MOLECULAR ELECTRON MICROSCOPY
OF SIGNALING PROTEIN COMPLEXES

by

Gerwin Henri Westfield

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in The University of Michigan
2013

Doctoral Committee:

      Assistant Professor Georgios Skiniotis, Chair
      Assistant Professor Aaron C. Goldstrohm
      Associate Professor Roger K. Sunahara
      Associate Professor Raymond C. Trievel
      Associate Professor Zhaohui Xu

Dedication

This dissertation is dedicated to my parents George and Janis Westfield. I was always given the freedom to pursue whatever paths I wanted, and was always given the love and support to achieve those goals. I would certainly not have been able to accomplish everything that I have without them.

Table of Contents

Chapter 3 In celltro purification and architecture of β-arrestin in complex with a GPCR   78

Chapter 4   Structural Analysis of the COMPASS Family of Histone H3K4 Methylases 100

List of Figures

List of Tables

List of Abbreviations

7-TM	seven transmembrane
AC	adenylyl cyclase
AH	alpha helical
ARF	ADP ribosylation factor
$\beta_2 AR$	beta-2 adrenergic receptor
$\beta_2 V_2$	beta-2-adrenergic vassopressin-2 chimeric receptor
$\beta arr1$	beta-Arrestin-1
cAMP	3',5'-cyclic adenosine monophosphate
CC	cross correlation
CCD	charged coupled device
COMPASS	Complex Proteins Associated with Set1
cryo-EM	electron cryo-microscopy
CTF	contrast transfer function
DEER	double electron-electron resonance
DNA	deoxyribonucleic acid
ECL(1-3)	extracellular loops
EM	electron microscopy
ERK	extracellular signal-regulated kinase
FEG	field emission gun
FSC	Fourier shell correlation

G proteins	guanine nucleotide-binding regulatory proteins
Gαi	G protein inhibatory alpha subunit
GDP	guanosine diphosphate
GPCR	G protein coupled receptor
GPI	glycosylphosphatidylinositiol
GRK	G protein receptor kinase
GTP	guanosine triphosphate
GTPγS	guanosine 5'-O-[gamma-thio]triphosphate
Gαs	G protein stimulatory alpha subunit
H3K4	histone 3 lysine 4
ICL(1-3)	intracellular loops
LaB6	lanthanum hexaboride
LCP	lipidic cubic phase
MAPK	mitogen-activated protein kinases
MLL	mixed lineage leukemia
Nb	nanobody
NF-kB	nuclear factor-kB
NMR	nuclear magnetic resonance
PDB	Protein Data Bank
PI3K	phosphoinositide 3-kinase
PKA	protein kinase A
PPi	pyrophosphate
RCT	random conical tilt
SET	Su (var), Enhancer of Zeste, and Trithorax
T4L	T4 Lysozyme
TEM	transmission electron microscope
TM(1-7)	transmembrane helix

Abstract

G protein-coupled receptors (GPCRs) are the primary signal transduction units responsible for the communication between cells and their environment. Thus, GPCRs are involved in every aspect of human physiology and are currently the target of more than 30% of drugs. Understanding the activation, signaling, and silencing of GPCRs is crucial for designing therapeutic strategies and modern agents for treating several pathologies. Here we applied single particle electron microscopy (EM) analysis to obtain the first snapshots of agonist occupied β_2-adrenergic receptor (β_2AR), a well-studied model GPCR, in complex with its primary signal transducer, the heterotrimeric G protein Gs ($G\alpha s\beta\gamma$). EM 2D averages and 3D reconstructions of the detergent-solubilized complex revealed the complex architecture, and, unexpectedly, showed that the α-helical (AH) domain of $G\alpha s$ is highly flexible in the absence of nucleotide. In contrast, the presence of the pyrophosphate mimic foscarnet, and also the presence of GDP favor the stabilization of the AH domain on the Ras-like domain of $G\alpha s$. Furthermore, we employed single particle EM to study the 3D architecture of β_2AR in complex with β-arrestin-1 (βarr), which is responsible for GPCR silencing and internalization. To this end, we took advantage of an optimized β_2AR-βarr preparation that was based on complex stabilization by a Fab antibody. The 3D reconstructions and modeling of the β_2AR - βarr complex provide novel insights into arrestin binding on GPCRs. Thus, herein we describe a starting framework to understand the structural basis of GPCR activation, signaling and regulation.

