATPTSLLISWDAPAVTVDFYVITYGETGSGWFPGQTFEVPGSKSTATISGLKPGVDYTITVYTYGYSSLGPGSPISINYRT
Mb(hGPR56_β4)	β4	VSSVPTKLEVVAATPTSLLISWDAPAVTVDFYVITYGETGHGWFPGQTFEVPGSKSTATISGLKPGVDYTITVYAFYPRSSRPSPISINYRT
Mb(hGPR56_β5)	β5	VSSVPTKLEVVAATPTSLLISWDAPAVTVDHYVITYGETGVGWVPGQTFTVPGSKSTATISGLKPGVDYTITVYAWNASIFSYSPISINYRT
Mb(hGPR56_β6)	β6	VSSVPTKLEVVAATPTSLLISWDAPAVTVDHYVITYGETGVGWVPGQTFTVPGSKSTATISGLKPGVDYTITVYAYSEWSYFVINPISINYRT
Mb(hGPR56_β7)	β7	VSSVPTKLEVVAATPTSLLISWDAPAVTVYYVITYGETGHGGYYYQEFKVPGSKSTATISGLKPGVDYTITVYAYDDEYSSSPISINYRT
Mb(hGPR56_β8)	β8	VSSVPTKLEVVAATPTSLLISWDAPAVTVDLYYITYGETGMWYPSSYQEFAVPGSKSTATISGLKPGVDYTITVYAESGWGYDVSSPISINYRT
Mb(hGPR56_β9)	β9	VSSVPTKLEVVAATPTSLLISWDAPAVTVDYYVITYGETGGSWYGWQEFAVPGSKSTATISGLKPGVDYTITVYAYPDHHYQGRSPISINYRT
Mb(mGPR56_β10)	β10	VSSVPTKLEVVAATPTSLLISWDAPAVTVFFYFITYGETGGNSPVQKFTVPGSKSTATISGLKPGVDYTITVYALYRSQKSGQYDYSSPISINYRT
Mb(mGPR56_β11)	β11	VSSVPTKLEVVAATPTSLLISWDAPAVTVVLYVITYGETGGNSPVQEFTVPGSKSTATISGLKPGVDYTITVYAQYESGTWLYRGSPISINYRT
Mb(mGPR56_β12)	β12	VSSVPTKLEVVAATPTSLLISWDAPAVTVDFYFITYGETGWGYGSYQAFEVPGSKSTATISGLKPGVDYTITVYAYYDSQRFLHSGSPISINYRT
Mb(mGPR56_β13)	β13	VSSVPTKLEVVAATPTSLLISWDASSSSVSYYRITYGETGGNSPVQEFTVPGSSSTATISGLKPGVDYTITVYAQSGPYYWYWGDSPISINYRT
Mb(hGPR56_β14)	β14	VSSVPTKLEVVAATPTSLLISWDATGYYYRYYRITYGETGGNSPVQEFTVPGSSSTATISGLKPGVDYTITVYAQSGPYYWYWGDSPISINYRT
Mb(mGPR56_β15)	β15	VSSVPTKLEVVAATPTSLLISWDASSSSVSYYRITYGETGGNSPVQEFTVPGSSSTATISGLKPGVDYTITVYAGVGNYKYWWGSSPISINYRT
Mb(hGPR56_β16)	β16	VSSVPTKLEVVAATPTSLLISWDANYYYSYGDVIYYRITYGETGGNSPVQEFTVPYYYSTATISGLKPGVDYTITVYAYDEYYTYGWSSPISINYRT
Mb(mGPR56_β17)	β17	VSSVPTKLEVVAATPTSLLISWDAMKNDEDVQYYRITYGETGGNSPVQEFTVPGSSSTATISGLKPGVDYTITVYAGVSSYYYYWGSSPISINYRT
Mb(hGPR56_β18)	β18	VSSVPTKLEVVAATPTSLLISWDAPAVTVDYYVITYGETGVGWVPGQTFEVPGSKSTATISGLKPGVDYTITVYAYHEYYFISPISINYRT
Mb(hGPR56_β19)	β19	VSSVPTKLEVVAATPTSLLISWDAPAVTVDFYYITYGETGSSYWSYQEFTVPGSKSTATISGLKPGVDYTITVYAIDQWQYYYYEMGSPISINYRT

Table V.1 Clones obtained in domain-specific human and mouse GPR56 monobody engineering campaign.

Full Name	Abbreviated name	GPR56 fragment necessary for interaction (species)	Function in SRE assay
Mb(mGPR56_α5)*	α5	ECR (mouse)	allosteric inverse-agonist
Mb(hGPR56_β1)	β1	ECR (human)	allosteric inverse-agonist
Mb(hGPR56_β3)	β3	GAIN (human and mouse)	neutral allosteric ligand
Mb(hGPR56_β7)	β7	PLL (human)	allosteric agonist
Mb(mGPR56_β12)	β12	PLL (mouse)	neutral allosteric ligand

Table V.2 Summary of GPR56-binding monobodies.

*First published in Salzman et al., 2016 (93).

(Table V.1). These included human GPR56-specifc clones that bind to the PLL domain but not the GAIN domain (e.g. Mb(hGPR56_β7)), to the GAIN domain but not the PLL domain (e.g. Mb(hGPR56_β6), and to full-length ECR but not the PLL or GAIN domain in isolation (e.g. Mb(hGPR56_ β1); Figure V.3A). We obtained similar clones specific to mouse GPR56 (Figure 2A, Table V.2). Surprisingly, we also identified monobodies that bind the human and mouse GAIN domains (e.g. Mb(hGPR56_β3); Figure V.3A), although our selection strategy was not designed to enrich such cross-reactive clones. Four particularly interesting clones will be the focus hereafter: Mb(hGPR56_β1), Mb(hGPR56_β3), Mb(hGPR56_β7), and Mb(mGPR56_β12), hereafter abbreviated β1, β3, β7, and β12, respectively. Their properties are summarized in Figure V.3, Figure V.4, Table V.2 and Table V.3.

Given our interest in facilitating S4 characterization, we set out to test if a GAIN domain-specific monobody, β3, could detect S4 on the cell surface. To this end, we expressed a ΔPLL mouse GPR56 construct, equivalent to the predicted protein

product of S4, in HEK293T cells (93) and co-stained cells with α5 and β3. We measured a robust increase in β3 binding signal for both WT and ΔPLL GPR56 as compared with empty vector, strongly suggesting β3 detects the predicted protein product of S4 (Figure V.2B). As there are no other reagents, to our knowledge, that specifically detect the truncated ECR of S4, β3 may prove useful in future studies of S4 expression.

New monobodies modulate G protein signaling

We examined whether the new monobodies affect activity of GPR56 using a serum response element (SRE)-luciferase assay that measures signaling via $G\alpha_{13}$ (35, 93) (Figure V.5A-B). Monobody β1 decreased hGPR56 signaling (Figure V.5A). In contrast, β1 treatment of mGPR56- or empty vector-transfected cells resulted in no detectible change in signaling, demonstrating its specificity for human GPR56 (Figure V.5A-B). β1 decreased hGPR56 signaling with IC_{50} of 70 ±30 nM, resulting in a ~1.6-fold decrease in signaling relative to the basal activity (Figure V.5C). Notably, β1 had a profile similar to α5, a previously reported monobody directed to mouse GPR56, in that both bound the full ECR but not the isolated GAIN or PLL domain, and functioned as an allosteric inverse agonist of G protein signaling (34, 93).

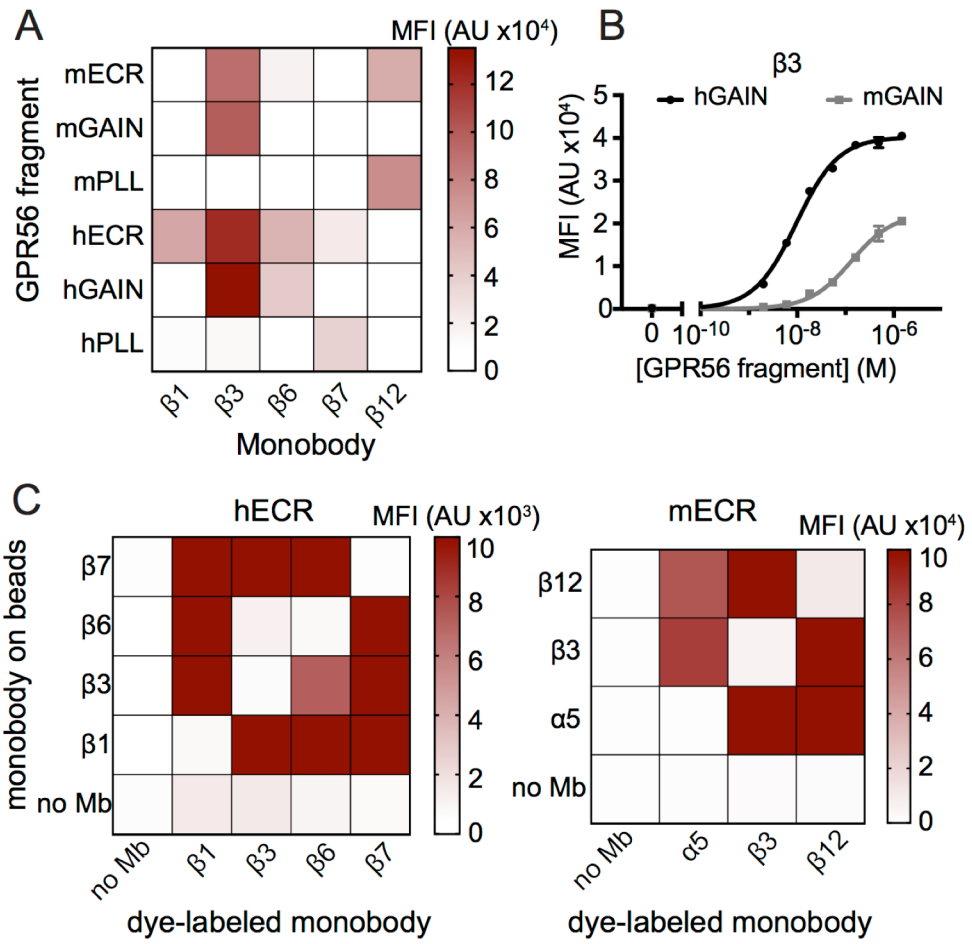


Figure V.3 Characterization of domain-specific GPR56-binding monobodies.

(A) 1.1 μM purified human and mouse GPR56 fragments binding to monobody-coated M280 beads. "ECR" fragment consists of PLL domain+PLL-GAIN linker+GAIN domain. (B) Yeast-displayed β3 binding to GPR56 GAIN domains. K_D values were determined for human and mouse GAIN domains to be 9.6 ±0.6 nM and 130 ±15 nM, respectively. Error bars indicated S.E.M., n=3. (C) "Sandwich" format binding assay whereby purified monobodies from the vertical axis were immobilized on M280 beads and incubated first with unlabeled human (left) or mouse (right) GPR56 ECR, followed by staining with purified monobodies from the horizontal axis, which were detected with fluorescently labeled neutravidin. By design, any protein pairs that that bind overlapping sites on the ECR of GPR56 result in low binding signal. Conversely, protein pairs with non-overlapping binding sites on GPR56 yield high binding signal.

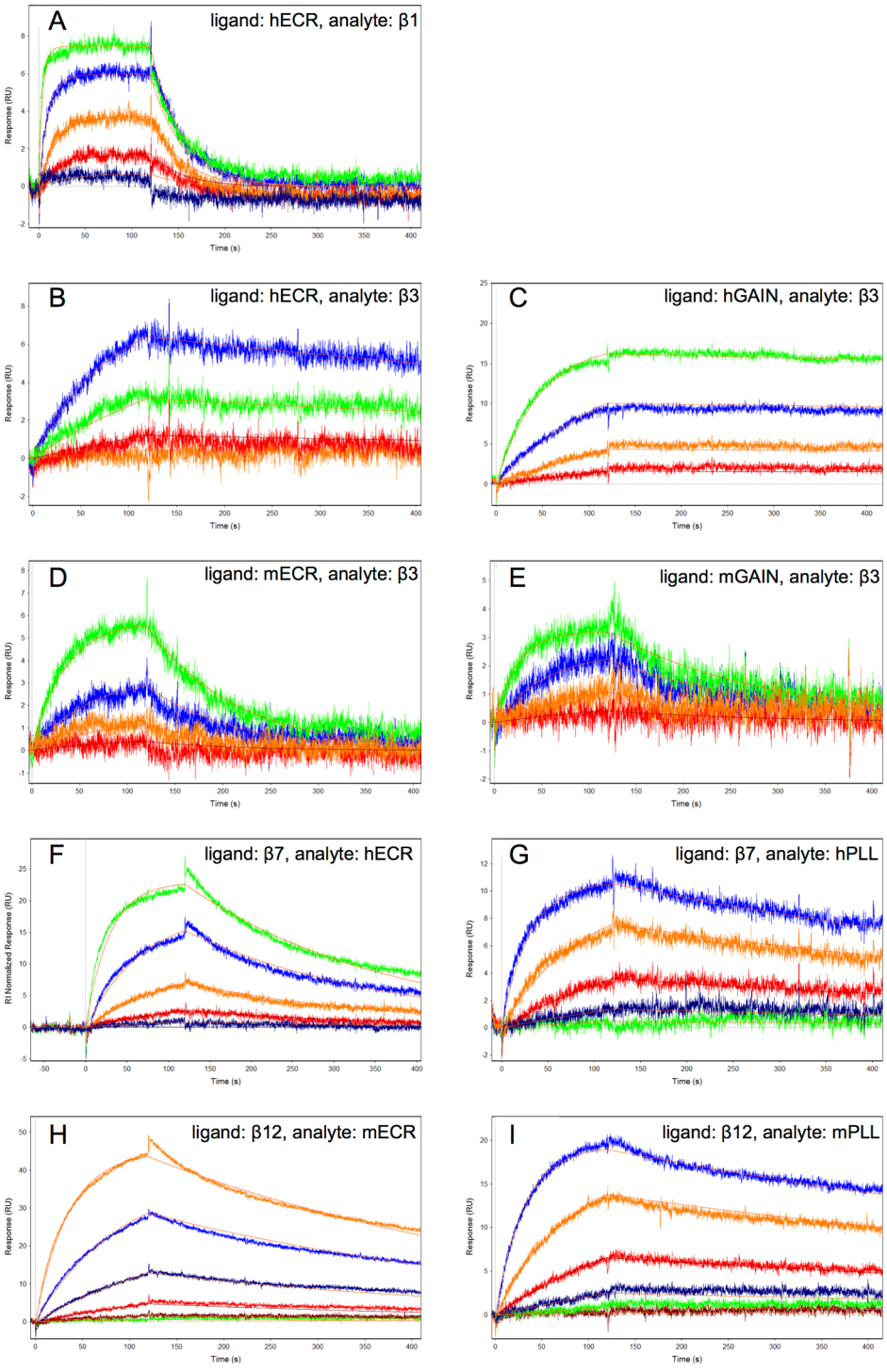


Figure V.4 SPR measurements of monobodies binding to purified GPR56 fragments.

(legend on the following page)

Figure V.4, continued.

The identities of the ligand (protein immobilized on chip) and analyte (protein in solution) are detailed on each plot. Each ligand concentration is shown in a different color trace. Within each plot, the multi-concentration global fit line is shown. See Table V.3 for k_{on}, k_{off}, and K_D values determined from the fit. In order from highest to lowest, the concentrations (in nM) of analyte used were: (A,G) 500.0, 166.7, 55.6, 18.5, and 6.2; (B,D) 40, 13.3, 4.4, and 1.5; (C,E) 13.3, 4.4, 1.5, and 0.5; (F) 1000.0, 333.3, 111.1, 37.0, and 12.3; (H) 1000.0, 333.3, 111.1, 37.0, 12.3, and 4.1; (I) 500.0, 166.7, 55.6, 18.5, 6.2, and 2.1. All measurements were taken at 25˚C.

In contrast to β1, β7, that targeted the PLL domain, increased signaling of hGPR56 with EC_{50} of 800 ±500 nM, resulting in a ~1.6-fold increase in signaling relative to basal activity (Figure V.5C), and thus may be classified as an allosteric agonist of G protein signaling (34). The effect of β7 was also specific to human GPR56 (Figure V.5A-B). This is the first stimulatory monobody of GPR56 to be characterized.

Monobody β3 that bound the GAIN domain of both human and mouse GPR56 exhibited no significant effect on signaling in this assay. Although the binding profiles in Figure 2A indicate that the epitopes for both β1 and β3 include the GAIN domain, the two epitopes do not overlap (Figure V.3C). This observation suggests that these two monobodies engaged the GAIN domain in distinct manners.

Activating and inhibiting monobodies modulate signaling of an autoproteolysis-defective GPR56 mutant

We next set out to determine whether these synthetic GPR56 ligands functioned in a *Stachel*-mediated or *Stachel*-independent manner. To this end, we introduced a single point mutation, H381S, in the GAIN domain of human and mouse GPR56,

ligand	analyte	k_{on} (M^{-1} s^{-1}) *10^5	k_{off} (s^{-1}) *10^{-3}	K_D (nM) [k_{off}/ k_{on}]
hECR	β1	4.1	29.3	71.5
hECR	β3	8.9	0.9	1.0
mECR	β3	4.6	10.8	23.2
hGAIN	β3	5.4	0.14	0.3
mGAIN	β3	9.2	5.8	6.3
β7	hECR	0.6	4.1	66.7
β7	hPLL	3.2	1.3	4.0
β12	mECR	0.8	2.2	27.9
β12	mPLL	1.9	1.1	5.6

Table V.3 GPR56-binding monobody affinity measurements by SPR.

All SPR measurements were taken at 25˚C.

previously shown to abolish GAIN domain-mediated autoproteolysis, and therefore *Stachel*-exposure, without affecting the overall structure of the ECR (55). We compared the effect of monobody treatment on G protein signaling of WT and H381S mutant GPR56, using the SRE-luciferase assay (Figure V.6). The effects of α5, β1, and β7 were essentially indistinguishable for both receptors (Figure V.6D), demonstrating that their effects on G protein signaling measured in this assay are independent of autoproteolysis.

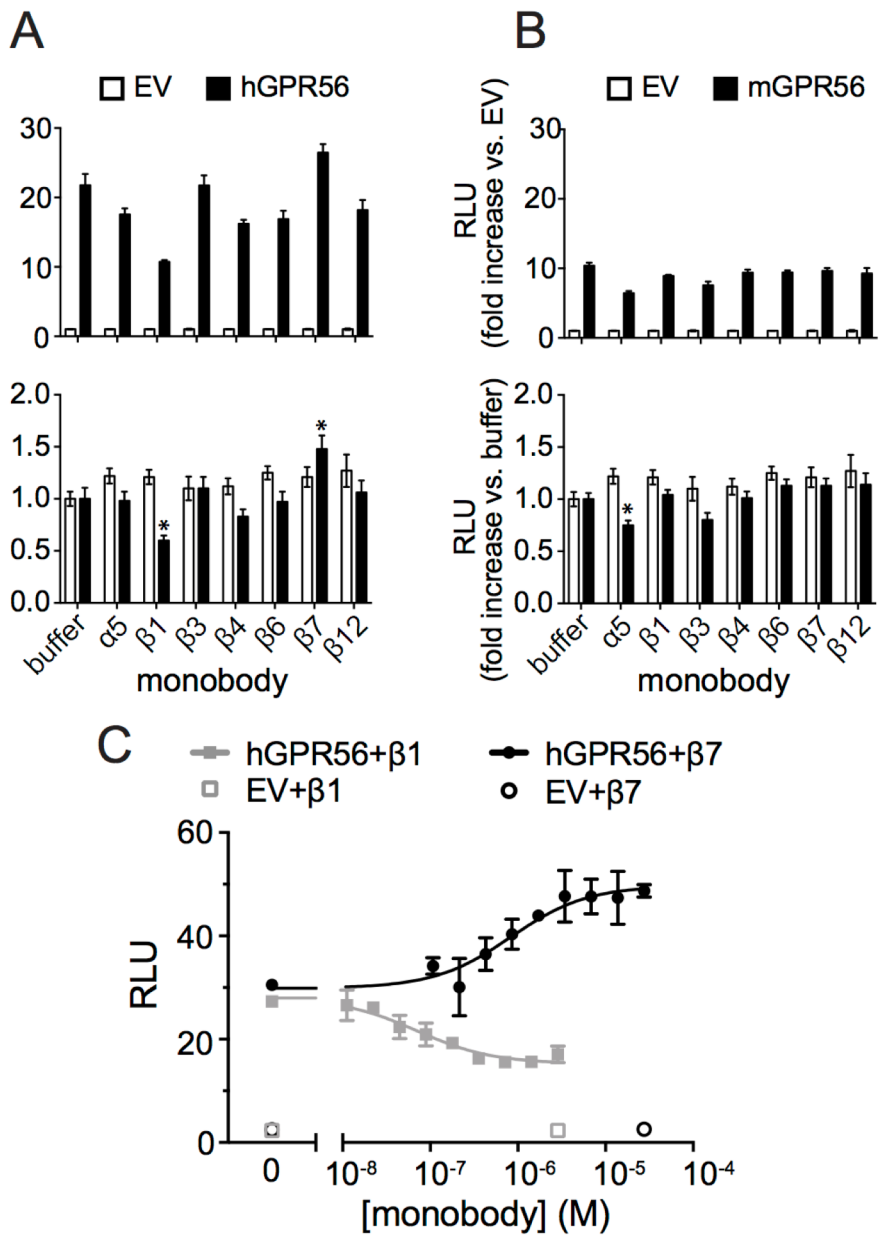


Figure V.5 Modulation of wt GPR56 signaling by monobodies.

(A-B) SRE-luciferase assay for $G\alpha_{13}$-mediated G protein signaling of (A) human and (B) mouse GPR56 in the presence of 0.7 μM purified monobodies presented as fold-increase vs. EV (top) and fold-increase vs. buffer (bottom). Significant effects are representative of many repeated experiments. *, $p<0.05$ vs. buffer-treatment by Student's two-tailed t-test. (C) Titration of monobodies β1 and β7 on SRE-luciferase activity of WT human GPR56. The IC_{50} value of β1 was determined to be 70 ±30 nM. The EC_{50} value of β7 was determined to be 800 ±500 nM. Error bars indicated S.E.M., n=3.

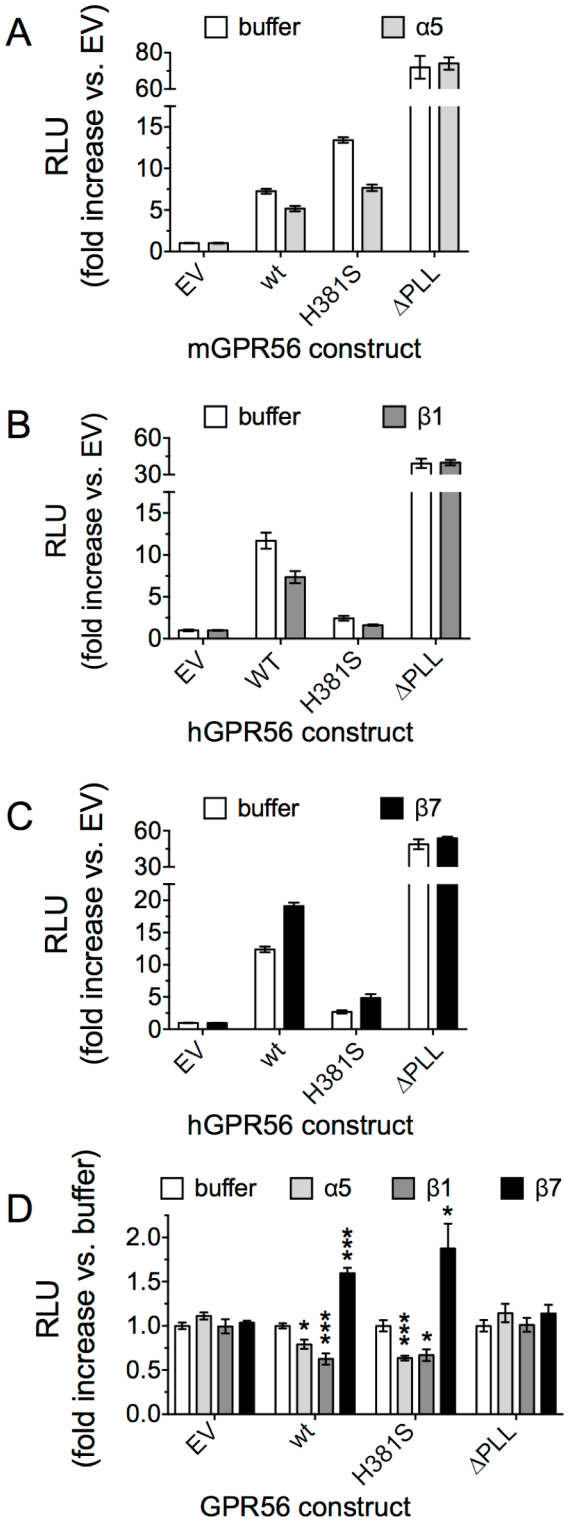


Figure V.6 Monobody-mediated activation and inhibition of GPR56 is autoproteolysis-independent.

(legend on the following page)

Discussion

Synthetic GPR56 ligands elucidate direct regulation of signaling by the ECR

In this study, we set out to engineer synthetic protein ligands targeted to the ECR of human GPR56 and obtained an allosteric agonist (β7) and an allosteric inverse agonist (β1) (Table V.2, Figure V.3 and Figure V.5). Thus, we have substantially expanded the available tools for modulating GPR56 signaling by targeting its ECR. Along with the previously characterized allosteric inverse agonist, the α5 monobody (93), we demonstrate that autoproteolysis is not required for each of these functional monobodies to modulate signaling (Figure V.6). These results strongly suggest that perturbations to the ECR are directly sensed by the 7TM, resulting in altered signaling without NTF shedding and *Stachel* exposure.

Binding characteristics of the new monobodies give further insights into the molecular mechanism of GPR56 regulation. The allosteric inverse agonists, α5 and β1, bound to the full-length ECR but not to the isolated GAIN or PLL domain (93) (Figure V.3). The X-ray crystal structure of the α5-ECR complex revealed that α5 interacts with both the PLL and GAIN domains (93), leading to the speculation that α5, and probably β1, decrease basal activity by restricting the inter-domain motions of the ECR. Unlike

Figure V.6, continued.

(A-D) SRE-luciferase assay of indicated GPR56 constructs in the presence of buffer or monobody. (A) mouse GPR56 constructs treated with 4.9 μM α5. (B-C) human GPR56 constructs treated with (B) 2.9 μM β1 and (C) 27.5 μM β7. (D) Data from panels A-C normalized to buffer treatment to account for differences in measured basal activity of GPR56 constructs, which we have previously shown is due, in part, to differences in cell-surface expression (93). *, $p<0.05$; ***, $p<0.001$ vs. buffer-treatment by Student's two-tailed t-test. Error bars indicated S.E.M., n=3.

these inverse agonists, β7, the allosteric agonist, binds more tightly to the PLL domain alone than it does for the full ECR (Figure V.3A), indicating that it binds to a region of the PLL domain that is less accessible in the full ECR, probably blocked by the GAIN domain or PLL-GAIN linker. As such, we speculate that by binding to the PLL domain within the full ECR, β7 disrupts the PLL-GAIN interface, thereby inducing a conformational change in the ECR or altering transient interactions between the ECR and 7TM, leading to increased basal activity. Taken together, the distinct binding profiles of agonist and inverse agonist monobodies suggest that alterations of the relative orientation between the GAIN and PLL domains contribute to regulation of GPR56 signaling. Alternatively, these results also suggest the possibility in which ECR-bound monobodies directly interact with 7TM and modulate signaling (Figure V.7). Future studies will determine the contributions of these complementary mechanisms.