Additionally, epigenetic regulation and signaling is a critical aspect of cellular function as well. We have reconstituted fully functional yeast and human Set1/COMPASS complex, the first H3K4 methylase, *in vitro* and have identified the minimum subunit composition required for histone H3K4 methylation. 3D cryo-EM reconstructions of the core yeast complex, combined with immunolabeling and 2D EM analysis of the individual

subcomplexes reveal a Y-shaped architecture with, the SET domain of Set1 is located at the juncture of Cps50, Cps30 and the Cps60-Cps25 module, lining the walls of a central channel that may act as the platform for catalysis and regulative processing of various degrees of H3K4 methylation.

Chapter 1   Introduction

1.1   Structure and function of membrane proteins

Membrane proteins are a tremendously important group of proteins that connects the extracellular environment to the interior environment of the cell. Membrane proteins serve a variety of functions including: transport of molecules and ions across the membrane, specific cell-cell interaction (ie. immune response), and signal relay between the cell's extracellular and intracellular environments. There are two major categories of membrane proteins; peripheral and integral. Peripheral membrane proteins are mostly transiently associated with the membrane through a combination of electrostatic, hydrophobic, and other non-covalent interactions. Additionally, peripheral proteins may be post-translationally modified with a variety of fatty acid chains or glycosylphosphatidylinositiol (GPI) that may anchor the protein to the lipid bilayer (Figure 1-1). Integral membrane proteins typically cross the plasma membrane with at least one pass possessing either α-helical or β-barrel fold. Their hydrophobic residues that interact with the fatty acyl groups of the phospholipid bilayer characterize these proteins as integral membrane proteins.
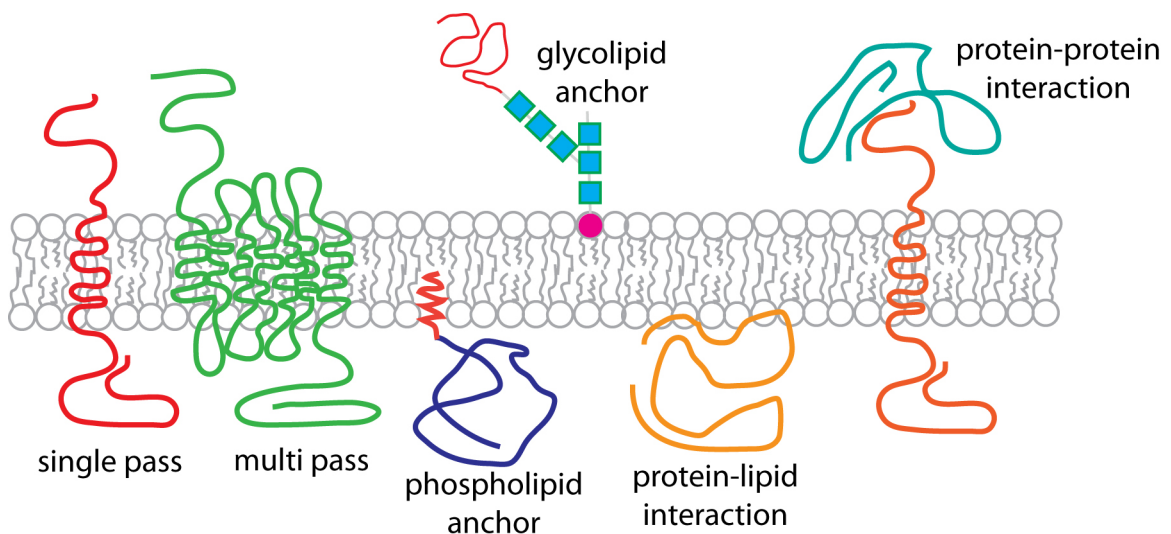
Figure 1-1. Illustration of membrane protein – plasma membrane interactions.

There are two types of membrane proteins; peripheral and integral. Peripheral membrane proteins are those that interact with one side (extracellular or intracellular) of the membrane, such as glycolipid or phospholipid anchor. Integral membrane proteins span across the lipid bilayer represented by the single-pass or multi-pass proteins.

While membrane proteins make up approximately 30% of the proteome, there are relatively few high- resolution structures determined because membrane protein structures are more difficult to crystalize than soluble proteins. Less than 2% of the coordinate files in the Protein Data Bank (PDB) consist of membrane proteins (Figure 1-2). There are many difficulties that cause this disparity, including protein expression and purification, solubilization, and further stabilization of hydrophobic regions of these membrane proteins that makes it difficult to obtain high-resolution structures. Recent advances in structural techniques have helped get past this roadblock, or at least alleviate some of these challenges.
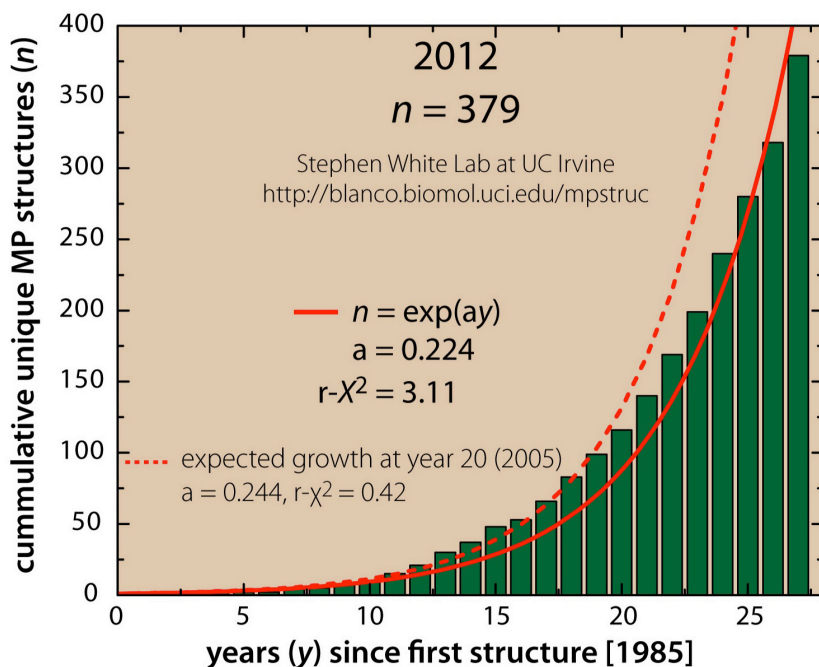
Figure 1-2. Number of membrane protein structures currently in the PDB data bank.