The fact that β3, which tightly binds the GAIN domain (Figure V.3A-B and Figure V.4), has no significant effect on signaling (Figure V.5A-B) suggests against the importance of shedding and *Stachel* exposure in basal activity. If basal activity were solely regulated by low levels of spontaneous shedding and *Stachel* exposure, one would expect that stabilization of the NTF-CTF associated state would result in decreased basal activity. Because the junction between the NTF and CTF resides within the GAIN domain, one would expect that a ligand for the GAIN domain would reduce basal activity by stabilizing the associated state, i.e. the folded conformation of the GAIN domain. Most importantly, the autoproteolysis-defective mutation did not eliminate the modulatory effects of the monobodies (Figure V.6). Thus, these results strongly suggest that the regulation of basal activity is *Stachel*-independent.

A unified model of ligand-mediated regulation of aGPCRs

The biological relevance of *Stachel*-mediated aGPCR activation is extremely clear (35, 83, 86, 130, 147, 149, 178). However, unanswered questions and recent observations necessitate the introduction of the complementary *Stachel*-independent model (91). For example, overexpression of autoproteolysis-deficient *lat-1*/ADGRL1 in *lat-1*-knockout *C. elegans* rescues the WT phenotype, suggesting some aGPCR functions do not require autoproteolysis (179). Additionally, there are several aGPCRs that lack the conserved residues critical for autoproteolysis and therefore remain uncleaved (55, 56). Furthermore, several aGPCRs including GPR56 are found partially uncleaved *in vivo* (55, 57, 93). For example, GPR56 in skeletal muscle was found to be almost completely uncleaved (57), although it plays critical roles in skeletal muscle cells (79, 148). Together these observations suggest that *Stachel*-independent mechanisms may play important roles in aGPCR signaling.

The irreversible activation of aGPCRs via the *Stachel* is mirrored by the relatively well-studied GPCRs called protease-activated receptors (PARs). PARs are cleaved by an external protease as opposed to the autoproteolytic cleavage in the aGPCR GAIN domain. Upon cleavage of a PAR, a tethered peptide agonist is exposed that irreversibly binds and activates the 7TM (180, 181). After activation, PARs must be endocytosed and degraded in order to turn off the signal (182). This signaling paradigm whereby each copy of the receptor is synthesized in order to be irreversibly activated exactly once before being degraded is associated with a relatively high energetic cost. For canonical GPCRs, reversible ligand-mediated activation events result in

endocytosis, ligand dissociation, and recycling to the plasma membrane (183). By recycling each receptor multiple times, cells avoid paying high energetic costs associated with synthesizing large proteins. This argument is especially relevant when considering the size of aGPCRs: the average length of the four human PARs is 395 amino acids (ranging from 374-425 amino acids), whereas the average length of the 33 human aGPCRs is 1414 amino acids (ranging from 457 to 6306 amino acids). Thus, the energetic cost of synthesizing an aGPCR for a single signaling event is significantly larger than that of a PAR. This reasoning supports the paradigm that in addition to an irreversible activation mode, aGPCR signaling is subject to reversible regulation. Our key observation that ECR ligands modulate signaling in an autoproteolysis-independent manner complements the recent studies by Kishore and colleagues, in which they measured the basal activities of GPR56 constructs with various ECR truncations through multiple signaling pathways (91, 184). Using an SRE-luciferase assay a construct lacking the NTF (i.e. 7TM with exposed *Stachel*; Figure V.1) had the highest activity, whereas one lacking both the NTF and the *Stachel* (i.e. just the 7TM) had the lowest activity (91), confirming the agonistic function of *Stachel* on the 7TM (35). In comparison, the full-length constructs of both the WT and an autoproteolysis-deficient mutant exhibited a moderate level of activity (91), suggesting that the full ECR modulates 7TM signaling.

The location of the *Stachel* sequence within the GAIN domain strongly suggests that *Stachel*-mediated activation is a single-turnover event. After autoproteolysis and before shedding the hydrophobic *Stachel* is buried within the hydrophobic core of the GAIN domain and forms extensive hydrogren-bond networks with adjacent β-strands

(55, 93). This architecture strongly suggests that the exposure of the *Stachel* requires substantial deformation of the GAIN domain. Furthermore, because the *Stachel* is a central part of the GAIN domain, the release of the *Stachel* from the GAIN domain most likely leads to a collapse of the original conformation, prohibiting re-association of the *Stachel* (55). In a similar vein, transient exposure of the *Stachel* to interact with the 7TM without causing irreversible NTF-CTF dissociation should be an extremely rare event, if not practically impossible. Therefore, we propose that *Stachel*-mediated aGPCR activation is irreversible.

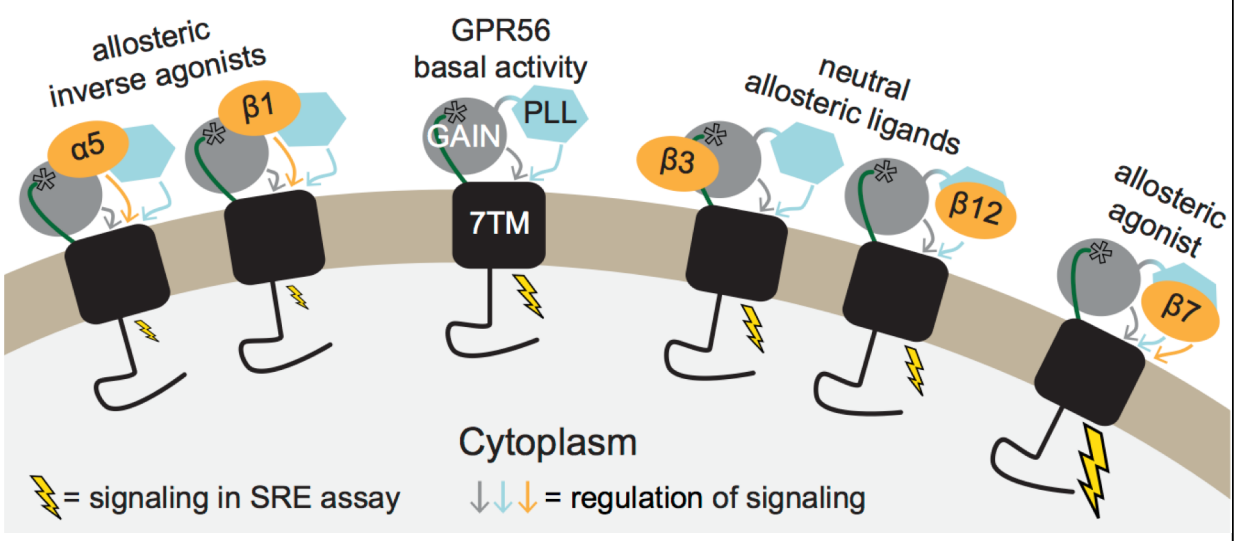


Figure V.7 Proposed mechanisms for distinct monobody-dependent modulation of GPR56 function.

Monobodies are arranged based on function and GPR56 domains necessary for binding are illustrated. Lightning bolt size represents signaling intensity. Autoproteolysis site indicated by *. Unneeded autoproteolysis indicated by outlined *. Rather than direct interactions, colored arrows represent regulation of 7TM signaling by the ECR or monobodies. Designations of “allosteric inverse agonist”, “neutral allosteric ligand”, and “allosteric agonist” based on definitions proposed by Christopoulos et al. (34).

In contrast to the *Stachel*-mediated activation, ECR-mediated *Stachel*-independent regulation is likely to be moderate and require no receptor turnover for resetting the signal, because ECR ligands interact with aGPCRs in a reversible manner. Thus, the *Stachel*-independent regulation should be suited for fine-tuning the signaling near the basal levels. The value of this mechanistic insight is clear from a pharmacological point of view, as the therapeutic benefits of inducing moderate and enormous changes, usually with allosteric and orthosteric ligands, respectively, in GPCR signaling have both been repeatedly demonstrated (30, 32, 185–187).

In addition to the mechanistic insight gleaned from these monobodies, the recent discovery of several synthetic small molecule ligands of GPR56 has furthered the potential for development of aGPCR-targeted therapeutics in the near future (89). The powerful combination of ECR-targeted synthetic proteins including monobodies and antibodies (83, 92, 172) with 7TM-targeted small molecule ligands will be invaluable in future mechanistic and pharmacological studies of aGPCRs.

CHAPTER VI

The conserved patch on the PLL domain of GPR56/ADGRG1 mediates tissue transglutaminase binding

Summary

A hallmark feature of an adhesion G protein-coupled receptor (aGPCR) is a long a diverse extracellular region, consisting of multiple protein domains. Though bioinformatics analysis suggests these domains interact with protein ligands, either from the extracellular matrix, or on neighboring cells, few such interactions have been biophysically characterized. Here, we present the first rigorous biochemical characterization of the interaction between the aGPCR GPR56/ADGRG1 and its native ligand, tissue transglutaminase (TG2), and show that PLL domain-binding monobodies block this interaction.

Introduction

Though many aGPCR ligands have been identified, few have been characterized with respect to precise binding site and binding affinity (68, 188). Determining the structural implications of ligand binding remains a key aspect of understanding aGPCR-mediated biology. Furthermore, as native aGPCR ligands tend to be large and unwieldy transmembrane or extracellular matrix proteins, designing precise experiments to elucidate the functional and mechanistic consequences of native ligand binding has proved challenging. In order to avoid this problem, we chose to engineer well-behaving synthetic ligands (i.e. monobodies) to provide insight into the mechanisms by which

ligands may regulate aGPCR signaling. However, in order to fully understand the biology mediated by aGPCRs, it is critical to characterize interactions with native ligands.

The pathophysiological relevance the aGPCR GPR56/ADGRG1 is detailed in the previous chapters. Three natural ligands for GPR56 have been identified including collagen III (75), heparin (88), and tissue transglutaminase (TG2) (87), an extracellular matrix crosslinking enzyme. Studies show that collagen III and heparin activate and inhibit receptor signaling by increasing and decreasing receptor shedding, respectively. TG2 has not been shown to directly modulate G protein signaling of GPR56. In initial characterization experiments, each ligand was shown to interact with a region of GPR56 later identified as the PLL domain (93). However, in the time since the structure of the PLL domain was solved, no precise natural ligand binding sites have been identified.

In our previous study of the GPR56 ECR, we showed that deletion of the N-terminal PLL domain was sufficient to increase basal activity of the receptor, a direct observation of ECR-mediated regulation of signaling. Furthermore, we identified a conserved, surface-exposed patch of the PLL domain necessary to promote GPR56-mediated myelination in a zebrafish model (93). Though we speculated this patch may be important for native ligand binding, we had no information regarding the identity of such a ligand, nor the tools to probe such an interaction.

In this study, we performed the most rigorous biochemical characterization of the GPR56-TG2 interaction to date and identified the precise binding site of TG2 on the PLL domain. We went on to demonstrate that two PLL-binding monobodies block the GPR56-TG2 interaction. Finally, we show that in our *in vitro* GPR56 signaling assay, the

fragment of TG2 necessary to bind GPR56 has no significant effect on G protein signaling.

Results

The two C-terminal domains of TG2 are sufficient to bind the PLL domain of GPR56

Using a baculovirus-based expression system as previously described (55, 93), recombinant mouse TG2 was purified from High Five insect cell cytoplasm. We first set out to confirm previous results that the C-terminal two domains of TG2 (D3D4) are sufficient to mediate GPR56-binding (87). We purified full-length TG2 and a construct corresponding to D3D4, each with an AVI-tag to facilitate biotinylation. We immobilized the TG2 constructs on M280 streptavidin-coated beads and measured binding to purified GPR56 constructs by flow cytometry as previously described (93). We found that TG2 (both full-length and D3D4) bound strongly to wt human and mouse GPR56 (Figure VI.1A), confirming that the D3D4 construct is sufficient to mediate GPR56 binding. As such, we used the D3D4 construct for all experiments unless otherwise indicated. We observed a marked decrease in binding signal with introduction of the H89A mutation (Figure VI.1A). This mutation to a surface-exposed conserved patch on the PLL domain (Figure VI.1B), was previously identified to cause GPR56 loss-of-function in a zebrafish model of CNS myelination without altering basal activity (93).

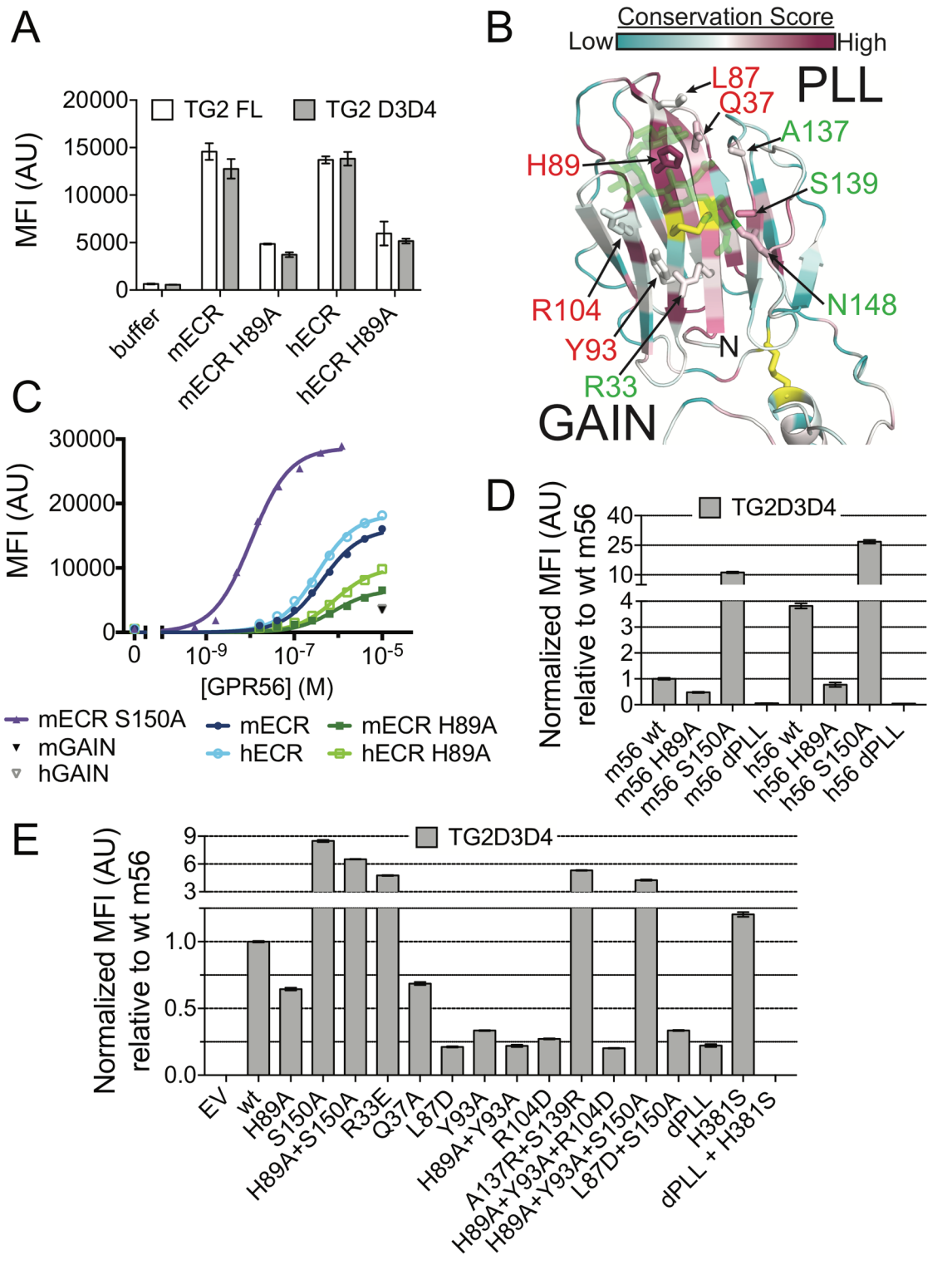


Figure VI.1 Characterization of the GPR56-TG2 interaction.

(legend on the following page)

Using mouse TG2D3D4-coated M280 beads, we measured the affinity of TG2 for the GPR56 ECR. We observed a dissociation constant (K_D) of 440±20 nM and 330±15 nM for wt mouse and human ECR, respectively. As suggested by the aforementioned single-point binding measurements, we observed K_D values of 880±80 nM and 790±11 nM for mouse and human ECR H89A, respectively (Figure VI.1C).

In addition to confirming the domains of TG2 responsible for binding, we set out to determine the region of GPR56 involved in binding. To more accurately represent a biologically relevant interaction, we developed a binding assay in which FL GPR56 was expressed in HEK293T cells and binding of purified TG2 D3D4 was quantified using flow cytometry. In this assay, we confirmed that wt human and mouse GPR56 bound strongly to TG2 and H89A decreases binding signal (Figure VI.1D). Furthermore, we observed that deletion of the PLL domain resulted in complete loss of binding, demonstrating that the PLL domain is necessary for TG2 binding. We also observed that deletion of the N-linked glycan at residue N148 (achieved by mutating the N-linked glycosylation consensus sequence N-X-S/T) by incorporating the S150A mutation

Figure VI.1, continued.

(A) Purified extracellular GPR56 fragments bind M280 beads coated with FL TG2 or its C-terminal pair of domains (D3D4). (B) Analysis of surface-exposed residues that comprise the conserved patch of the PLL domain as identified previously (93). Non-cysteine residue sidechains are colored by conservation score. Cystine residues are colored yellow. The N148-linked glycan is colored green. Residue labels are colored red or green if mutation of that residue results in decreased or increased TG2 binding signal, respectively (as determined in panel E). (C) Titration of GPR56 fragments to TG2D3D4-coated M280 beads. K_D values were determined for mECR, hECR, mECR H89A, hECR H89A, and mECR S150A to be 440±20nM, 330±15nM, 880±80nM, 790±11nM, and 11±0.7nM, respectively. (D-E) HEK293T cells transfected with indicated GPR56 constructs were stained with purified TG2D3D4. TG2D3D4 MFI is normalized to different expression levels of GPR56 mutants shown in Figure VI.2. (E) All constructs shown are mouse.

resulted in dramatically increased TG2 binding (Figure VI.1D). Using TG2-coated M280 beads and purified GPR56 ECR, we observed a K_D of 11±0.7 nM for mouse ECR S150A, roughly a 40-fold increase in affinity over wt (Figure VI.1C). As illustrated in Figure VI.1B, this glycan was previously observed to lie directly adjacent to H89 on the PLL domain (93). Therefore, it is not unreasonable that its deletion may dramatically affect an interaction already shown to involve H89. All these data suggest that TG2 interacts with the previously described “conserved patch” of the PLL domain, including H89 and the N148-linked glycan, both of which are highly conserved across species (93).

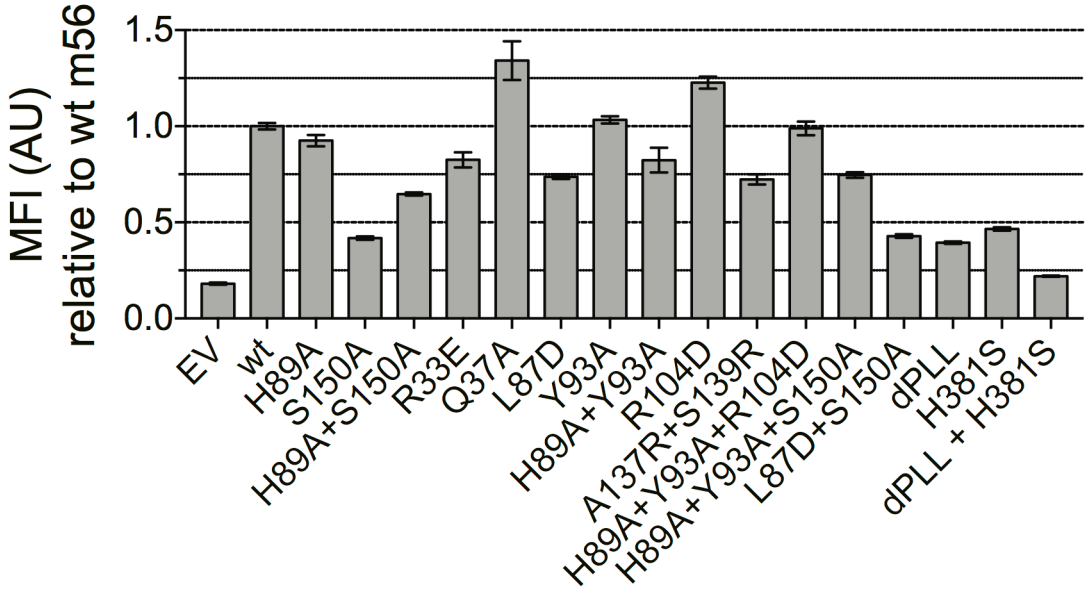


Figure VI.2 Cell-surface expression of mouse GPR56 mutants.

HEK293T cells transiently transfected with plasmids harboring the indicated GPR56 mutations were assayed for GPR56 surface expression using labeled Mb2-3 in flow cytometry. Cells were not permeabilized.

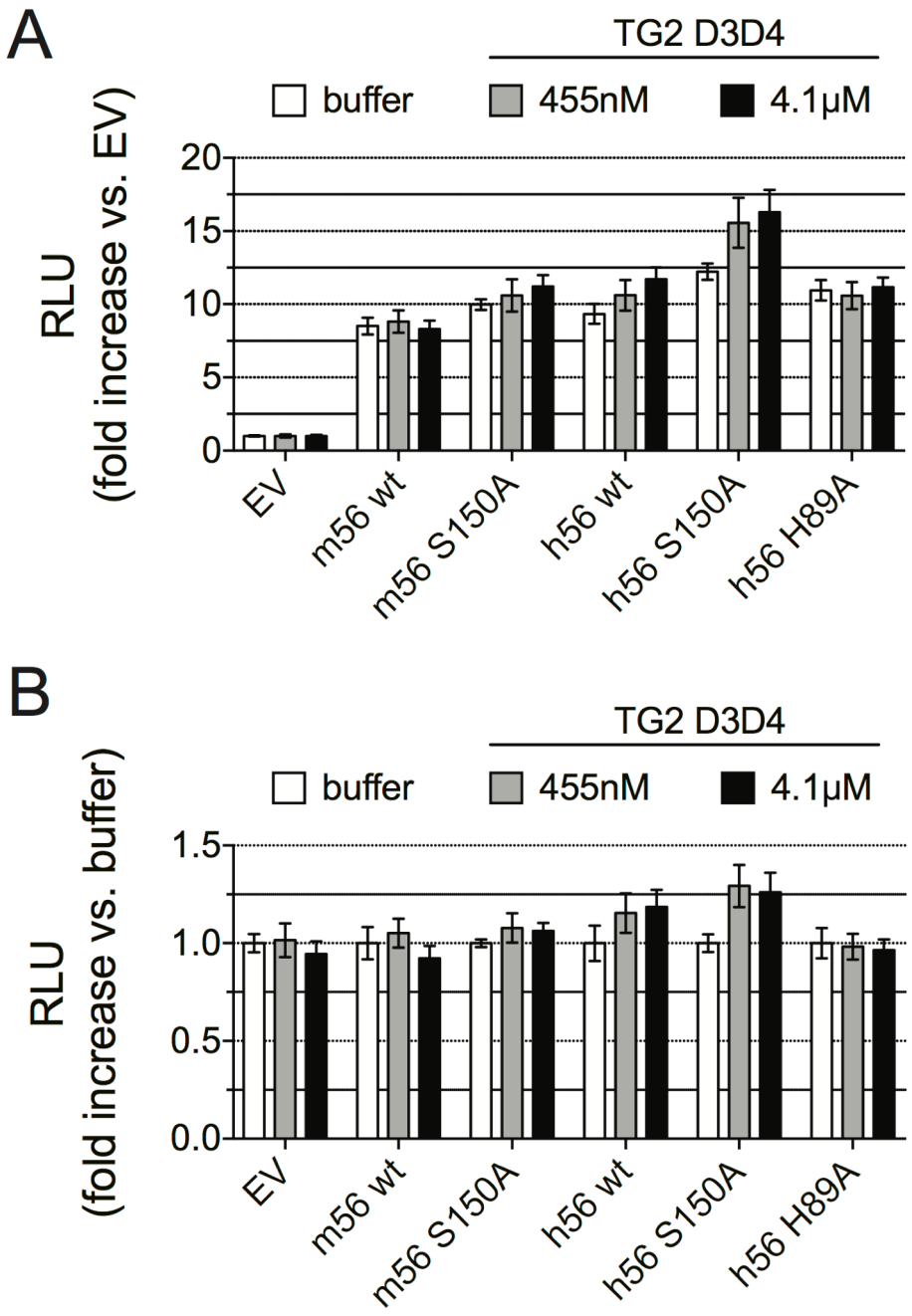


Figure VI.3 Purified TG2 D3D4 has minimal, if any, effect on GPR56 G protein signaling *in vitro*.

In the SRE-luciferase assay, treatment of cells expressing wt or mutant GPR56 constructs with high concentrations of TG2D3D4 resulted in negligible changes in signaling. The largest observed of TG2 treatment occurred with human GPR56 S150A, resulting in a ~1.2-fold increase in signaling. Data presented as fold-increase relative to empty vector (A) and fold-increase relative to untreated cells (B).

In an effort to determine if TG2 directly activates GPR56 upon binding, we treated GPR56-expressing HEK293T cells with TG2 D3D4 in the SRE-luciferase assay and observed no obvious effect on signaling of the wt receptor (Figure VI.3).

The previously identified conserved patch on the PLL domain mediates TG2 binding

In order to identify the specific residues on the PLL domain that mediate TG2 binding, we constructed a panel of GPR56 single, double, and triple mutants. We chose residues to mutate based on their inter-species conservation, surface exposure, and proximity to H89 and the N148-linked glycan. In addition to H89, we found that mutation of several nearby residues (L87D, Y93A, R104D) in the PLL domain resulted in decreased TG2 binding signal (Figure VI.1B,E). Though L87D results in undetectable TG2 binding, this mutation decreased cell-surface expression of GPR56 by ~25% Figure VI.2). We found that the H89A+Y93A+R104D triple mutant abolished TG2 binding without impacting cell-surface expression (Figure VI.2). Thus, we believe this would be an ideal loss-of-function mutant to use in future experiments designed to measure the function of the TG2-GPR56 interaction.

PLL-binding monobodies inhibit TG2 binding to human and mouse GPR56

Though we are able to break the GPR56-TG2 interaction by mutagenesis as discussed above, we believe tools to disrupt this interaction (or interactions between GPR56 and other natural ligands) in a system composed of wt molecules will be invaluable in future mechanistic and functional studies of the pathophysiology mediated

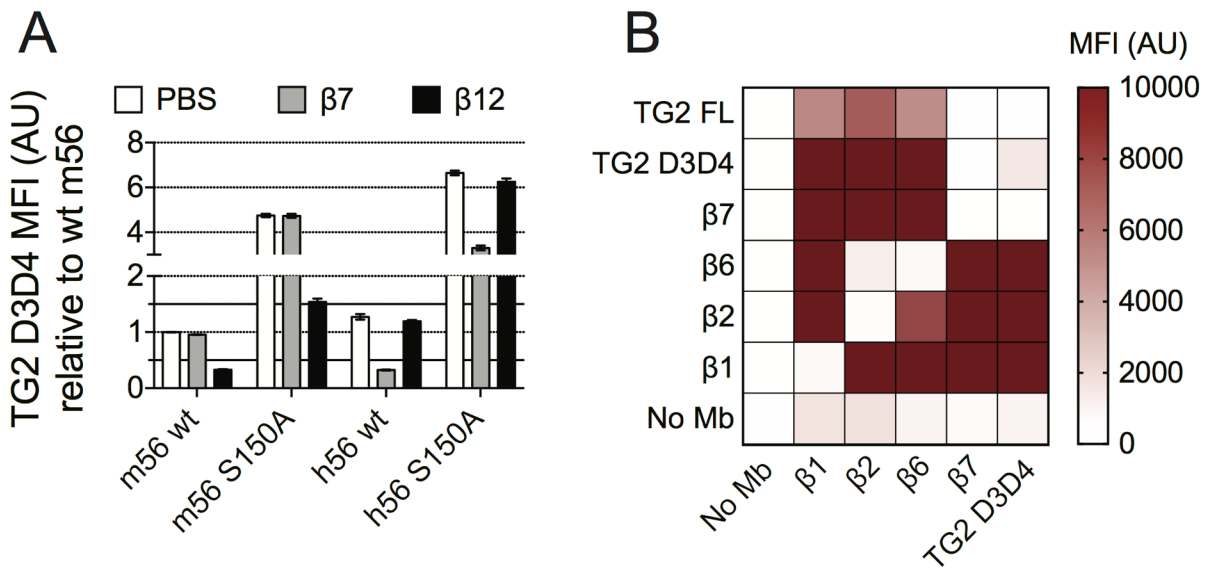


Figure VI.4 PLL-binding monobodies block TG2 binding.