1.2   G protein-coupled receptors as signal transduction units

G protein-coupled receptor overview

G protein-coupled receptors (GPCRs) are the largest and most diverse membrane protein families, encoding more than 800 genes in the human genome (Fredriksson Mol. Pharm, 2003).   GPCRs consist of seven transmembrane (7-TM) α-helices that arrange themselves into a tertiary barrel-like structure (Figure 1-3). These integral membrane proteins are involved in the signal transduction that mediates many cellular responses to an assortment of ligands including: hormones, nucleotides, peptides, and amines. Additionally GPCRs are responsible for olfaction, taste and vision among many other physiological processes.
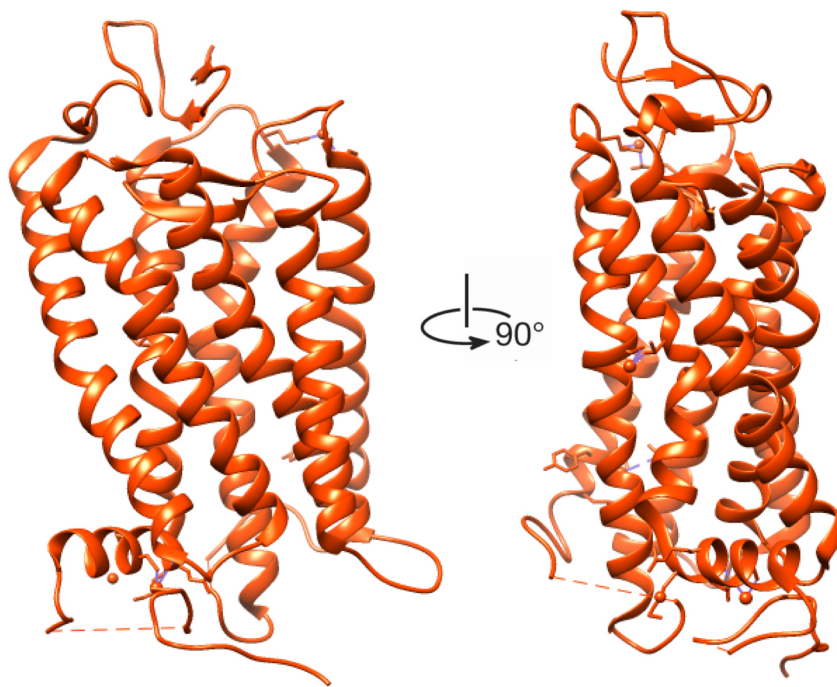
Figure 1-3. First high-resolution GPCR crystal structure.
The x-ray crystal structure of rhodopsin was solved in 2000 (Palczewski et al., 2000), and was the first high-resolution structure of a GPCR.

The 7-TM helices of GPCRs form a cavity near the plasma membrane that acts as a ligand-binding domain, threreby, mediating this cross-talk across the membrane with

intracellular interactions with guanine nucleotide-binding regulatory proteins (G proteins) (Strader et al., 1994). There are five main families of GPCRs classified based on their three-dimensional structural similarities and amino acid sequence: A, B, C, adhesion and Frizzled/Taste2 (Fredriksson et al., 2003).

GPCR structural characteristics

The GPCR architecture is typically divided in three regions: the extracellular region, consisting of the amino terminus and three extracellular loops (ECL1-ECL3); the transmembrane (TM) region containing seven α-helices (TM1-TM7); and the intracellular region possessing three intracellular loops (ICL1-ICL3), an intracellular amphipathic helix (H8), and the carboxy terminus (Venkatakrishnan et al., 2013) (Figure 1-4). The greatest homology across different GPCRs is within the transmembrane regions, with the most diversity coming on the carboxy terminus, the amino terminus, and the intracellular loop spanning between helix 5 (TM5) and helix 6 (TM6). The most variability occurs on the N-terminus where the sequence can be fairly short for monoamine and peptide receptors (10-50 amino acids), and much longer for glutamate and glycoprotein hormone receptors (350-600 amino acids) (Kobilka, 2007). The conserved tight barrel-like formation of the hydrophobic 7-TM helices is important for structural stability and is involved with ligand binding.
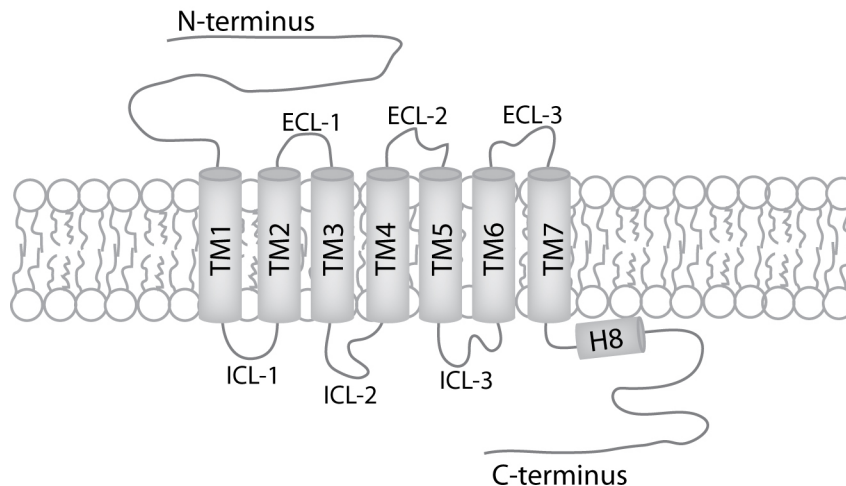
Figure 1-4. Schematic of a GPCR
A representative diagram of the general structure and domain assignment of a GPCR seven transmembrane protein (7-TM). The extracellular portion consists of an N-terminal domain and three extracellular loops (ECL1-3). There are seven transmembrane

helices (TM1-7). The intracellular portion includes 3 intracellular loops (ICL1-3), an amphipathic helix 8 (H8), and a C-terminal tail.