(A) HEK293T cells transfected with indicated GPR56 constructs were stained with purified TG2D3D4 in the presence of excess unlabeled monobody competitor. (B) "Sandwich" format binding assay whereby purified proteins from the vertical axis were immobilized on M280 beads and stained first with unlabeled hGPR56 ECR, followed by purified protein from the horizontal axis, which were detected with labeled neutravidin. By design, any protein pairs that that bind overlapping sites on hGPR56 ECR result in low binding signal. Conversely, protein pairs with non-overlapping binding sites on hGPR56 produce high binding signal.

by GPR56. Specifically, we hypothesized that monobodies that bind the PLL domain of GPR56 may competitively block TG2 binding.

Of the monobodies reported in the previous chapter, β7 and β12 specifically bind human and mouse PLL domain, respectively (Figure VI.4A). We carried out a GPR56-TG2 binding assay in the presence of excess β7 and observed a significant decrease in TG2 binding to hGPR56 but not mGPR56, as expected given β7 does not bind mPLL. For β12, which binds mPLL and not hPLL, we obtained the opposite result (Figure VI.4A). Thus, these data show that PLL-binding monobodies inhibit TG2 binding.

To confirm β7 inhibits TG2 binding, we carried out a “sandwich” binding assay. In short, we immobilized each protein of interest (i.e. monobody or TG2) on M280 beads, incubated with unlabeled hGPR56 ECR, and stained with a labeled version of each protein of interest (Figure VI.4B). In this format, any pair of proteins that binds non-overlapping sites on the hGPR56 ECR should yield a strong binding signal. Pairs of proteins that share overlapping binding sites on hGPR56 ECR should yield weak binding signal. In this assay, we confirmed β7 shares an overlapping binding site on hGPR56 ECR with TG2.

Discussion

TG2 belongs to the transglutaminase family, a group of cross-linking enzymes with effects on biological functions such as apoptosis, cell-matrix interactions, and tissue stability, as well as involvement in diseases such as autoimmune disorders, neurodegenerative conditions like Alzheimer’s disease, and cancer (189). TG2 catalyzes a Ca^{2+}-dependent transamidation reaction that crosslinks proteins in the extracellular matrix, a process that has been implicated in tumor progression. Studies have implicated TG2 in murine models of melanoma progression, and have shown that GPR56 may inhibit melanoma progression in a TG2-dependant manner (82).

We have confirmed previous results that the GPR56-TG2 interaction occurs between the GPR56 PLL domain and the C-terminal pair of TG2 domains (D3D4) (87) and we measured the affinity of the interaction between the wt proteins to be sub-micromolar (Figure VI.1). Furthermore, we have shown that PLL-binding monobodies block TG2 binding and mapped the TG2-binding site on the PLL domain to the

previously identified surface-exposed conserved patch (Figure VI.4). These observations suggest that the H89A-induced loss-of-function phenotype we previously observed in a zebrafish model of GPR56-mediated myelination (93) is due to abrogation of TG2 binding. Future studies of the GPR56-TG2 interaction in oligodendrocyte development and myelination may reveal the role of TG2 in these complex processes.

The vast majority of studies involving TG2 focus on its enzymatic activity (i.e. crosslinking proteins in the extracellular matrix). The active site of TG2 is not present in the GPR56-binding fragment of TG2. Thus, though it is clear that enzymatic activity is not required for GPR56 binding, the role of TG2-mediated crosslinking in the context of GPR56 binding remains unclear. When we added TG2 D3D4 to our SRE-luciferase assay for GPR56-mediated G protein signaling, we saw no effect on signaling of the wt receptor (Figure VI.3). This suggests GPR56 may function as a simple scaffold to recruit TG2. However, as the zebrafish myelination phenotype discussed above was shown to be G protein-signaling dependent, we speculate that TG2 may play an indirect role in GPR56 activation upon its recruitment *in vivo*. Future studies addressing such complicated macromolecular assemblies in a cellular context will provide mechanistic insight into the biological activation of GPR56.

CHAPTER VII

Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion[1]

Summary

FLRTs are cell-adhesion molecules with emerging functions in cortical development and synapse formation. Their extracellular regions interact with LPHNs to mediate synapse development, and with UNC5/Netrin receptors to control the migration of neurons in the developing cortex. Here, we present the crystal structures of FLRT3 in isolation and in complex with LPHN3. The FLRT3/LPHN3 structure reveals that LPHN3 binds to FLRT3 at a distinct site from UNC5. Structure-based mutations specifically disrupt FLRT3/LPHN3 binding, but do not disturb their interactions with other proteins or their cell-membrane localization. Thus, they can be used as molecular tools to dissect the functions of FLRTs and LPHNs in vivo. Our results suggest that UNC5 and LPHN3 can simultaneously bind to FLRT3 forming a trimeric complex and that FLRT3 may form trans-synaptic complexes with both LPHN3 and UNC5. These findings provide molecular insights for understanding the role of cell-adhesion proteins in synapse function.

---

[1] The text of this section is copied verbatim (with minor changes) from:
Lu YC, Nazarko OV, Sando R III, Salzman GS, Südhof TC, and Araç D. (2015) Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion. Structure 23(9):1678–1691. Y.C.L. designed and performed the crystallography, structure determination, mutagenesis, and bio-layer interferometry binding experiments. O.N. designed and performed the HEK cell expression and flow cytometry binding as- says. R.S. designed and performed the cell aggregation assays. G.S. designed and performed the structure determination, differential scanning fluorimetry assay, and designed and assisted with the flow cytometry binding assays. T.C.S. designed the cell-aggregation assays and contributed to the preparation of the manuscript. D.A. designed all the experiments and wrote the paper. (68)

Introduction

During neural development, immature neurons migrate from their birthplaces in the embryo to their final positions and complete synaptic circuits. Outgrowth of axons and dendrites from neurons, guidance of the motile growth cone through the embryo towards postsynaptic partners and finally the generation of synapses between these axons and their postsynaptic partners are essential landmarks of neural development that are each mediated by a complex interaction network of cell-surface proteins on the nerve cells. Distinct properties of each neuron and its synaptic connections are also a function of the cell-adhesion molecules expressed on its surface and the interactions these molecules are involved in. An imperfection in any of these steps may lead to malformations or inappropriate connectivity of the brain and is believed to be involved in many neurodevelopmental disorders such as attention deficit hyperactivity disorder, autism and brain cancers (190, 191) Although the role of binary protein-protein interactions are commonly studied and understood, it is likely that the presence or absence of ternary, quaternary and even higher-order protein-protein interactions are determinants for the build-up of a network as complex as is in the brain. However, the molecules and the underlying mechanisms of such complex phenomena are unclear.

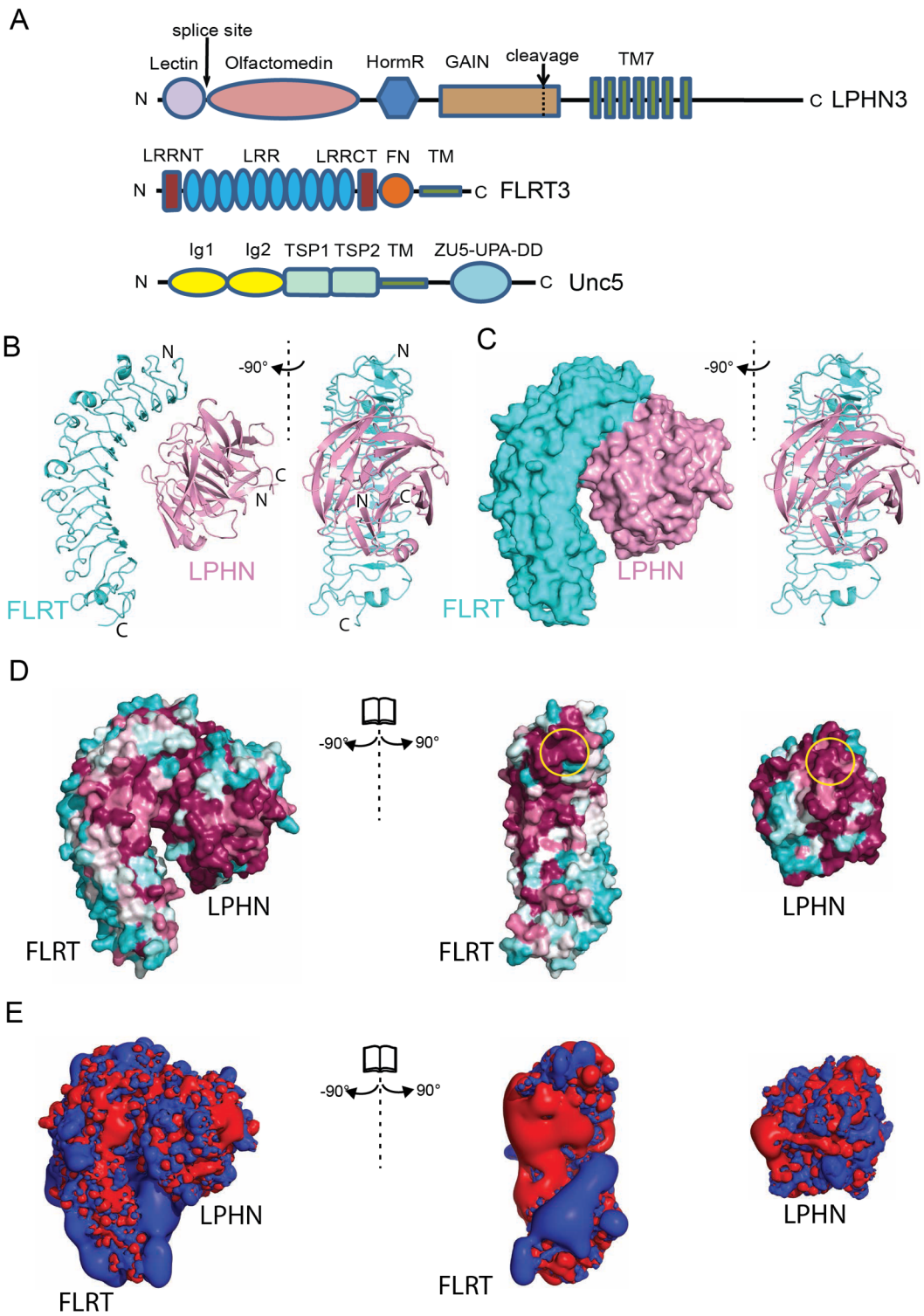


Figure VII.1 Crystal structure of the FLRT3/LPHN3 complex.

(legend on the following page)

Fibronectin leucine-rich repeat transmembrane (FLRTs) proteins are cell-surface molecules that contribute to early embryonic, vascular, and neural development(61, 67, 192–195). FLRTs are type-I membrane proteins with extracellular regions consisting of a leucine-rich repeat (LRR) domain with ten LRR repeats, a fibronectin type 3 domain followed by a juxtamembrane linker (196) (Figure VII.1A). FLRTs are expressed in many tissues, including brain (196, 197), and different FLRT isoforms (FLRT1-3) have different cell type-specific expression patterns in the hippocampus and cortex (Allen Mouse Brain Atlas, 2009). FLRTs interact with the axonal guidance receptors UNC5B and UNC5D proteins (67). Their ectodomains are suggested to shed from neurons to act as repulsive cues in axon guidance and neuron migration (67, 198). FLRTs have also been identified as high affinity endogenous ligands for latrophilins (LPHNs) and were suggested to play a role in glutamatergic synapse development (61, 199). Moreover, FLRT proteins interact with each other and may promote homotypic cell adhesion, and are additionally implicated in FGF (fibroblast growth factor) signaling during development (194, 197, 200). The N-terminal LRR domain of FLRTs is reported to be involved in all reported interactions of FLRTs (200, 201). The involvement of LRR

Figure VII.1, continued.

(A) Schematic diagram of vertebrate LPHN3, FLRT3, and UNC5D showing their domains. (B) Ribbon diagram of the FLRT3/LPHN3 heterodimer. (C) Surface representation of the FLRT3/LPHN3 heterodimer. (D) The structure of the FLRT3/LPHN3 complex is shown in surface representation on which the conservation of residues are mapped from most conserved (magenta) to least conserved (cyan) (using the ConSurf server (166)). The LPHN-binding site and the FLRT-binding site, where they come closest to each other, are indicated by yellow circles. (E) Positive (blue) and negative (red) electrostatic isosurfaces calculated at contours ±2 kT/e using PDB2QPR and APBS for LPHN3, FLRT3, and the complex structures. The structures in C, D and E are shown in a similar orientation as in B.

domains of FLRTs in numerous heterophilic interactions suggest a central role for FLRTs in neural development and raises the necessity to identify the specific binding sites for each of LRR domain interactions (such as Latrophilin, UNC5 and homophilic FLRT binding sites on FLRT) in order to unambiguously dissect the role of each interaction for FLRT function.

Latrophilins (LPHN1-3) are cell-surface molecules that belong to the adhesion-type G-protein coupled receptor (GPCR) family (59, 202). LPHN1 was identified as the calcium-independent receptor for α-latrotoxin, a black widow spider toxin that triggers massive neurotransmitter release from neurons and neuroendocrine cells (59, 203–206). Mutations of LPHNs have been linked to attention deficit hyperactivity disorder as well as numerous cancers (207–209). LPHNs are highly expressed in the brain (210), and were shown to function as heterophilic cell adhesion molecules in processes such as synapse formation or maintenance. They are the only adhesion-type GPCRs besides flamingo-like CESLR proteins that are conserved between vertebrates and invertebrates. In *C. elegans*, LPHN1 homolog Lat-1 is required for the alignment of cell division planes to the anterior-posterior axis during development (211). In *Drosophila*, Latrophilin/dCIRL sensitizes the chordotonal neurons to modulate the perception of mechanical signals (150). In vertebrates, LPHN3 and FLRT3 were reported to interact in trans through their ectodomains to mediate cell adhesion, an interaction that promotes the development of glutamatergic synapses (61). LPHNs have large extracellular sequences that contain an N-terminal lectin domain, a central olfactomedin (Olf) domain, a serine/threonine-rich region, a hormone-binding domain, and a C-terminal conserved GPCR Autoproteolysis Inducing (GAIN) domain that mediates

	FLRT3	*FLRT3/LPHN3 Complex*
Data collection		
Wavelength	0.96638	1.00883
Space group	$P2_1$	$P4_3$
Cell dimensions		
a, *b*, *c* (Å)	76.140 106.581 84.152	102.268 102.268 419.23
α, β, γ (°)	90 90.31 90	90 90 90
Resolution (Å)	50 - 2.601 (2.64 - 2.60)*	73.2 - 3.60 (3.66 - 3.60)*
R_{sym}	0.144 (0.640)	0.262 (0.634)
$CC_{1/2}$	0.763 (high res shell)	0.671(high res shell)
Completeness (%)	95.7 (73.1)	98.3 (79.3)
Redundancy	8.8 (1.6)	8.9 (1.5)
Refinement		
Resolution (Å)	45 - 2.60 (2.67 - 2.60)	73.20 - 3.60 (3.71 – 3.60)
No. reflections	39414 (1492)	45263 (1507)
R_{work}	0.2344 (0.2973)	0.2004 (0.2508)
R_{free}	0.2956 (0.4364)	0.2621 (0.3460)
No. atoms		
Protein	10,300	18360
Ligand/ion	56/0	56/4
B-factors		
Protein	48.138	117.938
Ligand/ion	40.311	118.261
R.m.s. deviations		
Bond lengths (Å)	0.0123	0.006
Bond angles (°)	1.330	0.983
Ramachandran statistics		
Favored (%)	90.28	93.5%
Outliers (%)	0	0

*Values in parentheses are for highest-resolution shell.

Table VII.1 Data collection and refinement statistics.

autoproteolysis (55, 59, 202) (Figure VII.1A). The LPHN Olfactomedin (Olf) domain is required for its synapse-promoting function and also for FLRT binding (199). Similar to FLRTs, LPHNs are also involved in numerous interactions. In addition to FLRTs, LPHNs interact with the members of the teneurin/ODZ family proteins (60, 65, 212) and with

neurexins (62).

Uncoordinated-5 (UNC5 A-D) receptors are type-I membrane proteins with extracellular regions containing two immunoglobulin domains and two thrombospondin domains (Figure VII.1A). UNC5 receptors bind to secreted Netrin ligands that act as repulsive cues during the development of the neural system to regulate neuronal migration (213). However, the migration of UNC5D-expressing neurons in the developing cortex, where netrin is not expressed, is regulated by FLRTs (67). The crystal structure of UNC5D in complex with the LRR domain of FLRT2 has revealed the binding interface, and it was suggested that during cortical development, the repulsive FLRT-UNC5 interaction mediates radial (upwards) migration of neurons, whereas the adhesive homophilic FLRT-FLRT interaction is important for the tangential (sideways) migration of neurons (214). However, the effect of other interactions on the function of these proteins, for example a possible simultaneous interaction of FLRT with LPHN3 was not clarified. In particular, the crystal structure suggests that FLRT dimers may exist in cis, casting doubt on the notion that hemophilic cell-adhesion can be mediated by FLRTs.

FLRTs and LPHNs work together with other proteins such as UNC5s to support the formation of proper neuronal connections. Although the FLRT/UNC5 crystal structure is available, the FLRT/ LPHN structure and well-characterized mutations that abolish a single interaction of each protein while maintaining the other interactions are not known. The presence of multiple binding partners and their compatibilities with each other, the formation of cis, trans or both interactions between proteins, and the pre-synaptic or post-synaptic localization of each protein are additional complications that

need to be addressed in order to decipher the mechanism of action of these proteins in neural development. We determined the three-dimensional structures of FLRT3 in isolation and in a complex with LPHN3 by X-ray crystallography. The structure of the complex reveals a large LPHN3-binding interface on the concave surface of FLRT3, which was previously reported as the dimerization surface. We designed multiple mutations on the LPHN3-binding surface of FLRT3 and on the FLRT3-binding surface of LPHN3 to specifically disrupt the interaction between them. We confirmed the wild type-like folding and trafficking of the mutant FLRT3s and mutant LPHN3s. The mutations that disrupt the FLRT3/ LPHN3 interaction without interfering with folding or trafficking confirm the binding interface revealed by the FLRT3/ LPHN3 complex structure. These mutants can be used as molecular tools to dissect the FLRT3/ LPHN3 interaction *in vivo*. We also studied the previously reported FLRT3 mutants called "FLRT3 dimerization mutant" and "UNC5 binding mutant" (214), and found that the FLRT3 dimerization mutant that was reported to abolish tangential (sideways) migration of neurons during cortex development is indeed defective in LPHN3 binding more severely than in dimerization. Our flow cytometry experiments showed that FLRT3, LPHN3 and UNC5 form a trimeric complex. FLRT3 binds the other two proteins simultaneously and bridges them, although LPHN3 and UNC5 do not directly interact with each other. Finally, we demonstrate that FLRT/LPHN pair as well as FLRT/UNC5 pair can interact in trans and cause cell-aggregation, suggesting that these protein-protein complexes are trans-synaptic complexes.

Results

Structure of the FLRT3/ LPHN3 complex

To determine the structure of the FLRT3/ LPHN3 complex, the extracellular LRR domain of human FLRT3 (residues K29-D357, corresponding to the N-terminal cap, LRR repeats and C-terminal cap) and the olfactomedin domain of LPHN3 (residues V132-G392) were produced in insect cells using the baculovirus expression system (residue numbering includes signal peptide even when it is absent). After purification of individual proteins, the complex was formed, and purified by size exclusion chromatography. The formation of the complex is independent of calcium as observed by native gel analysis (data not shown). Crystals of the complex were obtained in space group $P4_3$ with eight FLRT3/ LPHN3 complexes in the asymmetric unit, and diffracted to d_{min} = 3.6 Å (Table VII.1). The complex structure was obtained by molecular replacement using the available mouse FLRT3 structure (PDB ID 4V2E) and the mycolin olfactomedin structure (PDB ID 4WXQ, kindly provided by Raquel Lieberman before its release) and was further refined by the recently released human LPHN3 olfactomedin/lectin structure (PDB ID 5AFB) (214–216).

We also determined the crystal structure of the isolated human FLRT3 at a space group different than the previously published mouse FLRT3 (Figure S1). Crystals of FLRT3 were obtained in space group P2 with four molecules in the asymmetric unit and diffracted to d_{min}=2.6 Å. The structure was solved by molecular replacement using the available mouse FLRT3 structure as a model (PDB ID 4V2E) (Table VII.1). As described previously, FLRT3 has a horseshoe shape with dimensions ~83Å x 32Å x 40Å and contains ten LRR repeats (214).

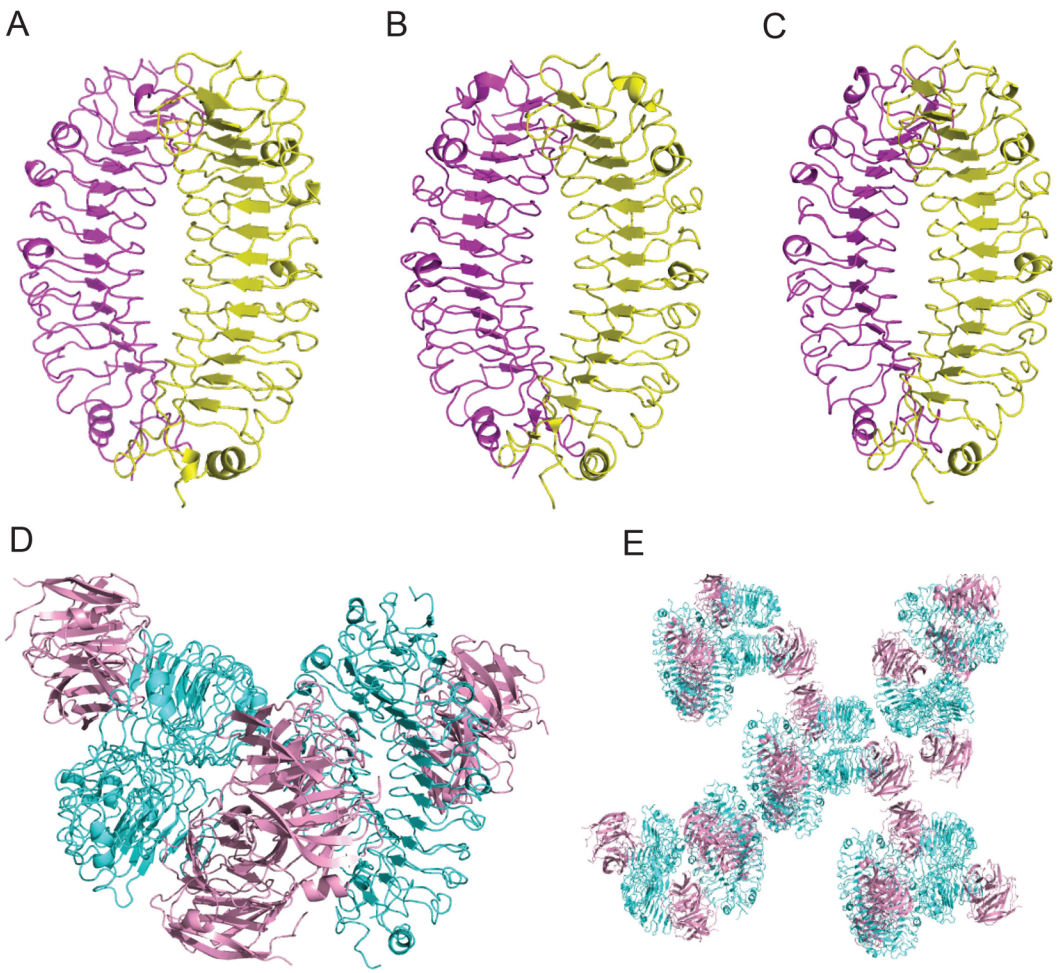


Figure VII.2 Crystal structure of the isolated FLRT3 LRR, and the crystal lattice of the FLRT3/LPHN3 complex structure.

(A) Ribbon diagram of the isolated human FLRT3 LRR in $P2_1$ space group. Two FLRT3 molecules make an interface on the side surface of FLRT3 where UNc5 binds. The previously reported mouse FLRT3 LRR dimer interface does not exist in our new isolated human FLRT3 structure. (B) Ribbon diagram of the published mouse FLRT3 LRR structure (PDB ID: 4V2E) showing the same interface as in A exists in both isolated FLRT3 structures. (C) Ribbon diagram of the FLRT3/LPHN3 complex structure showing only the FLRT3 LRR. Figure shows the same interface as in A and B also exists in the FLRT3/LPHN3 complex structure. The previously reported FLRT3 LRR dimer interface does not exist in the FLRT3/LPHN3 complex structure. These observations suggest that the observed interface in A,B and C can indeed be the dimer interface. It is also likely that the observed interface is a crystallization artifact that is preferred to generate crystal contacts. Careful mutagenesis experiments will answer these questions. (D) Ribbon diagram showing four FLRT3/LPHN3 complexes in the asymmetric unit of the complex structure. (E) A view of the FLRT3/LPHN3 complex structure crystal lattice.

The crystal structure of the FLRT3/LPHN3 complex comprises a heterodimer in which the five-bladed beta-propeller LPHN3 olfactomedin domain (dimensions ~52 Å x 43 Å x 43 Å) docks into the groove on the concave surface of the horseshoe-shaped FLRT3 LRR domain, creating a large buried surface area of 1550 $Å^2$ (Figure VII.1B,C). There were no other significant hetero-typic interactions in the crystal lattice (Figure Figure VII.2D,E, see Figure VII.2A-C for a discussion of homotypic interactions). Apart from sidechain rotamer changes, no major conformational changes are observed when the complex structure is compared with the individual structures of FLRT3 or LPHN3. N-linked glycosylation is observed at one (residue N226) glycosylation site on FLRT3, and three disulfide bonds (C31-C37, C35-C44, C309-C334) are observed in the FLRT3 LRR domain. A single disulfide bond (C135-C317) and no glycosylation is observed in LPHN3 Olfactomedin domain. The relative orientation of LPHN3 to FLRT3 is consistent with the previous prediction based on mutagenesis (216). However, numerous differences exist in the fine details of the molecular interactions.

The LPHN3-binding region on FLRT3 is located at the top of the inside concave surface of the LRR domain close to the N-terminal cap (Figure VII.1B,C). This region is away from the UNC5-binding region on FLRT3 (214). However, it is at the same side as the previously reported dimerization surface (see further below for the compatibility of these interactions with each other). The FLRT3-binding region on LPHN3 is located at the top/side surface of the olfactomedin domain and is mediated by the conserved loops primarily in the second and third blades. The blades of the olfactomedin domain are asymmetrical, and the fifth blade from where the N- and C-termini emerge is larger than the other blades, spanning as much space as almost two blades. This type of beta-

propeller structure was recently observed in the structure of olfactomedin domain of gliomedin as well (217).

There is electron density at the same position as of the Ca2+ ion in the isolated Olf-Lectin structure (PDB ID 5AFB). The FLRT3 binding site is opposite to the N- and C-termini of the olfactomedin domain which connect to the lectin and STP/HormR/GAIN/TM domains, respectively. The alternatively spliced sequence (residues K127-K131) between the lectin and olfactomedin domains of LPHN3 is located immediately N-terminal of the olfactomedin domain, distant from the FLRT3 binding region, explaining why its presence does not affect FLRT3 binding to LPHN3 (65). The FLRT3/ LPHN3 complex structure also explains why FLRT binding is mediated by only the olfactomedin domain (199), as opposed to teneurin binding that is mediated by both the lectin and olfactomedin domains and the splice insert between them (65). As both the N- and C-termini of the olfactomedin domain reside next to each other, it is not possible to conclude from the FLRT/LPHN structure whether the FLRT3/ LPHN3 interaction is in trans or cis.

The sequence identities between FLRT LRR domains and the sequence identity between the LPHN Olfactomedin domains are high (Table VII.2, Table VII.3). Although invertebrate LPHN homologs lack an olfactomedin domain, it is the most conserved extracellular domain of LPHNs in vertebrates, suggesting an essential function. Similarly, no direct FLRT homolog is readily identifiable in invertebrates. To visualize conserved and variable regions of FLRT LRR domains and LPHN Olfactomedin domains, we mapped the conservation of residues on the FLRT3/LPHN3 complex structure, and colored residues from most conserved to least conserved. The binding

surfaces of FLRTs and LPHNs to each other correspond to one of the most conserved regions (yellow ovals in Figure VII.1D). Visualization of the electrostatic surface potential on the complex structure shows that the inside concave surface of the LRR domain is positively and negatively charged on separate sides, suggesting that charge-mediated interactions are likely to be mediated by this concave surface (Figure VII.1E). The binding interface of FLRT and LPHN is not highly charged.

Mutations at the binding surface disrupt the interaction of FLRT3 and LPHN3

The high affinity of the FLRT3/ LPHN3 complex (K_d = 220 nM-260 nM as determined by Bio-layer interferometry (BLITZ), Figure S2,(61)) is achieved by a combination of interactions, comprised of hydrogen bonds, salt bridges, hydrophobic interactions and long-range electrostatic interactions (Figure VII.3A). The docking of the complementary surfaces of the round-shaped Olfactomedin domain into the concave surface of the LRR domain creates a large interface. Our observed binding affinity and the observed interface area of the LPHN3/FLRT3 complex are in line with a study that demonstrates correlation of the binding affinity with interface area (218). Numerous polar residues are at the binding interface (Figure VII.4A).

Mapping the residues that are mutated in some attention deficit hyperactivity disorder (A247S in human LPHN3 (219)) and cancer cases (R196C, A278P and A278T in human LPHN2 (208, 209)) on the FLRT3/ LPHN3 structure reveals that A247S and A278P/T mutations map close to the binding interface, and thus may cause a direct defect in binding to FLRT3 (Figure VII.4B,C). The R196C mutation is at the distant

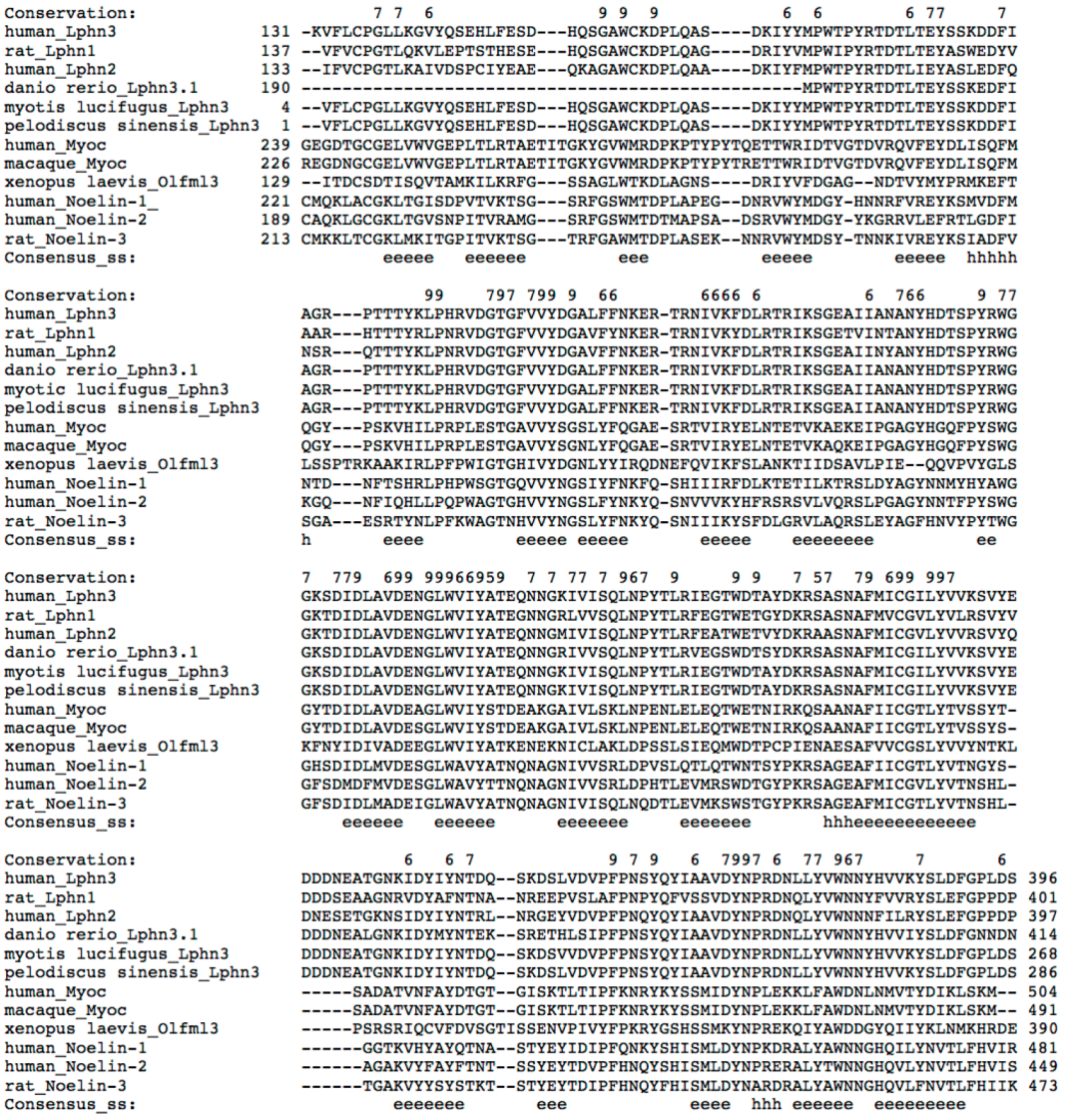


Table VII.2 Sequence alignment of olfactomedin domains

Sequence alignment of olfatomedin domains using PROMALS3D: human LPHN3 (Uniprot ID: Q9HAR2), rat LPHN1 (O88917), human LPHN2 (O95490), *daniorerio* LPHN3.1 (F1QAS7), *myotislucifugus* LPHN3 (G1P220), *pelodiscussinensis* LPHN3 (K7F4U8), human Myocilin (Q99972), *macacafascicularis*Myocilin (Q863A3), *xenopuslaevis* Olfml3 (B5MFE9), human Noelin1 (Q99784), human Noelin2 (O95897), rat Noelin2 (Q568Y7). The first line in each block shows conservation indices with 9 being the most conserved and 7 and 6 being less conserved. Consensus predicated secondary structures are represented by *h*, alpha-helix, or *e*, beta-strandview of the FLRT3/LPHN3 complex structure crystal lattice.

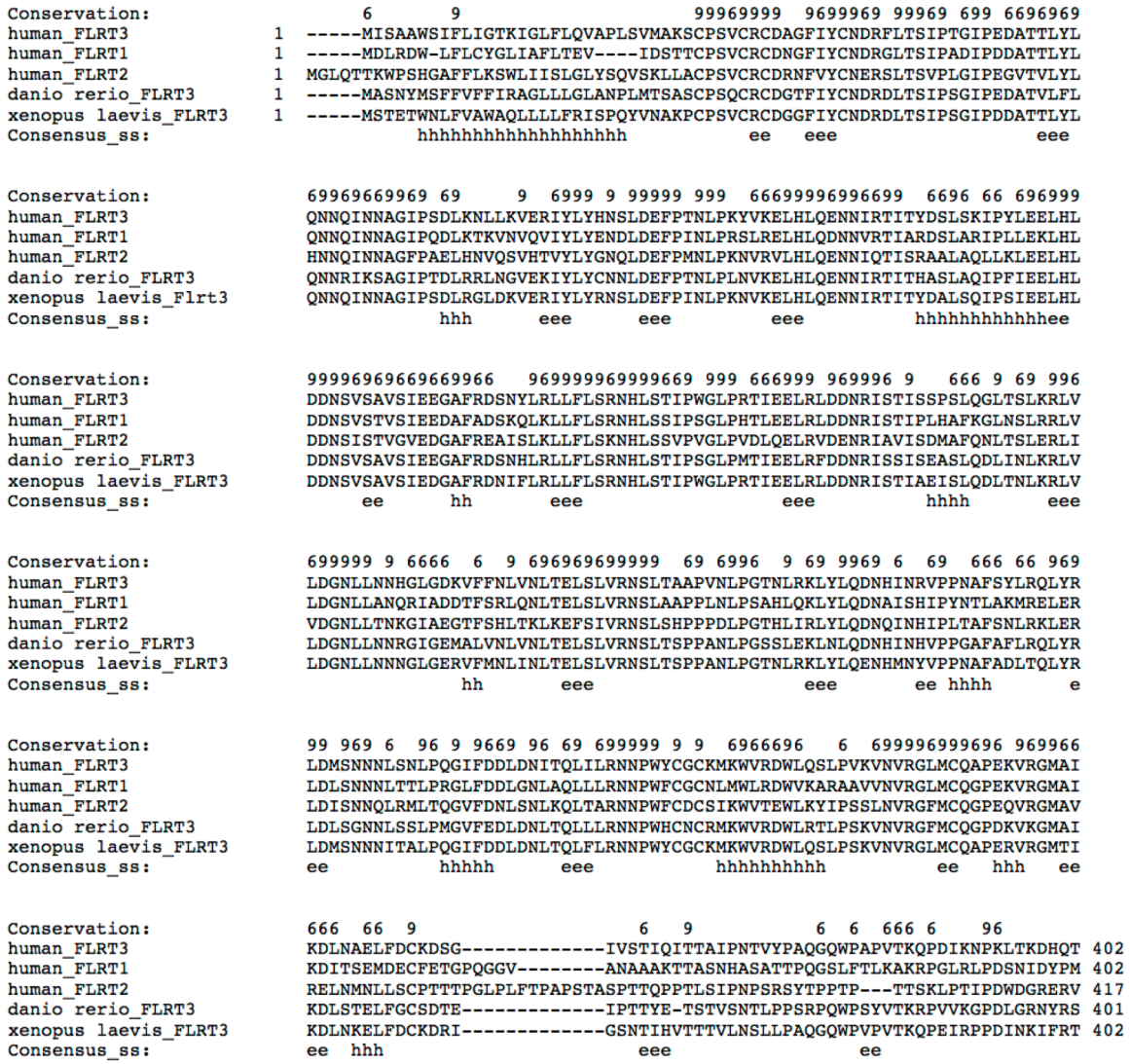


Table VII.3 Sequence alignment of LRR repeats of FLRT homologs

Sequence alignment of FLRT3 LRR domains using PROMALS3D: human FLRT3 (Uniprot ID: Q9NZU0), *xenopus laevis* (B7ZR53), *danio rerior* (B8A507), human FLRT2 (O43155), and human FLRT1 (Q9NZU1). The first line in each block shows conservation indices with 9 being the most conserved and 7 and 6 being less conserved. Consensus predicated secondary structures are represented by h, alpha-helix, or e, beta-strand.

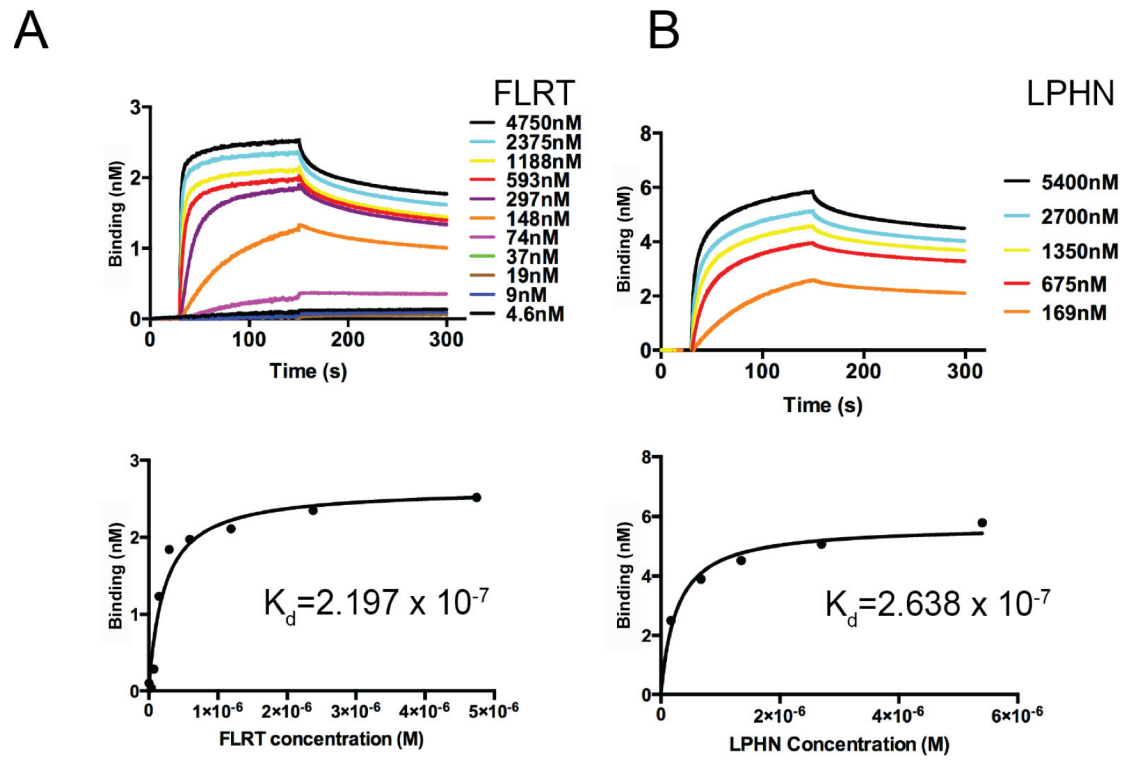


Figure VII.3 Bio-layer interferometry binding measurement of FLRT3 and LPHN3.

LPHN3 and FLRT3 bind to each other in the nanomolar range. Maximum binding of His-tagged FLRT3 LRR to immobilized biotinylaed LPHN3 Olf (A), or His-tagged LPHN3 to immobilized biotinylated FLRT3 LRR (B), measured at steady state during immersion of streptavidin sensor immobilized with biotinylated FLRT3 LRR or bitinylated LPHN3 Olf into various concentrations of His-LPHN3 Olf or His-FLRT3 LRR, respectively. Binding signal was plotted as a function of His-FLRT3 LRR or His-LPHN3 Olf concentration. The dissociation constants were calculated by fitting the curves to a single site binding model using GraphPad Prism.

edge of the binding interface. Interestingly, the A278P/T mutation that is very close to the binding interface is mutated to two different residues in cancers and is reported by two different studies (208, 209).

In order to specifically abolish a single interaction without interfering with other interactions or cell-surface localization of FLRT3 and LPHN3, and to confirm the validity of the binding interface that is observed in the FLRT/LPHN complex structure, we

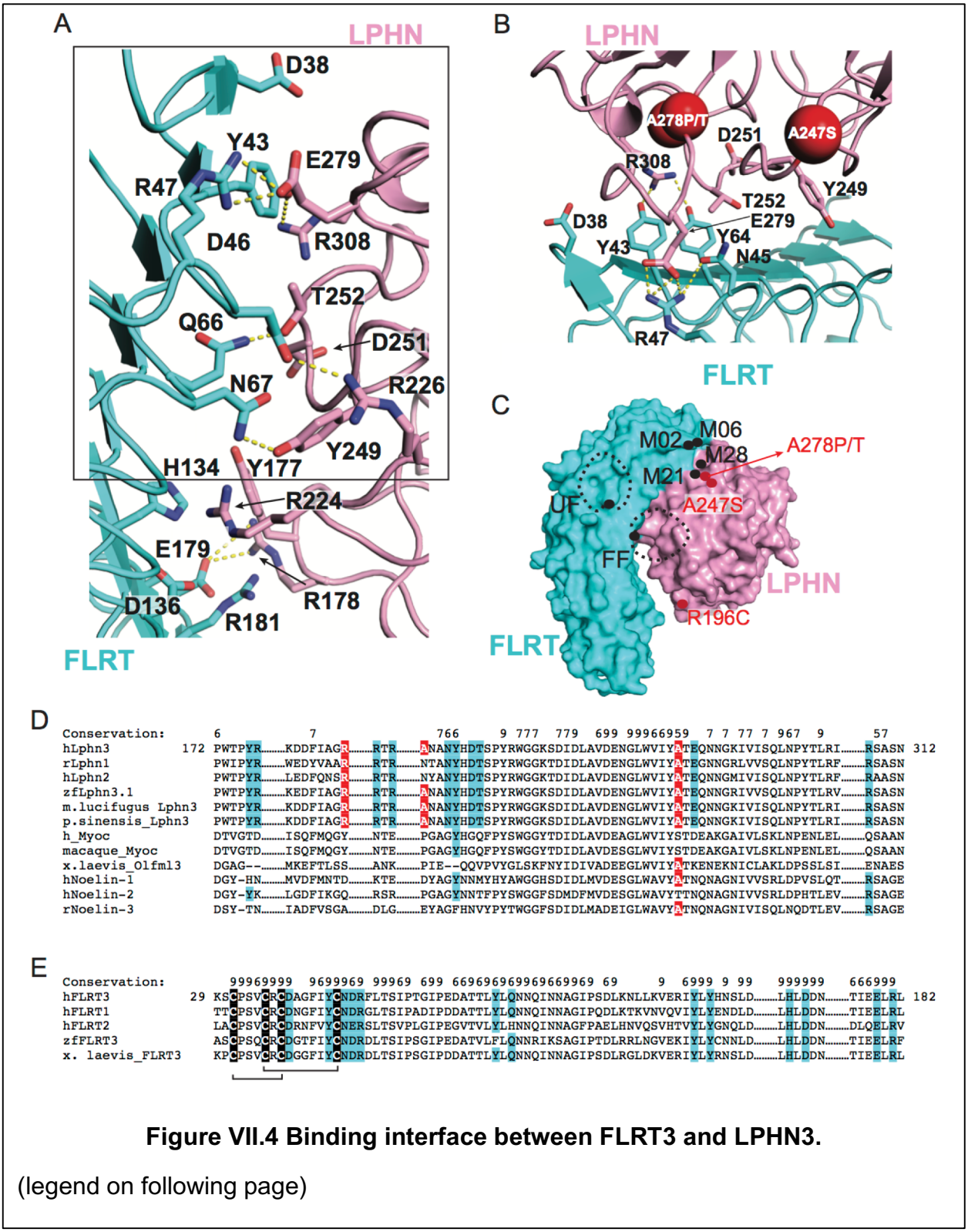


Figure VII.4 Binding interface between FLRT3 and LPHN3.

(legend on following page)

Figure VII.4, continued.

A) Ribbon diagram showing the entire interface between FLRT3 and LPHN3. Polar interactions are shown by dashed lines. Residues at the binding interface are shown as sticks. B) Close-up view of the binding interface between FLRT3 and LPHN3 (box in A). Labeled residues correspond to the residues that were mutated in the selected M02, M06, M21 and M28 mutants. Red spheres show the locations of the attention deficit hyperactivity disorder and cancer mutations on LPHNs. (C) Schematic drawing of locations of mutations studied in this manuscript (labeled black) and three disease mutations (labeled red). Dashed black spheres indicate the N-linked glycosylation moieties introduced as a result of mutagenesis. Carbohydrates on FLRT3-FF mutant clashes with LPHN3. (D) Sequence alignment of the FLRT binding site on different olfactomedin domain sequences. The key residues involved in binding to FLRT are highlighted in cyan. Disease mutations are highlighted in red. The residues essential for binding to FLRTs are conserved only in LPHNs. Conservation of each residue is labeled. Value 9 corresponds to highest conservation. (E) Sequence alignment of the LPHN binding site on FLRT sequences. The conserved cysteines are highlighted in black. The key residues involved in binding to LPHN are highlighted in cyan. Disulfide bonds are indicated by black lines.

designed surface mutations on full-length FLRT3 (FL-FLRT3) and full length LPHN3 (FL-LPHN3) that change only a few atoms on the protein surface rather than introduce large posttranslational modifications (Figure VII.4B). We designed mutations on LPHN3 to abolish binding to FLRT3 and mutations on FLRT3 to abolish binding to LPHN3. In designing all of these mutants, we tried to avoid any interference with the folding of the proteins. We also studied the previously published FLRT3 dimerization mutant (“FF”: R181N, D183T) and the FLRT3 mutant that abolishes UNC5 binding (“UF”: H165N). (See Table VII.4 for a full list of all mutations on the complex structure; see Figure VII.4C for a schematic representation of select important mutations; see Figure VII.4D,E for conservation of the residues at the binding sites).

hLPHN3

Mutant name	Mutations	Cell Surface Localization	FLRT3 LRR Binding
	wild type	+++	+++
M19	Y249A	+++	+++
M20	Y249A, D251A, T252A	+++	+
M21*	Y249A, D251A, T252A, R308A	+++	-
M22	Y177A, Y249A, R205A	++	+
M23	N248A, Y249A	+	-
M24	N248A, Y249A, E279A	+	-
M27	D251A, T252A	++	+
M28*	Y249A, D251A, T252A, E279A, R308A	+++	-

hFLRT3

Mutant name	Mutations	Cell Surface Localization	LPHN3 Olf Binding	Dimerization
	wild type	+++	+++	+++
M01	Y89A, Y91A	++	-	
M02	Y43A, Y64A	+++	-	+++
M03	Y43A, Y64A,Y89A, Y91A	+	-	
M04	Y64A, Q66A, N67A	-	-	
M05	Y43A, N45A, R47A	+++	-	
M06	Y43A, N45A, R47A, D38A	+++	-	
M08	R117A, E113A	+++	++	+++
M11	R181A, D183A			+++
M26	Q66A, N67A	+	-	
M35	Y43A, Y89A, Y91A	+	-	
FLRT UF	H165N	+++	+++	+++
FLRT FF	R181N, D183T	+++	-	+

* 100 nM tetramerized FLRT3 LRR binding shown for M21 and M28 mutants (Fig 3). 10 μM monomeric FLRT3 LRR binding shown for all other FL-LPHN3 mutants (Fig S4).

Table VII.4 Summary of mutations and their characteristics.

We first examined the expression levels and surface transport of all FL-FLRT3 and FL-LPHN3 mutants in order to eliminate misfolded mutants that are likely to be poorly expressed, and unlikely to reach the cell surface. The FL-FLRT3 and FL-LPHN3 had extracellular N-terminal Myc and Flag tags, respectively, to allow for detection of expression levels and cell-surface localization of wild type and mutant proteins. We expressed wild-type and mutant full-length FLRT3 and LPHN3 in transfected HEK293 cells. Cells were stained without detergent permeabilization (to label only the cell-surface localized protein), with an antibody suitable to react with the extracellular tag on the proteins, and the amount of surface-exposed FL-FLRT3 or FL-LPHN3 was detected by indirect immunofluorescence using flow cytometry (Figure VII.5, Figure VII.6, Figure VII.7, Figure VII.8, Figure VII.9, and Table VII.5).

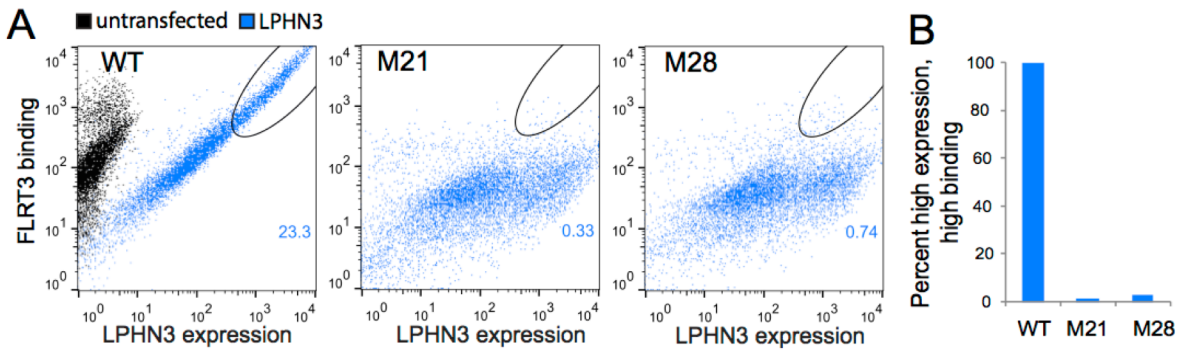


Figure VII.5 Mutations on LPHN3 abolish FLRT3 binding but do not interfere with proper cell-surface localization.

(A) Wild-type and mutant full-length LPHN3 proteins were tested for surface expression in HEK293 cells as well as their ability to bind soluble FLRT3 LRR domain using flow cytometry. Non-permeabilized HEK293 cells expressing N-terminally FLAG-tagged FL-LPHN3 were stained with mouse anti-FLAG primary antibody and anti-mouse FITC conjugated secondary antibody. LPHN3 surface expression was measured as green fluorescence from FITC (x-axis). FLRT3 LRR binding to FL-LPHN3-expressing cells was measured by monitoring red fluorescence of DyLight attached to neutravidin (y-axis). (refer to Figure S3B and Table S3 for detailed experimental setup and experimental conditions). Dot plot shows correlation between LPHN3 expression and FLRT3 binding in LPHN3-transfected cells (blue), or in untransfected cells (black). Black ovals on the plot show the “high LPHN3 expression and high FLRT3 binding” gate. The number on the plot represents the percent of all events that is in the high expression and high binding gate. 100nM purified wild-type His-tagged FLRT3 LRR domain was incubated with Biotin-Tris-NTA and was tetramerized with neutravidin to increase avidity before binding to cells (See setup in Figure S3B). Mutant FL-LPHN3 constructs (M21, M28) show proper surface expression (x-axis), but do not bind FLRT3 LRR (y-axis). (B) Quantification of cells that fall within the gate of “high LPHN3 expression and high FLRT3 binding” (black ovals) as indicated in A normalized to wild type. Bar height represents the percent of cells that fall within the gates shown.