Modes of GPCR activation

The large extent of sequence diversity amongst the N-terminus and extracellular loops of GPCRs is thought to partially dictate ligand specificity. The interaction between the ligand and the extracellular portion of the receptor is critical to initiate a transmembrane signal transduction event. Ligands are categorized by their effect on the receptor, and they typically fall into three major categories: (1) agonist, a compound that elicits maximal activation response of the receptor, (2) partial agonist, a compound that activates the receptor, but is unable to get maximal response, and (3) inverse agonist that has the same binding site as agonist, but exerts the opposite pharmacological effect of agonist or reverses constitutive activity of receptor. Upon ligand binding to the extracellular of the receptor, the receptor undergoes a conformational rearrangment. In the case of an agonist binding to a receptor, this ligand-binding event causes the receptor to rearrange to its activated state propagating the signal through the membrane. This also causes a conformation change to the intracellular portion of the receptor, allowing the receptor in its activated to state to interact with the intracellular G protein (Figure 1-5).
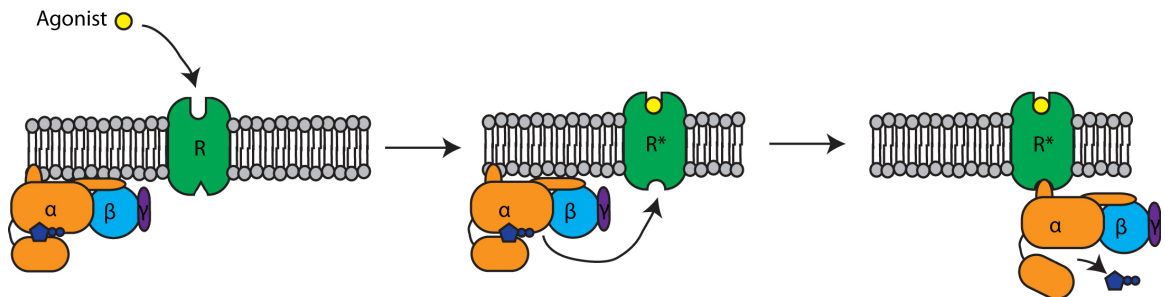
Figure 1-5. Diagram of GPCR activation and G protein interaction.
In their respective basal states, the G protein heterotrimer is in its GDP bound, trimeric form, and the receptor is in its unbound, inactive state. The G protein cannot interact stably with the inactive receptor conformation. Upon agonist binding to the extracellular portion of the receptor, this causes a conformational shift of the transmembrane helices changing the intracellular conformation. This intracellular conformation in the receptor's active form now allows the G protein to form a complex with the receptor, and further transmit the signal.