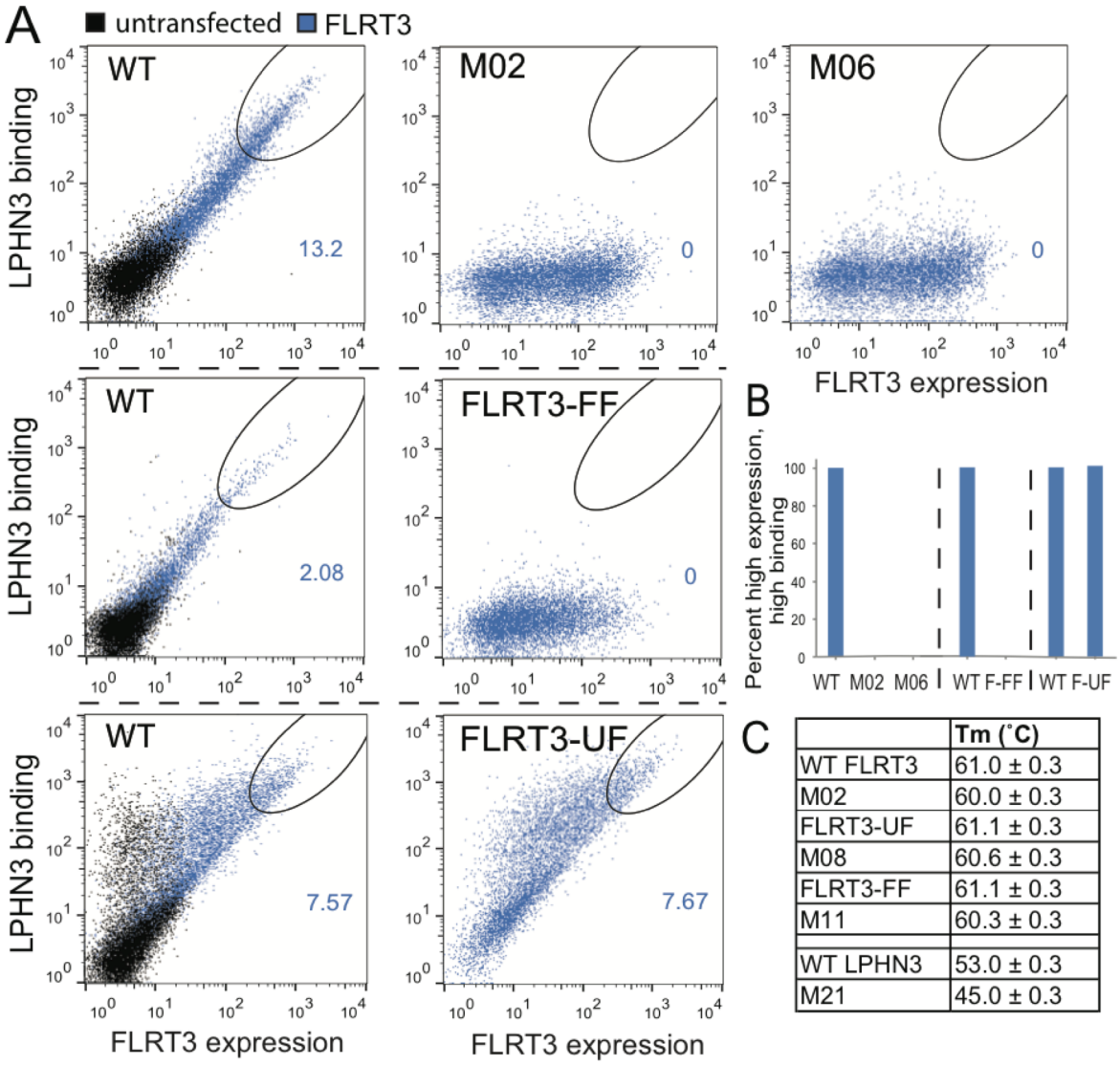


Figure VII.6 Mutations on FLRT abolish LPHN3 binding but do not interfere with proper cell-surface localization.

(legend on the following page)

Figure VII.6, continued.

(A) Wild-type and mutant full-length FLRT3 proteins were tested for surface expression in HEK293 cells as well as their ability to bind soluble LPHN3 Olf domain using flow cytometry. Non-permeabilized HEK293 cells expressing N-terminally Myc-tagged FL-FLRT3 were stained with mouse anti-c-Myc primary antibody and anti-mouse FITC conjugated secondary antibody. FLRT3 surface expression was measured as green fluorescence from FITC (x-axis). Biotinylated LPHN3 Olf binding to FL-FLRT3-expressing cells was measured by monitoring red fluorescence of DyLight attached to neutravidin (y-axis). (refer to Figure S3C and Table S3 for detailed experimental setup and experimental conditions). Dot plot shows correlation between FLRT3 expression and LPHN3 binding in FLRT3-transfected cells (blue), or in untransfected cells (black). See legend for Figure 3 for details. See setup in Figure S3C. Mutant FLRT3 constructs (M2, M6, FLRT-FF and FLRT-UF) show proper surface expression (x-axis), but do not bind LPHN3 (y-axis) except FLRT3-UF. Different sets of experiments are separated with dashed black lines. Wild type data is provided for each set of experiments as positive control. (B) Quantification of cells that fall within the gate of “high FLRT3 expression and high LPHN3 binding” (black ovals) as indicated in A normalized to wild type. Bar height represents the percent of cells that fall within the gates shown. See also Figure S3. (C) Differential scanning fluorimetry (DSF) of wild-type and mutant FLRT3 LRR and LPHN3 olfactomedin domains showing the melting temperature (Tm).

We next measured soluble FLRT3 LRR binding to all FL-LPHN3 mutants, using surface binding of recombinant FLRT3 LRR domain to HEK293 cells transfected with full-length wild type or mutant LPHN3s using flow cytometry (Figure VII.5 and Figure VII.8, see Figure VII.7 and Table VII.5 for details). To ensure the mutations completely abolish binding, low and high concentrations of wild type recombinant FLRT3 LRR were used in binding experiments. Specifically, by staining with purified FLRT3 LRR at concentrations as low as 10 nM, only high-affinity interactions can be detected. By staining at concentrations as high as 10 μM, lower affinity interactions can be detected. Additionally, by staining with tetramerized FLRT3 LRR, the avidity effect increases the

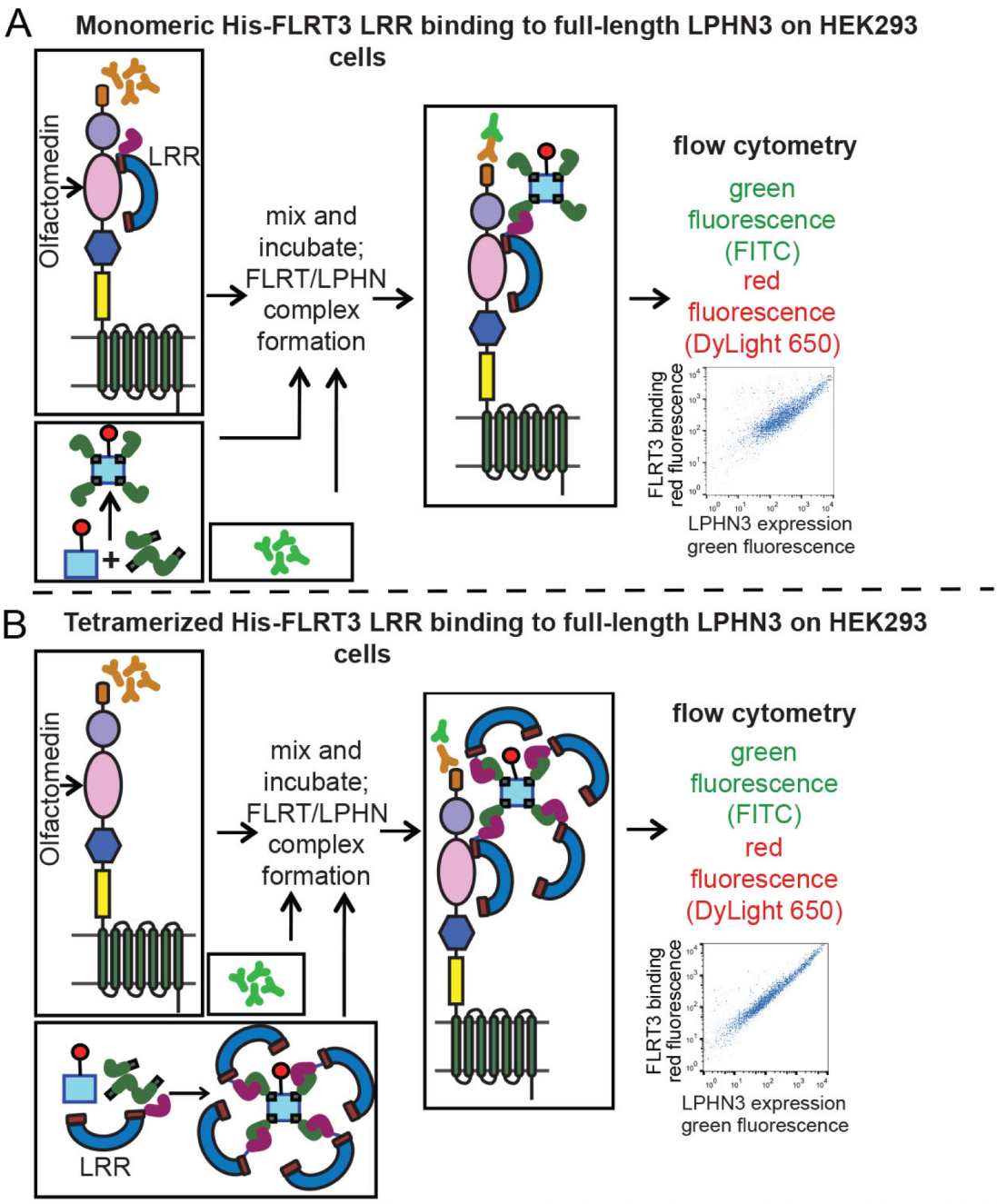


Figure VII.7 Experimental setup for flow cytometry experiments.

(continued on the following page)

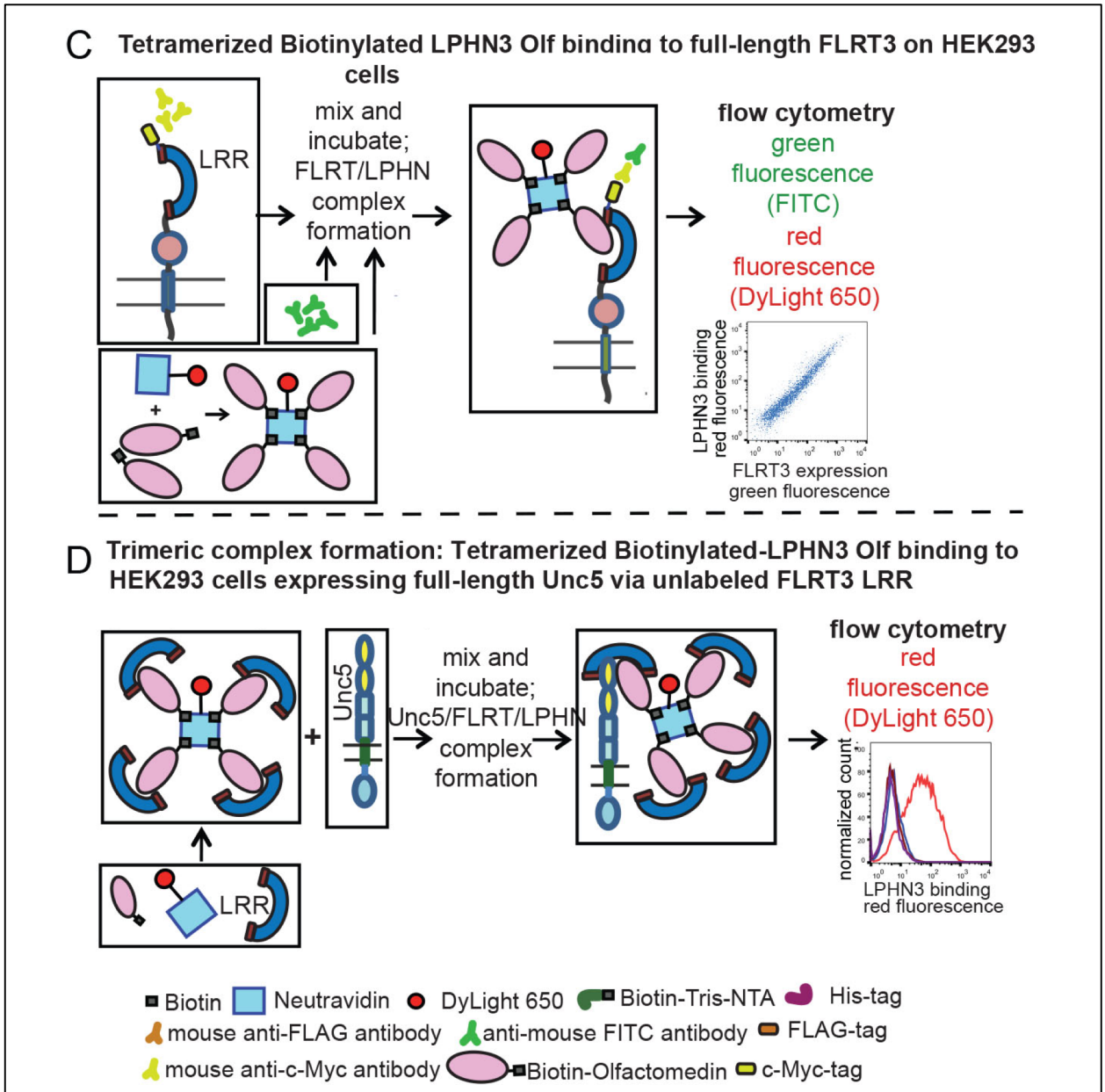


Figure VII.7, continued.

(legend on the following page)

Figure VII.7, continued.

N-terminally tagged full-length proteins (Flag-LPHN3, Myc-FLRT3, His-Unc5D or His-Unc5B) were expressed in HEK293 cells. Cells were then mixed with different combinations of soluble proteins and reagents before performing flow cytometry experiments. Black boxes around proteins or reagents indicate they were pre-incubated before mixing with the other components. Neutravidin is a tetrameric protein with four biotin-binding sites and was used to tetramerize the biotinylated proteins to increase the affinity of the protein due to the avidity effect. Biotin-Tris-NTA is a reagent (kindly provided by the S. Koide lab) to convert His-tagged soluble proteins to biotinylated proteins (It was used when expression/purification of biotinylated proteins was not optimal). Detection of full-length protein expression on the cell surface was monitored by green fluorescence of FITC-conjugated secondary anti-mouse antibody by flow cytometry. Monomeric or tetramerized soluble protein binding to the cells was monitored by red fluorescence of DyLight-conjugated neutravidin by flow cytometry.

Setup in Figure VII.8: Monomeric His-FLRT3 LRR binding to HEK293 cells expressing full-length wild type or mutant LPHN3. Biotin-Tris-NTA was used to recognize His tagged FLRT3 LRR.

Setup in Figure VII.5: Tetramerized His-FLRT3 LRR binding to HEK293 cells expressing full-length wild type or mutant LPHN3. His-FLRT3 LRR is tetrametrized by preincubation with Neutravidin and Biotin-Tris-NTA.

Setup in Figure VII.6 and Figure VII.9: Tetramerized biotinylated-LPHN3 olf binding to HEK293 cells expressing full-length wild-type or mutant FLRT3. Biotin-tagged LPHN3 Olf was first biotinylated by in vitro biotibyation. Biotinylated LPHN3 Olf is tetrametrized by preincubation with Neutravidin.

Setup for Figure VII.12 and Figure VII.14: Tetramerized biotinylated-LPHN3 Olf binding to HEK293 cells expressing full-length wild-type Unc5D or Unc5B. Biotinylated LPHN3 Olf and non-biotinylated His-tagged FLRT3 LRR is tetrametrized by preincubation with Neutravidin. The formation of a trimeric complex of His-FLRT3 LLR, biotinylated LPHN3 olf domain and FL-Unc5 expressed by HEK293 cells is observed.

effective concentration by over 100-fold (220) (His-FLRT3 was tetramerized by mixing with biotin-tris-NTA (BTtrisNTA) and neutravidin -a protein that binds to biotin and tetramerizes it). Thus, by observing no binding of 100 nM tetramerized FLRT3 LRR to

mutant FL-LPHN constructs, we concluded that the affinity for wild-type FLRT3 LRR for these LPHN3 mutants was very low (i.e. K_d > 10 μM).

These experiments showed that FL-LPHN3-M21 (Y249A, D251A, T252A, R308A) and FL-LPHN3-M28 (Y249A, D251A, T252A, E279A, R308A) mutants were defective in FLRT LRR binding and had no surface localization problems (Figure VII.5A,B). Similar experiments were performed for testing wild type recombinant LPHN3 Olf domain binding to HEK293 cells transfected with full-length wild type or mutant FLRT3s (Biotinylated LPHN3 Olf was tetramerized by mixing with neutravidin).. FL-FLRT3-M02 (Y43A, Y64A) and FL-FLRT3-M06 (D38A, Y43A, N45A, R47A) mutants were defective in LPHN3 Olf binding and had no surface localization problems (Figure VII.6A,B and Figure VII.9, see Figure VII.7 and Table VII.5 for details). To further confirm the proper folding and stability of the mutants, we also performed differential scanning fluorimetry experiments with select mutants that measure the stability of proteins by monitoring the melting temperature (Figure VII.6C). Mutant FLRT3 LRR and mutant LPHN3 Olf proteins expressed and purified in baculovirus expression system were used for these experiments. Together, these mutants that are transported to the cell surface and exhibit either a defect in FLRT3 binding or in LPHN3 binding, or no apparent defect provide us with a toolkit to explore the function of LPHN3 and FLRT3 in the neuronal activities of these cell-adhesion proteins.

Experiment purpose	Reagents used				
Cell-surface expression	**primary antibody**		**secondary antibody**		**Fluorophore**
FLAG-LPHN3 full length	mouse anti-FLAG 1:1000		anti-mouse FITC 1:100		FITC
Myc-FLRT3 full length	mouse anti-c-Myc 1:20		anti-mouse FITC 1:100		FITC
Detect soluble protein binding to cells	**neutravidin DyLight 650**	**biotin-tris-NTA**	**Biotinylated-LPHN3 Olf**	**His-FLRT3 LRR**	
monomeric His-FLRT3 LRR	100 nM	100 nM	-	10 µM	DyLight 650
tetramerized His-FLRT3 LRR	100 nM	100 nM	-	100 nM	DyLight 650
tetramerized Biotinylated-LPHN3	100 nM	-	100 nM	-	DyLight 650
Trimeric complex Unc5/FLRT3/LPHN3	5 nM/ 100 nM	-	5 nM/ 100 nM	5 nM/ 100 nM	DyLight 650

Table VII.5 Summary of experimental conditions for flow cytometry.

Table showing experimental conditions used for performing the flow cytometry experiments in the related figures. Experimental conditions for both detecting cell-surface expression and for detecting soluble protein binding to surface expressed proteins are summarized. The schematic diagram explaining the experiment and the data related to the experiment are indicated in separate columns.

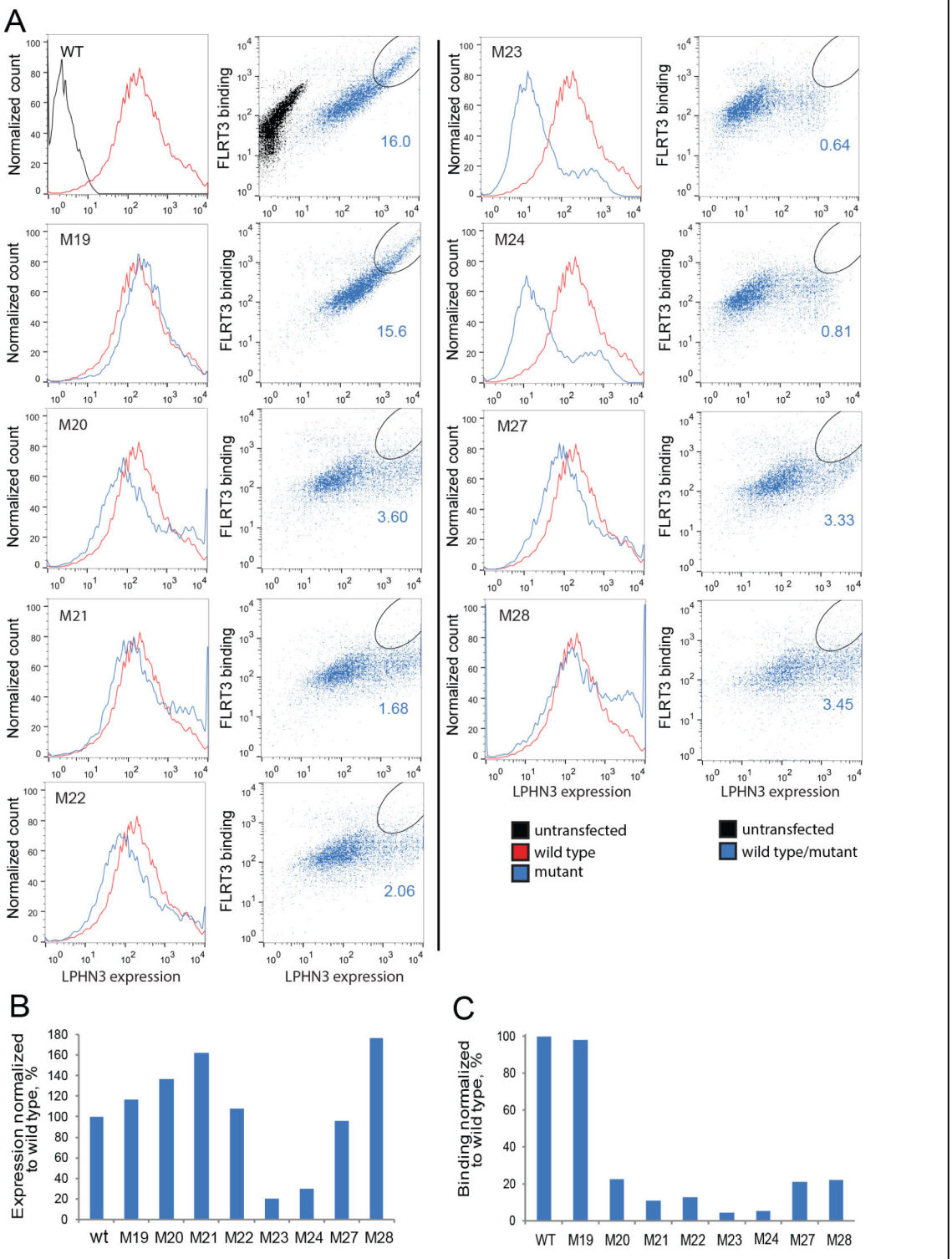


Figure VII.8 Expression and FLRT3 LRR-binding data for all FL-LPHN3 mutants.

(legend on the following pages)

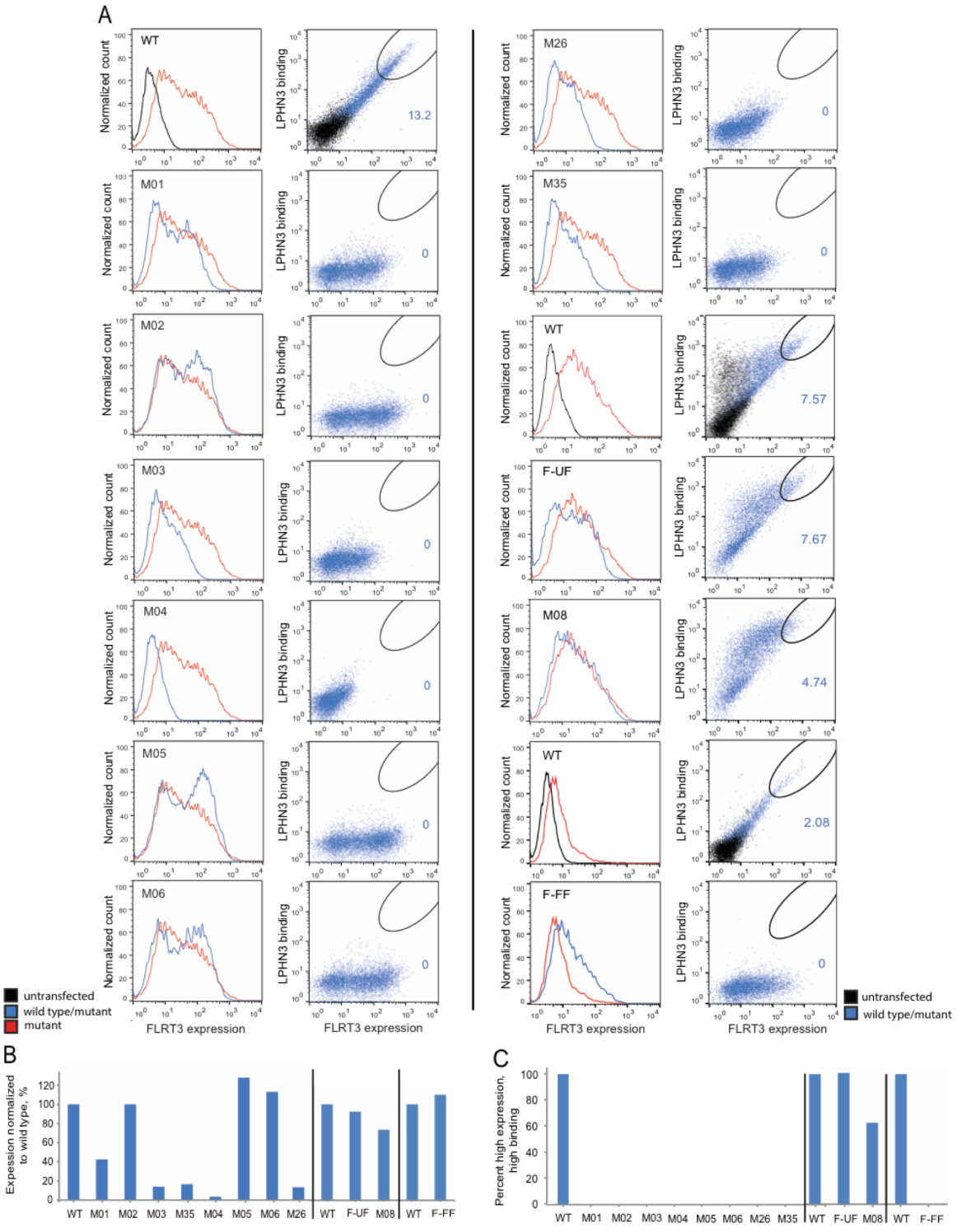


Figure VII.9 Expression and LPHN3 Olf-binding data for all FL-FLRT3 mutants.

(legend on the following page)

Figure VII.8, continued.

Wild-type and mutant full-length LPHN3 constructs were tested for surface expression in non-permeabilized HEK293 cells as well as their ability to bind to soluble His-FLRT3 LRR domain using flow cytometry. (A) FL-LPHN3 surface expression in wild-type FL-LPHN3-transfected cells (red) compared to untransfected cells (black), and to mutant FL-LPHN3-transfected cells (blue) (left panels). Non-permeabilized HEK293 cells expressing N-terminally FLAG-tagged FL-Lphn3 were stained with mouse anti-FLAG primary antibody and anti-mouse FITC conjugated secondary antibody. FL-LPHN3 surface expression was measured as green fluorescence from FITC. His-FLRT3 LRR binding to FL-LPHN3-expressing cells was measured by monitoring red fluorescence of DyLight attached to neutravidin (y-axis, right panles). 10 μM non-tetramerized His-FLRT3 LRR was mixed with the cells, then 100 nM Biotin-Tris-NTA and 100 nM Neutravidin was added for detection of His-FLRT3 LRR binding. Black ovals show the “high LPHN3 expression and high FLRT3 binding” gate. Numbers on the dot plot show percent of events in this gate. (Refer to Figure VII.7 and Table S3 for detailed experimental setup and experimental conditions). (B) Surface expression of mutant LPHN3 transfected cells normalized to wild type. (C) Quantification of cells that fall within the gate of “high LPHN3 expression and high FLRT3 binding” (black ovals) as indicated in A. Bar height represents the percent of cells that fall within the gates shown.