β2-adrenergic receptor as model system

The β2-adrenergic receptor (β$_2$AR) has been a model system for studying the diverse family of GPCRs over the past 40 years. β$_2$AR is a GPCR that is activated by adrenaline, and plays important roles in pulmonary and cardiovascular physiology. The sites of interaction between receptor and agonist have been extensively studied through mutagenesis (Strader et al., 1989; Wieland et al., 1996), and many biophysical studies characterizing agonist binding and conformational changes that lead to activation (Ghanouni et al., 2001a; Ghanouni et al., 2001b; Yao et al., 2006). Additionally there are several determined crystal structures of the β$_2$AR bound to partial inverse agonist and antibody (Rasmussen et al., 2007), bound to inverse agonist with mutations to help solubilization (Cherezov et al., 2007), bound to irreversible agonist in its active form (Rosenbaum et al., 2011), and in its active form bound to camelid antibody (Rasmussen et al., 2011) (Figure 1-6). These different high-resolution structures have aided in better understanding the conformational changes that occur upon β$_2$AR activation.
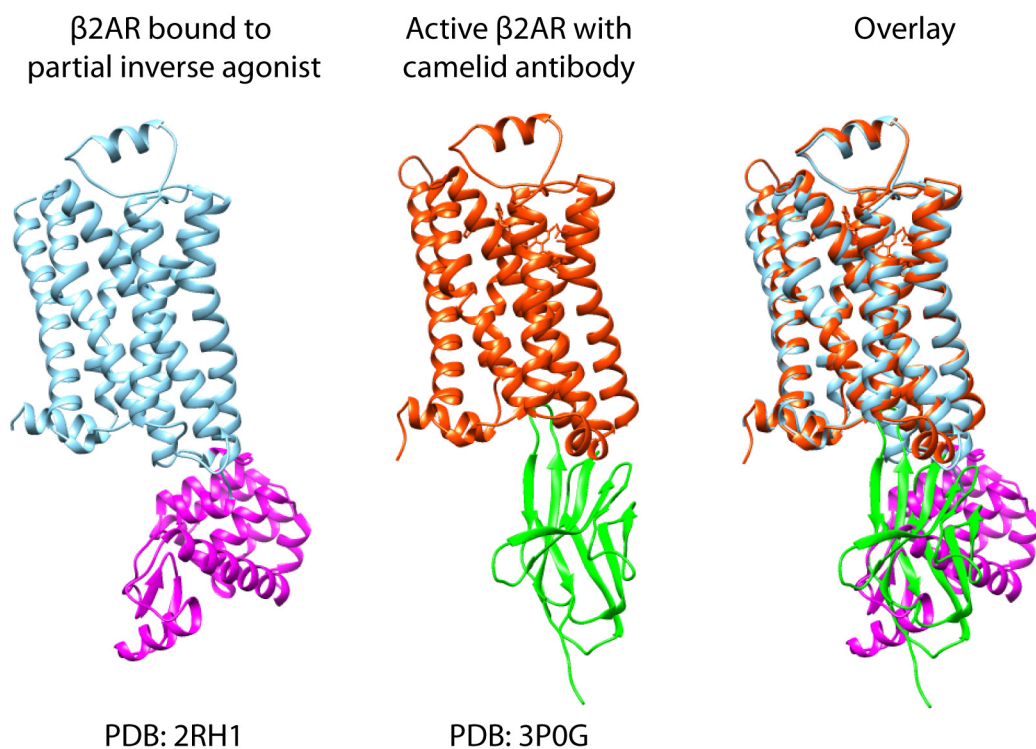
Figure 1-6. β2AR crystal structures.
Left, β2AR crystal structure stabilized by partial inverse agonist, and T4 Lysozyme (purple). (PBD: 2RH1). Middle, β2AR crystal structure in active form stabilized by camelid antibody, nanobody 80 (Nb80, green; PDB: 3P0G). Right, overlay of two β2AR crystal structures.

As mentioned above, the majority of GPCR signaling is mediated through GPCR/G protein interactions. β_2AR can couple to both the stimulatory subunit of the heterotrimeric G protein (Gαs) and the inhibitory subunit (Gαi) (Xiao et al., 1999), but can also signal in a G protein-independent manner through the MAP kinase pathway mediated by arrestin (Azzi et al., 2003). β_2AR desensitization involves many pathways that includes receptor phosphorylation, arrestin-mediated endosomal internalization, lysozomal degradation, and receptor recycling (Freedman and Lefkowitz, 1996).