Figure VII.9, continued.

Wild-type and mutant full-length FLRT3 constructs were tested for surface expression in non-permeabilized HEK293 cells and their ability to bind to biotinylated LPHN3 Olf domain using flow cytometry. (A) FL-FLRT3 surface expression in wild-type FL-FLRT3-transfected cells (red) compared to untransfected cells (black), and to mutant FL-FLRT3-transfected cells (blue; left panels). Non-permeabilized HEK293 cells expressing N-terminally Myc-tagged FL-FLRT3 were stained with mouse anti-c-Myc primary antibody and anti-mouse FITC conjugated secondary antibody. Biotinylated LPHN3 Olf binding to FL-FLRT3-expressing cells was measured by monitoring red fluorescence of DyLight attached to neutravidin (y-axis, right panles). 100nM purified wild-type biotinylated LPHN3 Olf domain was tetramerized with 100 nM Neutravidin to increase avidity before binding to cells. Black ovals show the “high FLRT3 expression and high LPHN3 binding” gate. Numbers on the dot plot show percent of events in this gate. (B) Surface expression of mutant FLRT3 transfected cells normalized to wild type.(C) Quantification of cells that fall within the gate of “high FLRT3 expression and high LPHN3 binding” (black ovals) as indicated in A. Bar height represents the percent of cells that fall within the gates shown. Separate sets of experiments are indicated by vertical lines. Wild type data is provided as positive control for each dataset.

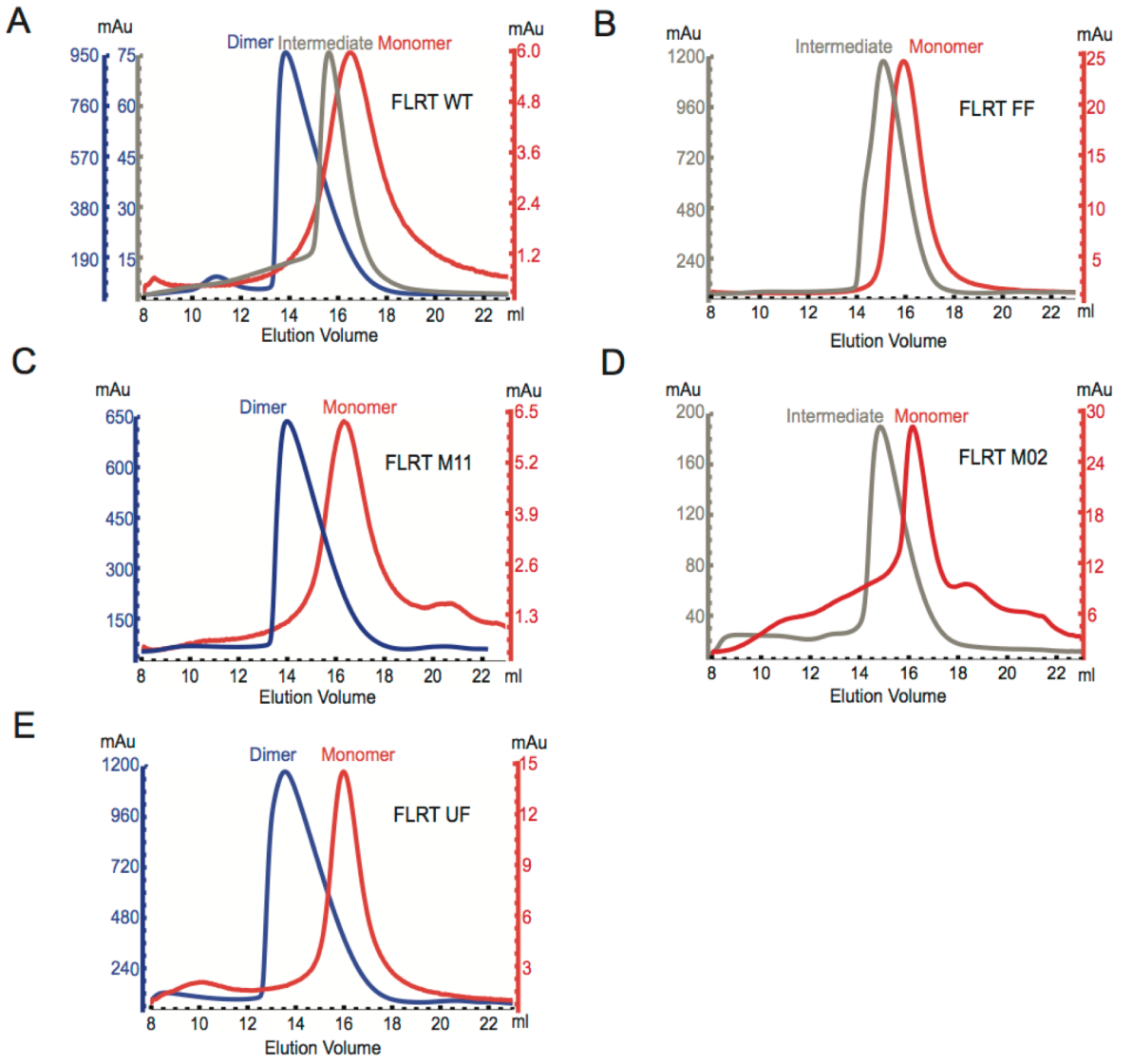


Figure VII.10 Mutations on LPHN3 abolish FLRT3 binding but do not interfere with proper cell-surface localization.

Size-exclusion gel filtration profile showing elution volumes of wild type FLRT3 LRR domain or mutants at various protein concentrations. Protein concentration is monitored by the UV absorbance at 280 nm (mAu) indicated by the y-axis. Different colors of the y-axis match the different elution peaks. Blue, gray, and red curves indicate dimer, intermediate, or monomer FLRT3 LRR, respectively. Note that M2 mutant displays wild type behavior. Due to low protein yield, the protein could not be concentrated to high enough concentrations to form dimers. However, the observed intermediate peak elutes at similar concentrations as wild type (note similar mAU).

We also studied the previously reported FLRT3 mutants called “FLRT3-FF dimerization mutant” and “FLRT3-UF UNC5 binding mutant” (see Figure VII.4C for a schematic illustration of their location on the complex structure) (214). These mutations, however, were generated by introducing large N-linked carbohydrate modifications into the protein surface to disrupt interactions. In vivo studies performed with these mutants had reported that FLRT dimerization is involved in tangential (sideways) migration of neurons during cortex development, whereas repulsive FLRT-UNC5 interaction is involved in radial (upward) migration of neurons. Our further analysis of these mutants showed that FLRT3-FF decreases dimerization although it does not abolish it (see below, Figure VII.10A,B). Importantly, it completely abolishes the binding of LPHN3 LRR to FL-FLRT3 (Figure VII.6A,B) suggesting that the in vivo effect of the FLRT3-FF mutant on tangential migration of neurons is likely due to the lack of LPHN3 binding to FLRT3. The UNC5-binding mutant FLRT3-UF, on the other hand, had no detectable effect on LPHN3 binding (Figure VII.6A,B).

Effect of mutations on FLRT3 dimerization

Next, we tested the effect of the FLRT3 mutations on FLRT3 dimerization. It was previously reported that gel filtration chromatography can detect the concentration dependent dimerization of FLRT3 LRR (214). As expected, the size exclusion chromatogram of purified FLRT3 LRR showed that the protein concentration affects the elution volume, indicating that at high protein concentrations FLRT3 LRR forms a dimer, whereas at low protein concentrations FLRT3 LRR is a monomer (Figure VII.10A, notice the correlation of the elution volume with the absorbance units (mAU) reflective of

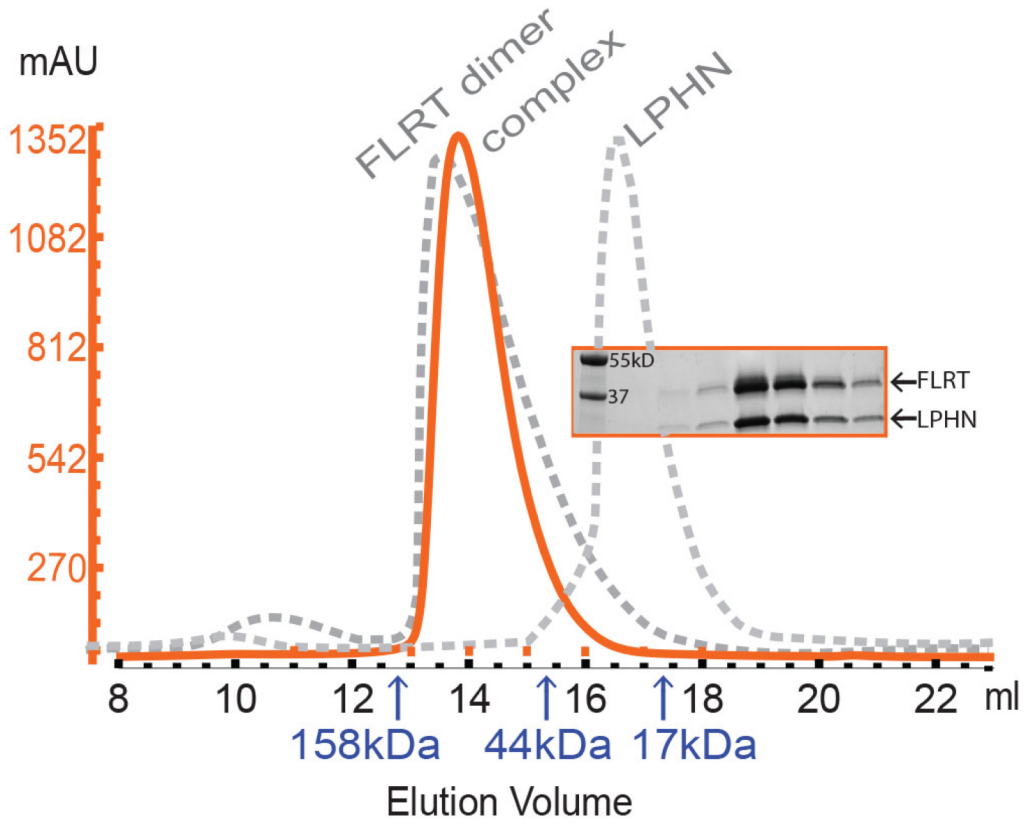


Figure VII.11 Size-exclusion gel filtration profile showing the formation of FLRT3/LPHN3 complex.

Dotted lines represent FLRT LRR dimer and LPHN3 Olf for reference. SDS-Page gel shows fractions corresponding to the complex peak. The elution volumes and corresponding molecular weights for the gel filtration standards are indicated by blue arrows.

protein concentration, see Figure VII.11 for elution profiles of gel filtration standards, dimeric FLRT3 LRR, LPHN3 Olf and the FLRT3/LPHN3 complex). At intermediate concentrations, the eluted peak was in between the dimer and monomer elution volumes. We speculate that the monomer/dimer has a fast exchange rate giving rise to a single intermediate peak rather than a mixture of monomer/dimer peaks. When dimer was diluted and reran on the gel filtration column, a monomer peak was observed. Similarly, when the monomer peak was concentrated and reran, a dimer peak was

observed, showing that the dimer-monomer formation is reversible. Based on the measured protein concentrations, a low affinity in the submicromolar range is predicted for the FLRT3 LRR dimer in vitro. However, the affective affinity of two FL-FLRT3 monomers involved in a possible cis-interaction on the cell surface is likely higher.

We used gel filtration chromatography to monitor the effect of the studied FLRT mutations on the ability of FLRT LRR to dimerize. The reported FLRT dimerization mutant (FLRT3-FF) decreased dimerization but did not completely abolish it (Figure VII.10B). When we mutated the same residues to alanine without introducing a glycosylation site (M11, R181A, D183A instead of R181N, D183T), FLRT3 LRR dimerization was not affected (Figure VII.10C), suggesting that the large glycosylation moiety introduced to the surface of the concave side of FLRT3 LRR in the FLRT-FF mutant is blocking FLRT3 dimerization due to steric hindrance rather than breaking specific molecular interactions. This result demonstrates that the introduction of carbohydrates to a site can block protein interactions dramatically. In addition, we tested our FLRT3 mutant that does not bind to LPHN3 (M02) for dimerization, and showed that its dimerization ability is not affected (Figure VII.10D). Thus, it can be used as a specific LPHN3-binding mutant in further experiments. The FLRT-UF mutant, on the other hand, showed wild-type like dimerization ability (Figure VII.10E).

FLRT3, LPHN3 and UNC5 form a trimeric complex

FLRT proteins are involved in heterodimeric interactions with LPHNs and with UNC5s, and in homodimeric interactions with themselves. However, whether all of these interactions are compatible is unclear. Next, we investigated whether

FLRT3/UNC5 interaction is compatible with FLRT3/ LPHN3 interaction, or in other words, whether FLRT3, LPHN3 and UNC5 can form a trimeric complex. The availability of the FLRT3/LPHN3 complex structure and the previously reported FLRT3/UNC5 complex structure (214) enabled us to compare structures and predict, then test, the compatibility of the possible interactions of FLRT3 with each other. Intriguingly, superimposition of the FLRT3/LPHN3 structure with the FLRT3/UNC5 structure suggests that UNC5 and LPHN3 bind to distinct surfaces on FLRT3 and that there are no clashes between UNC5 and LPHN3, suggesting that LPHN3 and UNC5 can bind to FLRT3 simultaneously (Figure VII.12A).

In order to test whether this model is correct, we expressed full-length UNC5D or full-length UNC5B in HEK293 cells and added premixed purified recombinant FLRT3 LRR, purified recombinant biotinylated LPHN3 Olf, and fluorescently labeled neutravidin (a protein that binds to biotin and tetramerizes it) onto the cells (see Figure VII.12B for a schematic representation of the experiment). Bound LPHN3 Olf was detected using flow cytometry. Our results showed that LPHN3 is detected on cells only when FLRT3, LPHN3 and neutravidin are all added, indicating the formation of a trimeric complex (Figure VII.12C,D and Figure VII.14). No LPHN3 binding was detected when FLRT3 was not added onto UNC5-expressing cells, suggesting that there is no direct interaction between UNC5 and LPHN3, instead, FLRT3 bridges the two proteins to form a trimeric complex. As mentioned above, due to avidity, by tetramerizing biotinylated LPHN3 Olf with neutravidin, the effective concentration of the FLRT3- LPHN3 affinity is more than 100-fold higher than the concentration used. Similar results were observed both with low or high protein concentrations with UNC5D. Flow cytometry experiments

performed with the purified FLRT3-UF LRR mutant showed that this mutant does not completely abolish UNC5 binding consistent with the previous experiments (Figure VII.12E and Figure VII.6A). These results suggest the formation of a trimeric complex between UNC5, FLRT and LPHN that forms in vitro.

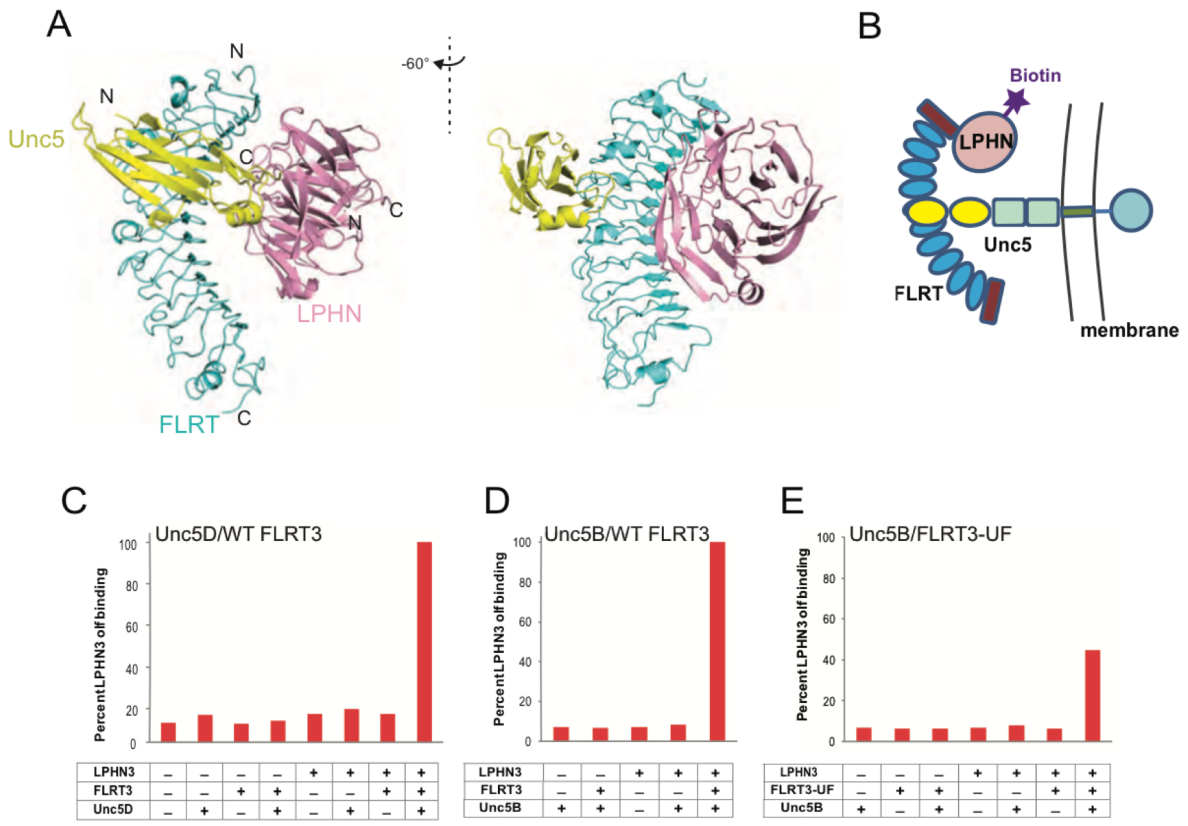


Figure VII.12 Formation of a trimeric complex between FLRT3, LPHN3, and UNC5 proteins.

(A) Superimposition of FLRT3/ LPHN3 and FLRT2/UNC5D (PDB ID 4V2C) structures. (B) Simple schematic diagram of the experimental design for the detection of the trimeric complex on HEK293 cells. (C) Quantification of purified biotinylated LPHN3 Olf binding to HEK293 cells transfected with FL-UNC5D. FLRT3 LRR domain (non-biotinylated), biotinylated LPHN3 Olf domain and neutravidin (allowing tetramerization) were mixed at different combinations at 5 nM or 100 nM concentrations and red fluorescence of DyLight650 attached to neutravidin was detected by flow cytometry. Pretetramerization of FLRT3 and LPHN3 with neutravidin increases avidity. (D) Same experiment as in C performed using UNC5B instead of UNC5D, using 100 nM pretetramerized complex. (E) Same experiment as in D using FLRT3-UF mutant. (Raw data in Figure VII.14.)

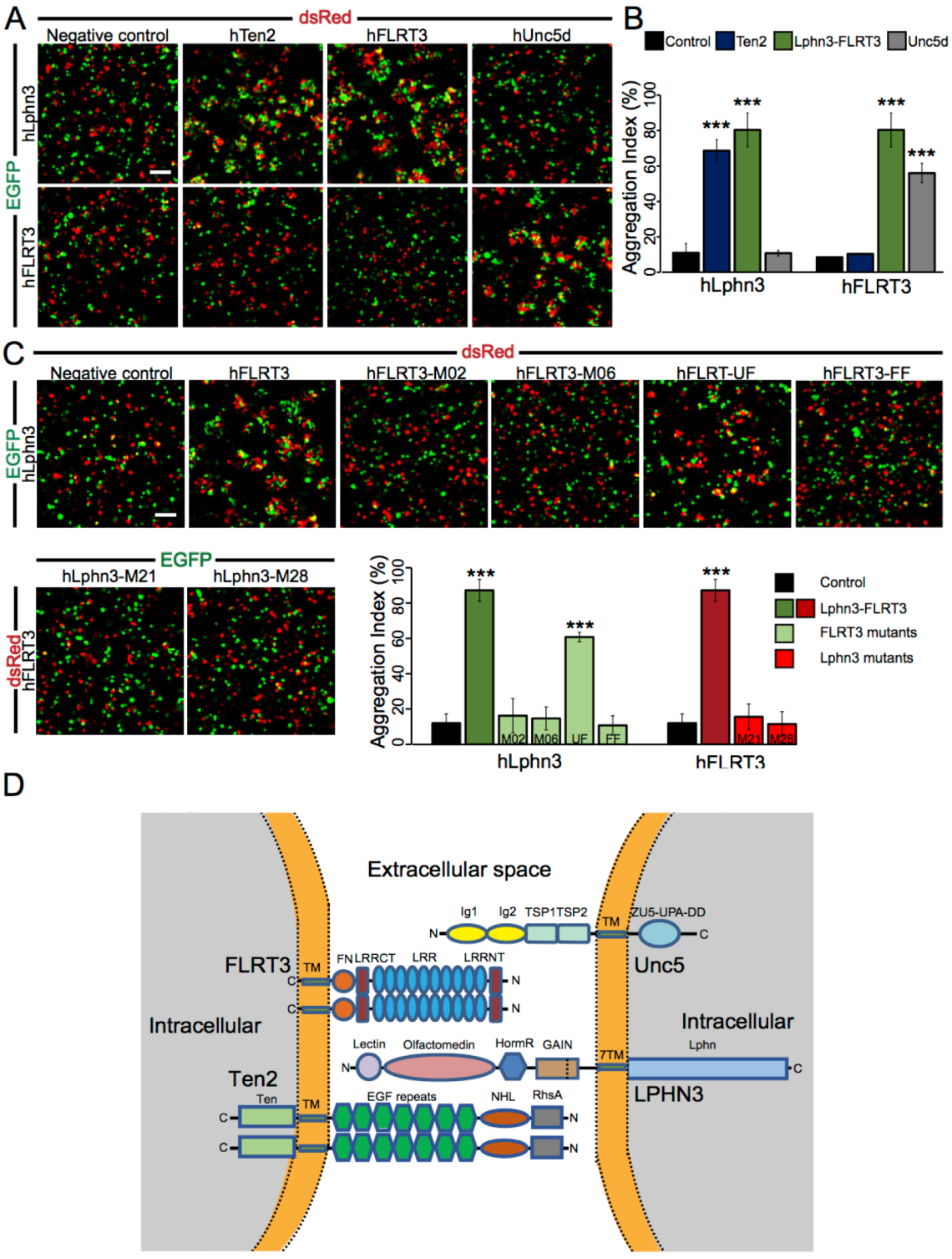


Figure VII.13 Analysis of LPHN3/FLRT3/UNC5d trans interactions.

(legend on the following page)

FLRT3/ LPHN3 and FLRT3/UNC5 mediate formation of intercellular contacts

Cell-adhesion molecules may function via two different ways: Two cell-adhesion molecules expressed on the same cell might be involved in cis-interactions or two cell-adhesion molecules each expressed on one of the two neighboring cells might be involved in trans-interactions. To examine if trans-interactions between LPHN3, FLRT3, or UNC5d can support cell-cell adhesion, we performed cell aggregation assays with non-adherent HEK293 cells in which each full-length protein is expressed on different cell populations and the cells are then mixed to monitor cell aggregation (Figure VII.13A-C). Previous studies have shown that LPHN1 binding to Teneurin2 in trans can promote cell aggregation (Boucard et al., 2014). Therefore, this condition was used as a positive control in cell aggregation experiments. Interestingly, we found that the binding of FL-LPHN3 to FL-FLRT3 and of FL-UNC5d to FL-FLRT3 both induce cell-cell adhesion in trans. Point mutations in both FL-LPHN3 and FL-FLRT3 in residues we

Figure VII.13, continued.

(A) Representative images from cell aggregation assays with wild-type full-length proteins. LPHN3 induces cell aggregation with FLRT3, as well as Ten2, a previously identified trans interaction partner (65). FLRT3 promotes trans-cellular adhesion with LPHN3 and UNC5d, but not with Ten2. (B) quantification of cell aggregation assays with wild-type proteins (means ± standard deviation). (C) Point mutations that disrupt LPHN3-FLRT3 binding abolish cell aggregation. Cell adhesion is preserved between LPHN3 and FLRT3 when the UNC5d-binding site on FLRT3 is mutated. All experiments were performed in three independent culture preparations (*** denotes $p<0.001$). (D) Diagram of LPHN3/FLRT3/UNC5d/Ten2 at a cellular junction. LPHN3 and UNC5d are localized to the opposing membrane which contains FLRT3 and Ten2. (Note: proteins are not drawn to scale). Adhesion will cause aggregates of two or more cells and therefore an increase in aggregation index. Aggregation index was calculated using the program ImageJ.

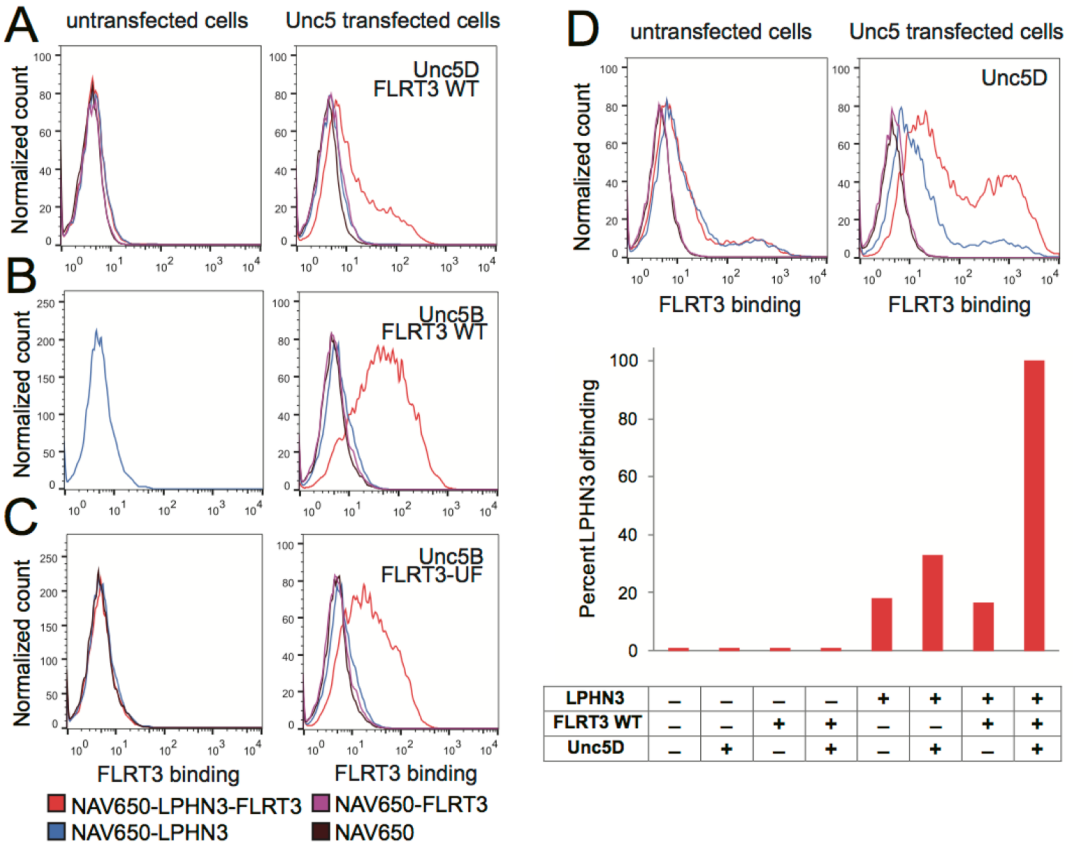


Figure VII.14 Formation of a trimeric complex between FLRT3, LPHN3, and UNC5 proteins.

(A to C) Untrasfected HEK293 cells or cells transfected with Unc5 were stained with a precomplex of biotinylated LPHN3 Olf domain, unlabeled His-FLRT3 LRR and neutravidin (NAV650) in different combinations to detect binding of LPHN3 Olf to cells with flow cytometry. Unc5D or Unc5B transfected HEK293 cells were tested for binding with 100 nM LPHN3 olf and wild type FLRT3, or LPHN3 Olf and FLRT-UF mutant precomplex. Panel A data was performed with 5 nM protein concentration with tetramerization, whereas panels B and C were performed with 100 nM protein concentration. Panel A was repeated with 100 nM protein concentration with tetramerization as shown in (D). Similar results were obtained but some non-specific binding is observed when 100 nM concentration was used.

found to be essential for complex formation in vitro abolish cell aggregation (Figure VII.13A-C). LPHN3, FLRT3, and UNC5d form a heterotrimeric complex, where LPHN3 and UNC5d bind to distinct binding sites on FLRT3. We observed no cell aggregates

mediated by homophilic interactions of any of these molecules, in particular FL-FLRT3 or FL-Ten2, suggesting that the hemophilic binding of these molecules operates in an exclusively cis configuration (Figure VII.13D). Altogether, these results support the notion that LPHN3, FLRT3, and UNC5d function in trans-cellular adhesion.

Discussion

Neural development is a complex phenomenon that is mediated by the coordinated interactions of numerous cell-surface proteins on neurons and glial cells. Studying binary protein interactions is a required but insufficient step to understand neural development as other interaction partners of the binary interaction in question likely affect the functional outcome of the specific protein-protein interaction. This outcome also depends on whether the cell-surface proteins are presented from two different cells making a trans-interaction or they are presented on the same cell making a cis interaction – in many cases, cis- and trans-interactions are mutually exclusive in that only one is possible when the respective proteins are embedded in the plasma membrane. Studying the basics of the specific protein-protein interactions and designing binding-mutants that specifically abolish one interaction without interfering with the other interactions of a protein are essential first steps before moving forward with functional understanding of each protein.

In this study, using a combination of biophysical, biochemical and cell-based approaches, we studied the structure and specificity of the interaction between FLRT3 and LPHN3, and also considered the interaction of FLRT3 with UNC5 and with itself. The structure of the FLRT3/LPHN3 complex revealed that the olfactomedin domain was

bound to the concave surface of the FLRT3 LRR horseshoe (Figure VII.1). The interaction surface spreads over the extensive complementary surfaces of both proteins. However, the major interaction hotspot is at the N-terminal top side of the LRR horseshoe and at the long loops emerging from the second and third blades of the LPHN3 olfactomedin beta-propeller (Figure VII.4). The concave surface of the FLRT3 LRR domain was previously reported to mediate FLRT3 dimerization, thus it is essential to study the effect of any mutations in this region on both LPHN3 binding and FLRT3 dimerization before commenting on the effect of the mutation on the function of the protein. This surface is spatially separated from the UNC5 binding site on FLRT3, and thus UNC5 binding and LPHN3 binding are likely not exclusive. On the other hand, the FLRT3 binding site on LPHN3 is likely distant from the teneurin binding site on LPHN3 as teneurin binding is largely mediated by the lectin domain and the splice insert, both of which are at the N-terminal side of the olfactomedin domain. Mapping of the mutations of LPHN3 genes that are linked to attention deficit/hyperactivity disorder and cancers on the FLRT3/LPHN3 complex structure reveals that most mutations are at the FLRT3/LPHN3 binding interface where the two proteins come closest to each other, and one mutation is at the edge of the binding interface, suggesting protein-protein interaction defects (Figure VII.4B,C).

Mapping the conserved and variable residues on the surface of the LPHN3 and FLRT3 structures and relating their location to interface between FLRT3 and LPHN3 in the FLRT3/ LPHN3 structure shows that the FLRT3 binding site on LPHN3 and the LPHN3 binding site on FLRT3 are highly conserved. This observation again suggests a critical role for the FLRT/LPHN interaction in neural development (Figure VII.1D).

Mapping the electrostatics surface potential on the FLRT/LPHN3 structure shows a large positive and a large negative surface at the concave inside surface of the LRR horseshoe, suggesting that non-specific charge-charge interactions might be mediated via this surface, especially at high protein concentrations (Figure VII.1E).

To confirm that the binding interface that is revealed by the FLRT/LPHN complex structure is biologically relevant, we designed FLRT3 mutants and LPHN3 mutants that specifically disrupt the FLRT/LPHN interaction without interfering with membrane localization (Table VII.4, Figure VII.5, and Figure VII.6). Flow cytometry binding experiments showed that mutations located on the interaction surface break the interaction. Flow cytometry experiments monitoring cell-surface expression of the mutants on non-permeabilized cells showed that some mutations affect the proper folding/trafficking of the proteins (Figure VII.7 and Figure VII.8). Thus, only well-expressed and trafficked mutants were used in further experiments. All selected FLRT3 mutants were tested for their ability to dimerize to make sure only one specific interaction is broken in any given mutant (Figure VII.10). Thus, we generated well-characterized FLRT3 mutants and LPHN3 mutants that can be used as molecular tools in further experiments to specifically dissect the mechanism of these multi-interaction proteins.

Importantly, we studied the previously reported “FLRT3 dimerization mutant-FF” (214) and showed that the LPHN3-binding ability of this mutant is more severely affected than its dimerization ability (Figure VII.6A and Figure VII.10B). Considering that previously performed in vivo studies reported the FLRT-FF mutant impairs tangential (sideways) migration of neurons during cortex development, we suggest that the lack of

LPHN3 binding to the FLRT3-FF mutant might be responsible for this defect. The high-conservation of the binding interfaces and the localization of ADHD and cancer mutations at the FLRT/LPHN binding surface also suggest a critical role for this interaction. Similar considerations might apply to the mutants studied in another recent paper (216) as all these mutations include introduction of a new N-linked glycosylation site into the protein sequence.

The availability of the FLRT3/LPHN3 complex structure allowed us to make comparisons with the FLRT2/UNC5D complex structure and predict that LPHN3 and UNC5D can simultaneously interact with FLRT (Figure VII.12A). Our binding experiments showed that FLRT, LPHN and UNC5 form a trimeric complex and FLRT binds the other two proteins simultaneously and bridges them where LPHN and UNC5 do not directly interact with each other (Figure VII.12). Our cell-adhesion assays showed that LPHN3/ FLRT3 binding induces trans-cellular adhesion. Mutations in the binding interface of LPHN3 and FLRT3 abrogated cell adhesion. Interestingly, while FLRT3 / UNC5d binding also induces cell aggregation, LPHN3 /UNC5d binding does not. Altogether, these results suggest that LPHN3 and UNC5d are localized to one side of the cellular junction in which they participate, whereas FLRT3 is localized to the other side (Figure VII.13D). At least LPHN3 and Flrt3 have been localized to synapses in mature brain, suggesting that the FLRT3/LPHN3 and the FLRT3/UNC5d interaction may, among others, contribute to synapse formation and/or synaptic transmission. However, the precise pre- vs. post-synaptic localization of these proteins, as well as if they are necessary and/or sufficient for synaptic adhesion, remains to be elucidated – for none of these proteins has it actually been shown directly whether they are pre- or

postsynaptic. While FLRT3 appears to form a homodimer, we were unable to observe FLRT3-FLRT3 cell aggregation. Thus, this homodimer likely occurs in cis and is incapable of supporting trans-cellular adhesion. Interestingly, the concurrent trans-cellular interaction of FLRT3 with both UNC5d and LPHN3, and the likely simultaneous trans-cellular interaction of LPHN3 in turn with both FLRT3 and teneurins, creates an interaction network in which the cells are linked at an intercellular junction such as the synapse not by a simple one-to-one complex, but by a large complex composed of very different simultaneous interactions that may be independently regulated, and may transduce distinct trans-cellular signals.

In conclusion, the available high-resolution structure of the FLRT3/LPHN3 complex provides the basis for further advances in understanding their mechanism of action in brain function which may lead to the treatment of diseases that are caused by mutations in these proteins.

CHAPTER VIII

Conclusions and Future Directions

Targeting GPR56 with drug-like molecules

The majority of the work presented in this thesis focuses on the structural, functional, and mechanistic characterization of GPR56. Though we are still years away from GPR56-targeted therapeutics in the clinic, we are far closer than we were at the start of this work in 2013. Our panel of GPR56-targeted monobodies has demonstrated the principal that synthetic GPR56 ligands are capable of tuning receptor activity. Furthermore, that these monobodies bind to the ECR and not the 7TM, the conventional site of GPCR ligands. This result, in combination with the observed inhibitory role of the PLL domain, directly implicates the ECR as a prime regulator of GPR56 function and suggests that ECR-targeted therapeutics may be clinically desirable.

In fact, targeting the ECR may be more desirable than targeting the 7TM due to the high conservation of the 7TM. Because the GPCR 7TM is so conserved, obtaining highly specific 7TM-targeted GPCR ligands can be difficult. Drugs with poor specificity may lead to undesirable side-effects. On the other hand, aGPCR ECRs are very diverse, such that obtaining an ECR-targeted molecule with high specificity may be relatively straightforward. Overall, our results suggest that efforts to obtain ECR-targeted aGPCR ligands, such as our monobodies, may result in therapeutically desirable molecules.

With the goal of targeting GPR56, perhaps the optimal context involves its recently characterized role in acute myelogenous leukemia (AML) (49, 84), wherein GPR56 expression has been shown to decrease survival, reduce efficacy of

conventional chemotherapies, and promote cancer growth. As such, a GPR56 inhibitor is explicitly sought out by groups working on AML. Though all of the functional monobodies exhibit high potency, they also exhibit rather weak efficacy. However, in a case like AML, it is unclear what decree of efficacy is required to achieve the desired response. Testing inhibitory monobodies like β1 may not only assist in more targeted drug development of GPR56 inhibitors, but may also shed light on the biological relevance of modulating aGPCRs with potent ligands with relatively low efficacy.

aGPCRs in the extracellular milieu

The meshwork of the extracellular matrix is overwhelmingly complex, especially given its composition of non-soluble proteins, such as collagen. As cells consistently remodel the ECM and migrate through it, its structure and composition changes over time. It is critical to picture the extracellular regions of cell surface receptors interacting with their ligands and transducing signals in this crowded and dynamic environment. It is also important to keep in mind that any given cell may express hundreds or thousands of cell surface receptors, some of which may interact with the same ligand, or signal through the same downstream signaling pathway. Receptors may even interact with other receptors, resulting in modulating of signaling. For example, just as with the canonical dimerization-mediated activation mechanism of enzyme-linked receptors, some GPCRs are reported to dimerize (221, 222). Additionally, some GPCRs are reported to interact directly with RTKs, causing altered signaling (223). Taken together, the proteins that comprise the ECM participate in a web of overlapping and

complex functions, some of which are undoubtedly impossible to study in the reductionist environments of a test tube, or even a tissue culture dish. Given the severe technical complications associated with performing in-depth biochemical and biophysical measurements in native ECM, we must extrapolate data obtained *in vitro* to the native environment to the best of our ability.

GPR56 is reportedly activated by collagen III (75, 86). Given the difficulty of studying non-soluble proteins such as collagen III with standard biochemical techniques, even *in vitro*, characterization of this important interaction remains problematic. Many other aGPCR ligands are similarly problematic. Until robust biochemical and pharmacological assays in native ECM are developed, it is important to utilize existing technologies to characterize complementary aspects of aGPCR function. For example, we may study the contribution of mechanical force to aGPCR activation, like the force that may be transduced by migrating cells in native ECM. We speculate that mechanical force may induce NTF dissociation and *Stachel*-exposure (i.e. shedding), hence the development of biochemical and biophysical assays to quantitatively measure shedding will be impactful. Upon exposure of the *Stachel*, the field still questions the structural and allosteric implications of *Stachel*-mediated activation of the 7TM. The application of structural biology techniques may shed light on this vital aspect of aGPCR activation. In a similar vein, these experiments may provide insight into the conformational dynamics of a full-length aGPCR, including the identification of any putative interactions between extracellular domains and the 7TM that may govern ECR-mediated regulation of basal activity that has been observed.

It will also prove useful to apply techniques developed over the past five decades to study canonical GPCR function to aGPCRs, including screening small molecule libraries for functional compounds and characterizing hallmark GPCR biological phenomena such as β-arrestin recruitment, endocytosis, recycling, and degradation. The idea of aGPCRs participating in biased signaling is also very compelling, as their ECRs may serve as large scaffolds, interacting with diverse ligands, and regulating downstream signaling in complex ways. Given that GPCRs without large ECRs exhibit biased signaling, the additional regulatory capacity of the ECR may indeed mediate even more complicated modes of biased signaling for aGPCRs. Excitingly, over the past two years, several of these techniques have been successfully attempted for some of the more well-studied aGPCRs, like GPR56 (89, 91). Over the upcoming decade, exciting results will stem from a more thorough characterization of many more aGPCRs, aided by these tried-and-true techniques as well those still under development.

The mechanics of GAIN domain shedding and *Stachel*-exposure

Though it is widely hypothesized that GAIN domain shedding may be induced by mechanical force exerted on the NTF, such experiments have not been carried out. However, several feasible experimental approaches to directly measure force-induced GAIN domain shedding or even force-induced aGPCR activation are technically possible. We have already demonstrated the feasibility of purifying large quantities of aGPCR GAIN domains and full ECRs from insect cells (55, 68, 93). In this recombinant system, we can add N- and C-terminal epitope tags and even introduce point mutations that inhibit autoproteolysis. Single molecule force spectroscopy experiments (224) with

such constructs may answer some fundamental questions about GAIN domain architecture and shedding. For example, it is critical to understand if the amount force required to remove the *Stachel* is of a biologically relevant order of magnitude. Additionally, by comparing the force observed for the full ECR to that observed for the GAIN domain alone, we may learn if interactions between N-terminal adhesion domains and the GAIN domain, including interdomain disulfide bonds, influence shedding propensity, a phenomenon that has been hypothesized (93) but never proven. Finally, by comparing the wild-type ECR with an autoproteolysis-deficient mutant, we may learn if global GAIN domain unfolding is required for shedding. This importance of this experiment is underscored by the observation that some aGPCRs, including GPR56, undergo tissue-depednat autoproteolysis, such that some receptors remain partially or completely uncleaved in different tissues (57). Though this has been observed, its underlying biological relevance remains unclear.

It is conceivable that hydrogen exchange mass spectrometry (HX MS) (225) could also be used to more accurately understand the conformational space sampled by the GAIN domain, both before and after shedding. By applying this technique to wild-type (and autoproteolysis-deficient) purified GAIN domains, we would get a sense of the solvent accessibility of the *Stachel* prior to shedding (or in the absence of shedding). As it has been proposed that, even in the absence of autoproteolysis, the *Stachel* may transiently escape the GAIN domain to interact with the 7TM (35, 147), an HX MS experiment would elucidate if this proposed model is viable.

Experimental setups have been developed that allow force to be transduced to cells cultured in 96-well plates (226). In short, cultured cells are incubated with magnetic

beads coated with a ligand for the receptor if interest. A 96-well magnet array is then placed onto the plate, such that a single magnetic post is in contact with each well. By adjusting the distance between the magnets and the cells, the force can be titrated. Using such devices, the effect of mechanical force on signaling has been measured for several cell surface receptors including integrins (227, 228) and Notch (229). In addition to the biophysical measurements obtained using single molecule force spectroscopy and HX MS as described above, complementary experiments using one of these devices, for example, to measure the effect of mechanical force on aGPCR signaling will be a critical to support this hypothesis of aGPCR activation.

Structure of the aGPCR 7TM and its interactions with the *Stachel* and the NTF

Recent advances in cryo-electron microscopy and x-ray crystallography have enabled atomic-resolution structural studies of many non-adhesion GPCRs (24, 37, 38, 122, 230, 231). Among these are structures of the GLP-1 receptor, a secretin family GPCR, bound to its agonist peptide. Sucessful studies like these suggest that atomic-resolution aGPCR 7TM structures will be solved in the near future. Of course, it is critical to visualize the allosteric changes within the aGPCR 7TM upon activation, as it is for all GPCRs. Also, like the GLP-1 structure, any structure illustrating *Stachel*-7TM interaction and activation will be highly impactful. One of the more interesting aspects of aGPCR structural studies, however, will be the visualization of any putative NTF-7TM interactions. Though such interactions have been hypothesized as a mechanism behind *Stachel*-independent regulation of signaling (91, 93, 232), they have never been proven. Information from these structural studies could be used to target drug-like molecules

(e.g. monobodies) to perturb interactions at specific locations within the aGPCR, and thereby elucidate finer functional details.

Conclusion

aGPCRs are a biochemically fascinating class of cell-surface receptors, including an autoproteolytic GAIN domain, some or many N-terminal cell adhesion domains, and the classical GPCR 7TM. The biochemical and biophysical characterization of GPR56 presented in this thesis is important, but must be complemented with functional studies *in vivo*. With the ultimate goal of treating diseases mediated by aGPCRs, interdisciplinary collaborations are absolutely critical. aGPCRs are modular and complicated proteins; they are difficult to work with even for the most advanced protein biochemistry laboratories. Perhaps non-coincidentally, their functions are also multifaceted and complex, transducing cues from the ECM and neighboring cells into directions for proper cell migration, proliferation, and differentiation. Assays designed to characterize these phenomena are even more complex than those used to characterize the proteins. In the case of aGPCRs, specialists are required at every level, from foundational biochemistry, to pathophysiology, to pharmacology and medicinal chemistry. Let the work described herein contribute to the structural and biochemical foundation for future studies that ultimately result in the development many successful aGPCR-targeted therapeutics.

BIBLIOGRAPHY

1. Mahrle G, Orfanos CE (1977) the Plasma Unit Membrane. *Br J Dermatol* 96(2):215–223.

2. Finer-Moore J, Stroud RM (1984) Amphipathic analysis and possible formation of the ion channel in an acetycholine receptor. *Procedings Nat Acad Sci USA* 81(January):155–159.

3. Schneider WJ, Nimpf J (2003) LDL receptor relatives at the crossroad of endocytosis and signaling. *Cell Mol Life Sci* 60(5):892–903.

4. Hatta K, Takeichi M (1986) Expression of N-cadherin adhesion molecules associated with early morphogenetic events in chick development. *Nature* 320(6061):447–449.

5. Wheelock MJ, Johnson KR (2003) Cadherin-mediated cellular signaling. *Curr Opin Cell Biol* 15(5):509–514.

6. Alexander S, Mathie A, Peters J (2011) Guide to Receptors and Channels, 5th edition. 158(November):2008.

7. Hynes NE, MacDonald G (2009) ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol* 21(2):177–184.

8. Ihle JN, Kerr IM (1995) Jaks and Stats in signaling by the cytokine receptor superfamily. *Trends Genet* 11(2):69–74.

9. Rawlings JS (2004) The JAK/STAT signaling pathway. *J Cell Sci* 117(8):1281–1283.

10. Kawai T, Akira S (2007) TLR signaling. *Semin Immunol* 19(1):24–32.

11. Miggin SM, O'Neill L a J (2006) New insights into the regulation of TLR signaling. *J Leukoc Biol* 80(2):220–226.

12. Fesnak AD, June CH, Levine BL (2016) Engineered T cells: the promise and challenges of cancer immunotherapy. *Nat Rev Cancer* 16(9):566–581.

13. Brownlie RJ, Zamoyska R (2013) T cell receptor signalling networks: branched, diversified and bounded. *Nat Rev Immunol* 13(4):257–269.

14. Arnaout MA, Mahalingam B, Xiong J-P (2005) Integrin Structure, Allostery, and Bidirectional Signaling. *Annu Rev Cell Dev Biol* 21(1):381–410.

15. Luo B-H, Carman C V., Springer TA (2007) Structural Basis of Integrin Regulation

and Signaling. *Annu Rev Immunol* 25(1):619–647.

16. Larsen M, Artym V V., Green JA, Yamada KM (2006) The matrix reorganized: extracellular matrix remodeling and integrin signaling. *Curr Opin Cell Biol* 18(5):463–471.

17. Yap AS, Kovacs EM (2003) Direct cadherin-activated cell signaling: A view from the plasma membrane. *J Cell Biol* 160(1):11–16.

18. Brieher WM, Yap AS, Gumbiner BM (1996) Lateral dimerization is required for the homophilic binding activity of C-cadherin. *J Cell Biol* 135(2):487–496.

19. Kamiguchi H, Lemmon V (2000) IgCAMs: Bidirectional signals underlying neurite growth. *Curr Opin Cell Biol* 12(5):598–605.

20. Van Buul JD, Kanters E, Hordijk PL (2007) Endothelial signaling by Ig-like cell adhesion molecules. *Arterioscler Thromb Vasc Biol* 27(9):1870–1876.

21. McEver RP (2015) Selectins: Initiators of leucocyte adhesion and signalling at the vascular wall. *Cardiovasc Res* 107(3):331–339.

22. Fredriksson R (2003) The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* 63(6):1256–1272.

23. Stevens RC, et al. (2012) The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nat Rev Drug Discov* 12(1):25–34.

24. Rasmussen SGF, et al. (2011) Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* 477(7366):549–555.

25. Hoffman BB, Lefkowitz RJ (1982) Adrenergic receptors in the heart. *Annu Rev Physiol* 44(1):475–484.

26. Samama P, Pei G, Costa T, Cotecchia S, Lefkowitz RJ (1994) Negative antagonists promote an inactive conformation of the beta 2-adrenergic receptor. *Mol Pharmacol* 45(3):390–394.

27. Wettschureck N, Offermanns S (2005) Mammalian G Proteins and Their Cell Type Specific Functions. *Physilological Rev*:1159–1204.

28. Oldham WM, Hamm HE (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* 9(1):60–71.

29. Rankovic Z, Brust TF, Bohn LM (2016) Biased agonism: An emerging paradigm in GPCR drug discovery. *Bioorganic Med Chem Lett* 26(2):241–250.

30. Wootten D, Miller LJ, Koole C, Christopoulos A, Sexton PM (2017) Allostery and Biased Agonism at Class B G Protein-Coupled Receptors. *Chem Rev* 117(1):111–138.

31. Eglen RM, Reisine T (2011) GPCRs revisited: New insights lead to novel drugs. *Pharmaceuticals* 4(2):244–272.

32. Wootten D, Christopoulos A, Sexton PM (2013) Emerging paradigms in GPCR allostery: implications for drug discovery. *Nat Rev Drug Discov* 12(8):630–644.

33. Garland SL (2013) Are GPCRs Still a Source of New Targets? *J Biomol Screen* 18(9):947–966.

34. Christopoulos A, et al. (2014) International Union of Basic and Clinical Pharmacology. XC. Multisite Pharmacology: Recommendations for the Nomenclature of Receptor Allosterism and Allosteric Ligands. *Pharmacol Rev* 66(4):918–947.

35. Stoveken HM, Hajduczok AG, Xu L, Tall GG (2015) Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proc Natl Acad Sci* 112(19):6194–6199.

36. Cheng Z, et al. (2010) Luciferase Reporter Assay System for Deciphering GPCR Pathways. *Curr Chem Genomics* 4:84–91.

37. Jazayeri A, et al. (2017) Crystal structure of the GLP-1 receptor bound to a peptide agonist. *Nature* 546(7657):254–258.

38. Zhang Y, et al. (2017) Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* 546(7657):248–253.

39. Zhang H, et al. (2017) Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* 546(7657):1–18.

40. Yang L, et al. (2015) Conformational states of the full-length glucagon receptor. *Nat Commun* 6:7859.

41. Koth CM, et al. (2012) Molecular basis for negative regulation of the glucagon receptor. *Proc Natl Acad Sci* 109(36):14393–14398.

42. Byrne EFX, et al. (2016) Structural basis of Smoothened regulation by its extracellular domains. *Nature* 535(7613):517–522.

43. Zhao L-H, et al. (2016) Differential Requirement of the Extracellular Domain in Activation of Class B G Protein-coupled Receptors. *J Biol Chem* 291(29):15119–15130.

44. Booe JM, et al. (2015) Structural Basis for Receptor Activity-Modifying Protein-Dependent Selective Peptide Recognition by a G Protein-Coupled Receptor. *Mol Cell* 58(6):1040–1052.

45. Langenhan T, Aust G, Hamann J (2013) Sticky Signaling--Adhesion Class G Protein-Coupled Receptors Take the Stage. *Sci Signal* 6(276):re3-re3.

46. Paavola KJ, Hall RA (2012) Adhesion G Protein-Coupled Receptors: Signaling, Pharmacology, and Mechanisms of Activation. *Mol Pharmacol* 82(5):777–783.

47. Liebscher I, Schöneberg T, Prömel S (2013) Progress in demystification of adhesion G protein-coupled receptors. *Biol Chem* 394(8):937–950.

48. Jin G, et al. (2017) The G-protein coupled receptor 56, expressed in colonic stem and cancer cells, binds progastrin to promote proliferation and carcinogenesis. *Oncotarget*:1–14.

49. Pabst C, et al. (2016) GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. *Blood* 127(16):2018–2027.

50. Faiz A, et al. (2017) Latrophilin receptors: novel bronchodilator targets in asthma. *Thorax* 72(1):74–82.

51. Balenga N, et al. (2017) Orphan Adhesion GPCR GPR64/ADGRG2 Is Overexpressed in Parathyroid Tumors and Attenuates Calcium-Sensing Receptor-Mediated Signaling. *J Bone Miner Res* 32(3):654–666.

52. Hardin M, et al. (2017) Sex-Based Genetic Association Study Identifies *CELSR1* as a Possible Chronic Obstructive Pulmonary Disease Risk Locus among Women. *Am J Respir Cell Mol Biol* 56(3):332–341.

53. Bayin NS, et al. (2016) GPR133 (ADGRD1), an adhesion G-protein-coupled receptor, is necessary for glioblastoma growth. *Oncogenesis* 5(10):e263.

54. Ribasés M, et al. (2011) Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: A replication study. *Genes, Brain Behav* 10(2):149–157.

55. Araç D, et al. (2012) A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO J* 31(6):1364–1378.

56. Prömel S, et al. (2012) Characterization and functional study of a cluster of four highly conserved orphan adhesion-GPCR in mouse. *Dev Dyn* 241(10):1591–1602.

57. Iguchi T, et al. (2008) Orphan G Protein-coupled Receptor GPR56 Regulates Neural Progenitor Cell Migration via a G 12/13 and Rho Pathway. *J Biol Chem* 283(21):14469–14478.

58. Promel S, et al. (2012) The GPS Motif Is a Molecular Switch for Bimodal Activities of Adhesion Class G Protein-Coupled Receptors. *Cell Rep* 2(2):321–331.

59. Lelianova VG, et al. (1997) α-Latrotoxin receptor, latrophilin, is a novel member of the secretin family of G protein-coupled receptors. *J Biol Chem* 272(34):21504–21508.

60. Silva J-P, et al. (2011) Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc Natl Acad Sci* 108(29):12113–12118.

61. O'Sullivan ML, et al. (2012) FLRT Proteins Are Endogenous Latrophilin Ligands and Regulate Excitatory Synapse Development. *Neuron* 73(5):903–910.

62. Boucard AA, Ko J, Südhof TC (2012) High affinity neurexin binding to cell adhesion G-protein-coupled receptor CIRL1/latrophilin-1 produces an intercellular adhesion complex. *J Biol Chem* 287(12):9399–9413.

63. Young TR, Leamey CA (2009) Teneurins: Important regulators of neural circuitry. *Int J Biochem Cell Biol* 41(5):990–993.

64. Tucker RP, Kenzelmann D, Trzebiatowska A, Chiquet-Ehrismann R (2007) Teneurins: Transmembrane proteins with fundamental roles in development. *Int J Biochem Cell Biol* 39(2):292–297.

65. Boucard AA, Maxeiner S, Südhof TC (2014) Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: Regulation by alternative splicing. *J Biol Chem* 289(1):387–402.

66. Woelfle R, D'Aquila AL, Pavlović T, Husić M, Lovejoy DA (2015) Ancient interaction between the teneurin C-terminal associated peptides (TCAP) and latrophilin ligand-receptor coupling: A role in behavior. *Front Neurosci* 9(APR):1–10.

67. Yamagishi S, et al. (2011) FLRT2 and FLRT3 act as repulsive guidance cues for Unc5-positive neurons. *EMBO J* 30(14):2920–2933.

68. Lu YC, et al. (2015) Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion. *Structure* 23(9):1678–1691.

69. Liu M, et al. (1999) GPR56, a novel secretin-like human G-protein-coupled receptor gene. *Genomics* 55(3):296–305.

70. Piao X (2004) G Protein-Coupled Receptor-Dependent Development of Human Frontal Cortex. *Science (80-)* 303(5666):2033–2036.

71. Barkovich AJ, Hevner R, Guerrini R (1999) Syndromes of bilateral symmetrical

polymicrogyria. *Am J Neuroradiol* 20(10):1814–1821.

72. Jeong SJ, et al. (2013) GPR56 Functions Together with α3β1 Integrin in Regulating Cerebral Cortical Development. *PLoS One* 8(7):e68781.

73. Li S, et al. (2008) GPR56 Regulates Pial Basement Membrane Integrity and Cortical Lamination. *J Neurosci* 28(22):5817–5826.

74. Koirala S, Jin Z, Piao X, Corfas G (2009) GPR56-Regulated Granule Cell Adhesion Is Essential for Rostral Cerebellar Development. *J Neurosci* 29(23):7439–7449.

75. Luo R, et al. (2011) G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc Natl Acad Sci* 108(31):12925–12930.

76. Bae B-I, et al. (2014) Evolutionarily Dynamic Alternative Splicing of GPR56 Regulates Regional Cerebral Cortical Patterning. *Science (80-)* 343(6172):764–768.

77. Giera S, et al. (2015) The adhesion G protein-coupled receptor GPR56 is a cell-autonomous regulator of oligodendrocyte development. *Nat Commun* 6(May 2014):6121.

78. Ackerman SD, Garcia C, Piao X, Gutmann DH, Monk KR (2015) The adhesion GPCR Gpr56 regulates oligodendrocyte development via interactions with Gα12/13 and RhoA. *Nat Commun* 6(May 2014):6122.

79. Wu MP, et al. (2013) G-protein coupled receptor 56 promotes myoblast fusion through serum response factor- and nuclear factor of activated T-cell-mediated signalling but is not essential for muscle development in vivo. *FEBS J* 280(23):6097–6113.

80. Peng Y-M, et al. (2011) Specific expression of GPR56 by human cytotoxic lymphocytes. *J Leukoc Biol* 90(4):735–740.

81. Chang G-W, et al. (2016) The Adhesion G Protein-Coupled Receptor GPR56/ADGRG1 Is an Inhibitory Receptor on Human NK Cells. *Cell Rep* 15(8):1757–1770.

82. Yang L, Friedland S, Corson N, Xu L (2014) GPR56 inhibits melanoma growth by internalizing and degrading its ligand TG2. *Cancer Res* 74(4):1022–1031.

83. Chiang NY, et al. (2017) GPR56/ADGRG1 Activation Promotes Melanoma Cell Migration via NTF Dissociation and CTF-Mediated G??12/13/RhoA Signaling. *J Invest Dermatol* 137(3):727–736.

84. Saito Y, et al. (2013) Maintenance of the hematopoietic stem cell pool in bone marrow niches by EVI1-regulated GPR56. *Leukemia* 27(8):1637–1649.

85. Tritarelli A, et al. (2004) p53 Localization at Centrosomes during Mitosis and Postmitotic Checkpoint Are ATM-dependent and Require Serine 15 Phosphorylation. *Mol Biol Cell* 15(April):3751–3737.

86. Luo R, et al. (2014) Mechanism for adhesion G protein-coupled receptor GPR56-mediated RhoA activation induced by collagen III stimulation. *PLoS One* 9(6):e100043.

87. Xu L, Begum S, Hearn JD, Hynes RO (2006) GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. *Proc Natl Acad Sci U S A* 103(24):9023–9028.

88. Chiang N-Y, et al. (2016) Heparin interacts with the adhesion GPCR GPR56, reduces receptor shedding, and promotes cell adhesion and motility. *J Cell Sci* 129(11):2156–2169.

89. Stoveken HM, et al. (2016) Dihydromunduletone Is a Small-Molecule Selective Adhesion G Protein-Coupled Receptor Antagonist. *Mol Pharmacol* 90(3):214–224.

90. Paavola KJ, Stephenson JR, Ritter SL, Alter SP, Hall RA (2011) The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. *J Biol Chem* 286(33):28914–28921.

91. Kishore A, Purcell RH, Nassiri-Toosi Z, Hall RA (2016) Stalk-dependent and stalk-independent signaling by the adhesion G protein-coupled receptors GPR56 (ADGRG1) and BAI1 (ADGRB1). *J Biol Chem* 291(7):3385–3394.

92. Ohta S, et al. (2015) Agonistic antibodies reveal the function of GPR56 in human glioma U87-MG cells. *Biol Pharm Bull* 38(4):594–600.

93. Salzman GS, et al. (2016) Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. *Neuron* 91(6):1292–1304.

94. Birtalan S, et al. (2008) The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *J Mol Biol* 377(5):1518–1528.

95. Fellouse FA, Barthelemy PA, Kelley RF, Sidhu SS (2006) Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J Mol Biol* 357(1):100–114.

96. Fellouse FA, et al. (2007) High-throughput Generation of Synthetic Antibodies

from Highly Functional Minimalist Phage-displayed Libraries. *J Mol Biol* 373(4):924–940.

97. Fellouse FA, et al. (2005) Molecular recognition by a binary code. *J Mol Biol* 348(5):1153–1162.

98. Koide S, Sidhu SS (2009) The importance of being tyrosine: Lessons in molecular recognition from minimalist synthetic binding proteins. *ACS Chem Biol* 4(5):325–334.

99. Gilbreth RN, Koide S (2012) Structural insights for engineering binding proteins based on non-antibody scaffolds. *Curr Opin Struct Biol* 22(4):413–420.

100. Nord K, et al. (1996) Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nat Biotechnol* 15(8):772–777.

101. Sandström K, Xu Z, Forsberg G, Nygren P (2003) Inhibition of the CD28-CD80 co-stimulation signal by a CD28-binding affibody ligand developed by combinatorial protein engineering. *Protein Eng* 16(9):691–697.

102. Wikman M, et al. (2004) Selection and characterization of HER2/neu-binding affibody ligands. *Protein Eng Des Sel* 17(5):455–462.

103. Nord K, Nilsson J, Nilsson B, Uhlén M, Nygren PÅ (1995) A combinatorial library of an α-helical bacterial receptor domain. *Protein Eng Des Sel* 8(6):601–608.

104. Schlehuber S, Beste G, Skerra A (2000) A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin11Edited by R. Huber. *J Mol Biol* 297(5):1105–1120.

105. Skerra A (2000) Lipocalins as a scaffold. *Biochim Biophys Acta - Protein Struct Mol Enzymol* 1482(1–2):337–350.

106. Mercader J V., Skerra A (2002) Generation of anticalins with specificity for a nonsymmetric phthalic acid ester. *Anal Biochem* 308(2):269–277.

107. Beste G, Schmidt FS, Stibora T, Skerra a (1999) Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc Natl Acad Sci U S A* 96(5):1898–1903.

108. Koide A, Bailey CW, Huang X, Koide S (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol* 284(4):1141–1151.

109. Gilbreth RN, et al. (2014) Stabilization of the third fibronectin type III domain of human tenascin-C through minimal mutation and rational design. *Protein Engineering, Design and Selection*, pp 411–418.

110. Diem MD, et al. (2014) Selection of high-affinity Centyrin FN3 domains from a simple library diversified at a combination of strand and loop positions. *Protein Engineering, Design and Selection*, pp 419–429.

111. Forrer P, Stumpp MT, Binz HK, Plückthun A (2003) A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Letters*, pp 2–6.

112. Kohl A, et al. (2003) Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proc Natl Acad Sci* 100(4):1700–1705.

113. Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A (2003) Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol* 332(2):489–503.

114. Koide A, Wojcik J, Gilbreth RN, Hoey RJ, Koide S (2012) Teaching an old scaffold new tricks: Monobodies constructed using alternative surfaces of the FN3 scaffold. *J Mol Biol* 415(2):393–405.

115. Wojcik J, et al. (2010) A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol* 17(4):519–527.

116. Wojcik J, et al. (2016) Allosteric inhibition of Bcr-Abl kinase by high affinity monobody inhibitors directed to the Src homology 2 (SH2)-kinase interface. *J Biol Chem* 291(16):8836–8847.

117. Koide S (2009) Engineering of recombinant crystallization chaperones. *Curr Opin Struct Biol* 19(4):449–457.

118. Hunte C, Michel H (2002) Crystallisation of membrane proteins mediated by antibody fragments. *Curr Opin Struct Biol* 12(4):503–508.

119. Wilson IA, Stanfield RL (1994) Antibody-antigen interactions: New structures and new conformational changes. *Curr Opin Struct Biol* 4(6):857–867.

120. Nady N, et al. (2015) ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* 4(NOVEMBER2015):1–28.

121. Stockbridge RB, et al. (2015) Crystal structures of a double-barrelled fluoride ion channel. *Nature* 525(7570):548–551.

122. Rasmussen SGF, et al. (2011) Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* 469(7329):175–180.

123. Sha F, et al. (2013) Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. *Proc Natl Acad Sci*

110(37):14924–14929.

124. Oganesyan N, Kim R, Kim S-H (2004) On-column Chemical Refolding of Proteins. *PharmaGenomics* (September). Available at: http://www.lbl.gov/tt/publications/2004pubPharmagen.pdf.

125. Miyatake H, Hasegawa T, Yamano A (2006) New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL). *Acta Crystallogr Sect D Biol Crystallogr* 62(3):280–289.

126. Ness SR, De Graaff RAG, Abrahams JP, Pannu NS (2004) Crank: New methods for automated macromolecular crystal structure solution. *Structure* 12(10):1753–1761.

127. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* 203(3):253–310.

128. Thisse C, Thisse B (2008) High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* 3(1):59–69.

129. Lyons DA, et al. (2005) erbb3 and erbb2 are essential for Schwann cell migration and myelination in zebrafish. *Curr Biol* 15(6):513–524.

130. Petersen SC, et al. (2015) The Adhesion GPCR GPR126 Has Distinct, Domain-Dependent Functions in Schwann Cell Development Mediated by Interaction with Laminin-211. *Neuron* 85(4):755–769.

131. Tsumoto K, et al. (2004) Role of Arginine in Protein Refolding, Solubilization, and Purification. *Biotechnol Prog* 20(5):1301–1308.

132. De Genst E, et al. (2006) Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci* 103(12):4586–4591.

133. Gilbreth RN, Esaki K, Koide A, Sidhu SS, Koide S (2008) A Dominant Conformational Role for Amino Acid Diversity in Minimalist Protein-Protein Interfaces. *J Mol Biol* 381(2):407–418.

134. Spencer-Smith R, et al. (2016) Inhibition of RAS function through targeting an allosteric regulatory site. *Nat Chem Biol* 13(1):62–68.

135. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* 280(1):1–9.

136. Nave K-A, Trapp BD (2008) Axon-Glial Signaling and the Glial Support of Axon Function. *Annu Rev Neurosci* 31(1):535–561.

137. Snaidero N, et al. (2014) Myelin membrane wrapping of CNS axons by

PI(3,4,5)P3-dependent polarized growth at the inner tongue. *Cell* 156(1–2):277–290.

138. Noseworthy JH, Lucchinetti C, Rodriguez M, Weinshenker BG (2000) Multiple Sclerosis. *N Engl J Med* 343(13):938–952.

139. Olmos-Serrano JL, et al. (2016) Down Syndrome Developmental Brain Transcriptome Reveals Defective Oligodendrocyte Differentiation and Myelination. *Neuron* 89(6):1208–1222.

140. Bahi-Buisson N, et al. (2010) GPR56-related bilateral frontoparietal polymicrogyria: Further evidence for an overlap with the cobblestone complex. *Brain* 133(11):3194–3209.

141. Hamann J, et al. (2015) International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G Protein-Coupled Receptors. *Pharmacol Rev* 67(2):338–367.

142. Monk KR, et al. (2009) A G Protein-Coupled Receptor Is Essential for Schwann Cells to Initiate Myelination. *Science (80-)* 325(5946):1402–1405.

143. Mogha A, et al. (2013) Gpr126 Functions in Schwann Cells to Control Differentiation and Myelination via G-Protein Activation. *J Neurosci* 33(46):17976–17985.

144. Nishimori H, et al. (1997) A novel brain-specific p53-target gene, BAI1, containing thrombospondin type 1 repeats inhibits experimental angiogenesis. *Oncogene* 15(18):2145–2150.

145. Kuhnert F, et al. (2010) Essential Regulation of CNS Angiogenesis by the Orphan G Protein-Coupled Receptor GPR124. *Science (80-)* 330(6006):985–989.

146. Lin HH, et al. (2004) Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *J Biol Chem* 279(30):31823–31832.

147. Liebscher I, et al. (2014) A Tethered Agonist within the Ectodomain Activates the Adhesion G Protein-Coupled Receptors GPR126 and GPR133. *Cell Rep* 9(6):2018–2026.

148. White JP, et al. (2014) G protein-coupled receptor 56 regulates mechanical overload-induced muscle hypertrophy. *Proc Natl Acad Sci* 111(44):15756–15761.

149. Demberg LM, Rothemund S, Sch??neberg T, Liebscher I (2015) Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor GPR64/ADGRG2. *Biochem Biophys Res Commun* 464(3):743–747.

150. Scholz N, et al. (2015) The Adhesion GPCR Latrophilin/CIRL Shapes

Mechanosensation. *Cell Rep* 11(6):866–874.

151. Singer K, Luo R, Jeong SJ, Piao X (2013) GPR56 and the developing cerebral cortex: Cells, matrix, and neuronal migration. *Mol Neurobiol* 47(1):186–196.

152. Fujii Y, Ishikawa N, Kobayashi Y, Kobayashi M, Kato M (2014) Compound heterozygosity in GPR56 with bilateral frontoparietal polymicrogyria. *Brain Dev* 36(6):528–531.

153. Kim JE, et al. (2010) Splicing variants of the orphan G-protein-coupled receptor GPR56 regulate the activity of transcription factors associated with tumorigenesis. *J Cancer Res Clin Oncol* 136(1):47–53.

154. Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* 10(11):741–754.

155. Kalsotra A, Cooper TA (2011) Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* 12(10):715–729.

156. Irimia M, et al. (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159(7):1511–1523.

157. Bustanji Y, Samorí B (2002) The mechanical properties of human angiostatin can be modulated by means of its disulfide bonds: A single-molecule force-spectroscopy study. *Angew Chemie - Int Ed* 41(9):1546–1548.

158. Jin Z, et al. (2007) Disease-associated mutations affect GPR56 protein trafficking and cell surface expression. *Hum Mol Genet* 16(16):1972–1985.

159. Chiang NY, et al. (2011) Disease-associated GPR56 mutations cause bilateral frontoparietal polymicrogyria via multiple mechanisms. *J Biol Chem* 286(16):14215–14225.

160. Holm L, Rosenström P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(SUPPL. 2):W545-9.

161. Kim B-H, Cheng H, Grishin N V. (2009) HorA web server to infer homology between proteins using sequence and structural similarity. *Nucleic Acids Res* 37(Web Server):W532–W538.

162. Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ (2013) Dynamic integration of splicing within gene regulatory pathways. *Cell* 152(6):1252–1269.

163. Anderson GR, et al. (2015) β-Neurexins Control Neural Circuits by Regulating Synaptic Endocannabinoid Signaling. *Cell* 162(3):593–606.

164. Fuccillo M V., et al. (2015) Single-Cell mRNA Profiling Reveals Cell-Type-Specific Expression of Neurexin Isoforms. *Neuron* 87(2):326–341.

165. Schülein R, Westendorf C, Krause G, Rosenthal W (2012) Functional significance of cleavable signal peptides of G protein-coupled receptors. *Eur J Cell Biol* 91(4):294–299.

166. Celniker G, et al. (2013) ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem* 53(3–4):199–206.

167. Rudenko G, Hohenester E, Muller YA (2001) LG/LNS domains: Multiple functions - One business end? *Trends Biochem Sci* 26(6):363–368.

168. Araç D, et al. (2007) Structures of Neuroligin-1 and the Neuroligin-1/Neurexin-1β Complex Reveal Specific Protein-Protein and Protein-Ca2+ Interactions. *Neuron* 56(6):992–1003.

169. Domogatskaya A, Rodin S, Tryggvason K (2012) Functional Diversity of Laminins. *Annu Rev Cell Dev Biol* 28(1):523–553.

170. Yang L, et al. (2011) GPR56 regulates VEGF production and angiogenesis during melanoma progression. *Cancer Res* 71(16):5558–5568.

171. Luo R, Jin Z, Deng Y, Strokes N, Piao X (2012) Disease-associated mutations prevent GPR56-collagen III interaction. *PLoS One* 7(1):e29818.

172. Yang TY, et al. (2015) Expression and immunoaffinity purification of recombinant soluble human GPR56 protein for the analysis of GPR56 receptor shedding by ELISA. *Protein Expr Purif* 109(November):85–92.

173. Schlyer S, Horuk R (2006) I want a new drug: G-protein-coupled receptors in drug development. *Drug Discov Today* 11(11–12):481–493.

174. Christopoulos A (2014) Advances in G Protein-Coupled Receptor Allostery: From Function to Structure. *Mol Pharmacol* 86(5):463–478.

175. Shi D, et al. (2014) Celsr1 is required for the generation of polarity at multiple levels of the mouse oviduct. *Development* 141(23):4558–4568.

176. Fukuzawa T, et al. (2013) Lung Surfactant Levels are Regulated by Ig-Hepta/GPR116 by Monitoring Surfactant Protein D. *PLoS One* 8(7):e69451.

177. Coin I, et al. (2013) XGenetically encoded chemical probes in cells reveal the binding path of urocortin-i to CRF class B GPCR. *Cell* 155(6):1258–69.

178. Wilde C, et al. (2016) The constitutive activity of the adhesion GPCR GPR114/ADGRG5 is mediated by its tethered agonist. *FASEB J* 30(2):666–673.

179. Promel S, et al. (2012) The GPS motif is a molecular switch for bimodal activities of adhesion class G protein-coupled receptors. *Cell Rep* 2(2):321–31.

180. Chen J, Ishii M, Wang L, Ishii K, Coughlin SR (1994) Thrombin receptor activation: Confirmation of the intramolecular tethered liganding hypothesis and discovery of an alternative intermolecular liganding mode. *J Biol Chem* 269(23):16041–16045.

181. Gerszten RE, et al. (1994) Specificity of the thrombin receptor for agonist peptide is defined by its extracellular surface. *Nature* 368(6472):648–651.

182. Jacob C, et al. (2005) c-Cbl mediates ubiquitination, degradation, and down-regulation of human protease-activated receptor 2. *J Biol Chem* 280(16):16076–16087.

183. Koenig JA, Edwardson JM (1997) Endocytosis and recycling of g protein-coupled receptors. *Trends Pharmacol Sci* 18(8):276–287.

184. Kishore A, Hall RA (2017) Disease-Associated Extracellular Loop Mutations in the Adhesion G Protein-Coupled Receptor G1 (ADGRG1; GPR56) Differentially Regulate Downstream Signaling. *J Biol Chem*:jbc.M117.780551.

185. Christopoulos A (2014) Advances in GPCR Allostery: From Function to Structure. *Mol Pharmacol* (November):463–478.

186. J Gingell J, et al. (2016) An allosteric role for receptor activity-modifying proteins in defining GPCR pharmacology. *Cell Discov* 2:16012.

187. Christopoulos A (2002) G Protein-Coupled Receptor Allosterism and Complexing. *Pharmacol Rev* 54(2):323–374.

188. Paavola KJ, Sidik H, Zuchero JB, Eckart M, Talbot WS (2014) Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Sci Signal* 7(338):ra76-ra76.

189. Wang Z, Griffin M (2012) TG2, a novel extracellular protein with multiple functions. *Amino Acids* 42(2–3):939–949.

190. Asherson P, Gurling H (2012) Quantitative and molecular genetics of ADHD. *Curr Top Behav Neurosci* 9:239–272.

191. Südhof TC (2008) Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* 455(7215):903–911.

192. Egea J, et al. (2008) Genetic ablation of FLRT3 reveals a novel morphogenetic function for the anterior visceral endoderm in suppressing mesoderm differentiation. *Genes Dev* 22(23):3349–3362.

193. Leyva-Diaz E, et al. (2014) FLRT3 Is a Robo1-interacting protein that determines netrin-1 attraction in developing axons. *Curr Biol* 24(5):494–508.

194. Maretto S, et al. (2008) Ventral closure, headfold fusion and definitive endoderm migration defects in mouse embryos lacking the fibronectin leucine-rich transmembrane protein FLRT3. *Dev Biol* 318(1):184–193.

195. Muller P-S, et al. (2011) The fibronectin leucine-rich repeat transmembrane protein Flrt2 is required in the epicardium to promote heart morphogenesis. *Development* 138(7):1297–1308.

196. Lacy SE, Bönnemann CG, Buzney E a, Kunkel LM (1999) Identification of FLRT1, FLRT2, and FLRT3: a novel family of transmembrane leucine-rich repeat proteins. *Genomics* 62(3):417–426.

197. Böttcher RT, Pollet N, Delius H, Niehrs C (2004) The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling. *Nat Cell Biol* 6(1):38–44.

198. Söllner C, Wright GJ (2009) A cell surface interaction network of neural leucine-rich repeat receptors. *Genome Biol* 10(9):R99.

199. O'Sullivan ML, Martini F, von Daake S, Comoletti D, Ghosh A (2014) LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5. *Neural Dev* 9(1):7.

200. Karaulanov EE, Böttcher RT, Niehrs C (2006) A role for fibronectin-leucine-rich transmembrane cell-surface proteins in homotypic cell adhesion. *EMBO Rep* 7(3):283–290.

201. Karaulanov E, et al. (2009) Unc5B interacts with FLRT3 and Rnd1 to modulate cell adhesion in Xenopus embryos. *PLoS One* 4(5):e5742.

202. Krasnoperov VG, et al. (1996) The Calcium-Independent Receptor of α-Latrotoxin Is Not a Neurexin. *Biochem Biophys Res Commun* 227(3):868–875.

203. Deak F, et al. (2009) -Latrotoxin Stimulates a Novel Pathway of Ca2+-Dependent Synaptic Exocytosis Independent of the Classical Synaptic Fusion Machinery. *J Neurosci* 29(27):8639–8648.

204. Krasnoperov VG, et al. (1997) α-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. *Neuron* 18(6):925–937.

205. Sudhof TC (2001) α-Latrotoxin and Its Receptors: Neurexins and CIRL/Latrophilins. *Annu Rev Neurosci* 24(1):933–962.

206. Sugita S, Khvochtev M, Südhof TC (1999) Neurexins are functional alpha-

latrotoxin receptors. *Neuron* 22(3):489–96.

207. Arcos-Burgos M, et al. (2010) A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication. *Mol Psychiatry* 15(11):1053–1066.

208. Kan Z, Jaiswal BS, Seshagiri S (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Genome Biol* 11(Suppl 1):P37.

209. O'Hayre M, et al. (2013) The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. *Nat Rev Cancer* 13(6):412–424.

210. Sugita S, Ichtchenko K, Khvotchev M, Südhof TC (1998) α-Latrotoxin receptor CIRL/latrophilin 1 (CL1) defines an unusual family of ubiquitous G-protei-linked receptors: G-protein coupling not required for triggering exocytosis. *J Biol Chem* 273(49):32715–32724.

211. Langenhan T, et al. (2009) Latrophilin Signaling Links Anterior-Posterior Tissue Polarity and Oriented Cell Divisions in the C.??elegans Embryo. *Dev Cell* 17(4):494–504.

212. Levine A, et al. (1994) odd Oz: A novel Drosophila pair rule gene. *Cell* 77(4):587–598.

213. Sun H, et al. (2011) AAV-mediated netrin-1 overexpression increases peri-infarct blood vessel density and improves motor function recovery after experimental stroke. *Neurobiol Dis* 44(1):73–83.

214. Seiradake E, et al. (2014) FLRT Structure: Balancing Repulsion and Cell Adhesion in Cortical and Vascular Development. *Neuron* 84(2):370–385.

215. Donegan RK, et al. (2015) Structural basis for misfolding in myocilin-associated glaucoma. *Hum Mol Genet* 24(8):2111–2124.

216. Jackson VA, et al. (2015) Structural basis of latrophilin-FLRT interaction. *Structure* 23(4):774–781.

217. Han H, Kursula P (2015) The olfactomedin domain from gliomedin is a ??-propeller with unique structural properties. *J Biol Chem* 290(6):3612–3621.

218. Chen J, Sawyer N, Regan L (2013) Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* 22(4):510–515.

219. Domené S, et al. (2011) Screening of human LPHN3 for variants with a potential impact on ADHD susceptibility. *Am J Med Genet Part B Neuropsychiatr Genet* 156(1):11–18.

220. Wooldridge L, et al. (2009) Tricks with tetramers: how to get the most from multimeric peptide-MHC. *Immunology* 126(2):147–164.

221. Audet M, Bouvier M (2012) Restructuring G-Protein- Coupled Receptor Activation. *Cell* 151(1):14–23.

222. Wu H, et al. (2014) Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. *Science (80-)* 344(6179):58–64.

223. Natarajan K, Berk BC (2006) Crosstalk Coregulation Mechanisms of G Protein-Coupled Receptors and Receptor Tyrosine Kinases. *Transmembrane Signal Protoc* 332:51–78.

224. Noy A, Friddle RW (2013) Practical single molecule force spectroscopy: How to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope. *Methods* 60(2):142–150.

225. Marcsisin SR, Engen JR (2010) Hydrogen exchange mass spectrometry: What is it and what can it tell us? *Anal Bioanal Chem* 397(3):967–972.

226. Sniadecki NJ, et al. (2007) Magnetic microposts as an approach to apply forces to living cells. *Proc Natl Acad Sci* 104(37):14553–14558.

227. Provenzano PP, Keely PJ (2011) Mechanical signaling through the cytoskeleton regulates cell proliferation by coordinated focal adhesion and Rho GTPase signaling. *J Cell Sci* 124(8):1195–1205.

228. Gershlak JR, Black LD (2015) Beta 1 integrin binding plays a role in the constant traction force generation in response to varying stiffness for cells grown on mature cardiac extracellular matrix. *Exp Cell Res* 330(2):311–324.

229. Gordon WR, et al. (2015) Mechanical Allostery: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Dev Cell* 33(6):729–736.

230. Rasmussen SGF, et al. (2007) Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* 450(7168):383–387.

231. Liang Y-L, et al. (2017) Phase-plate cryo-EM structure of a class B GPCR–G-protein complex. *Nature* 546(7656):118–123.

232. Salzman GS, Zhang S, Koide A, Koide S, Arac D (2017) Stachel-independent activation and inhibition of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. *Procedings Nat Acad Sci USA*:Under Review.

233. Sha F, Salzman G, Gupta A, Koide S (2017) Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci* 26(5):910–924.

234. Reiser J, Sever S, Faul C (2014) Signal transduction in podocytes--spotlight on receptor tyrosine kinases. *Nat Rev Nephrol* 10(2):104–15.

235. Dorsam RT, Gutkind JS (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer* 7(2):79–94.

236. Rossmann MG, Argos P (1976) Exploring structural homologies of proteins. *J Mol Biol* 105:75.

237. Makalowski W, Boguski MS (1998) Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci* 95(16):9407–9412.