1.3   G protein heterotrimer and other down stream effectors

Guanine nucleotide-binding proteins (G proteins) are a family of proteins that are involved in intracellular signal transduction, primarily mediated by GPCRs. G proteins belong to the GTPase family of enzymes given that their activity is regulated by the ability to bind and hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP). There are two classes of G proteins: the first includes small monomeric proteins that function as GTPases (eg. Ras), and the second class includes those that form and function in complex as a G protein heterotrimer.

a
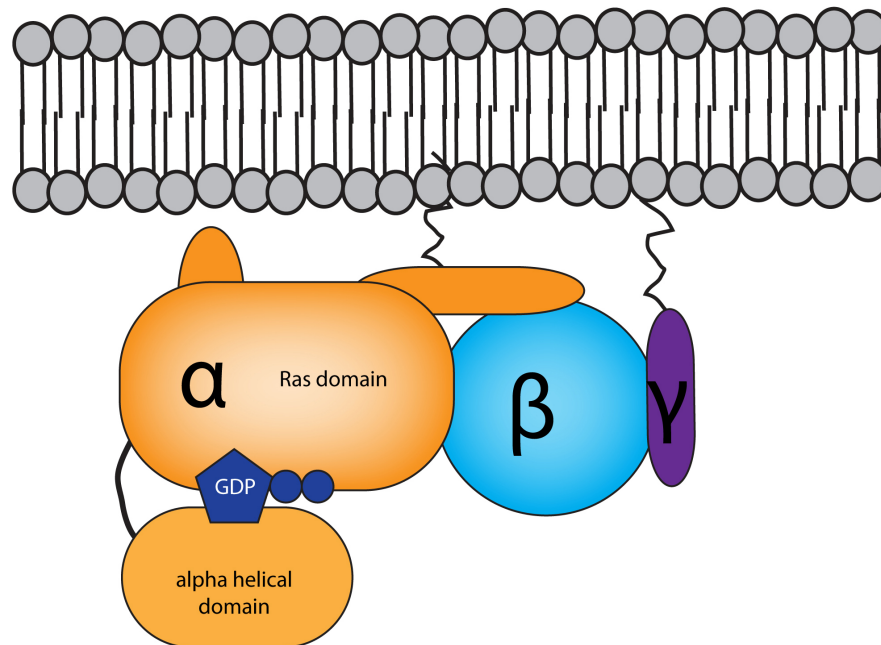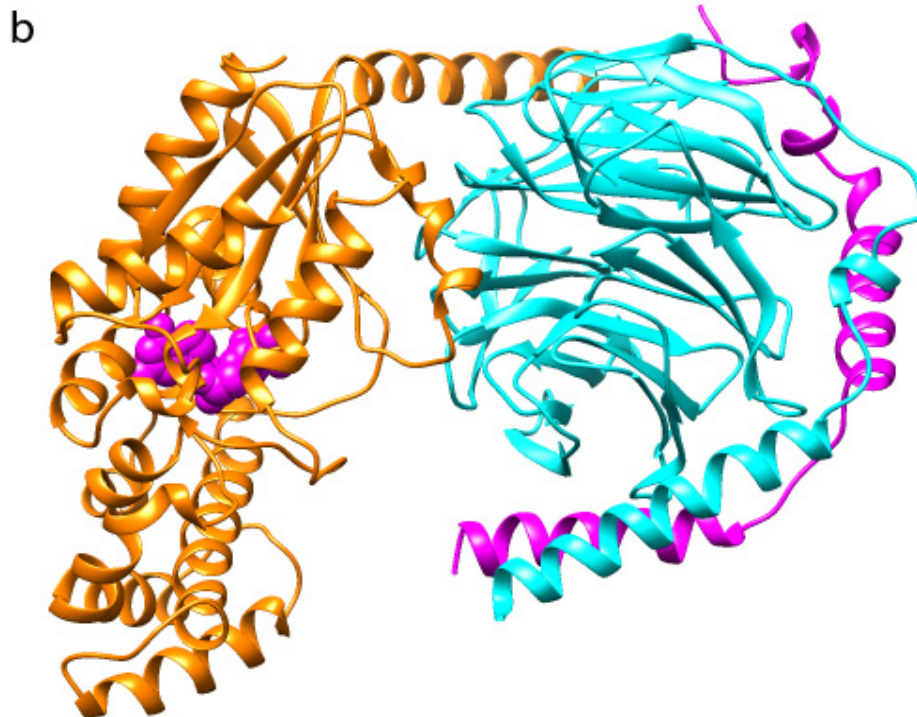
Figure 1-7. G-protein heterotrimer.
a, Diagram of a heterotrimeric G protein that consists of three subunits: Gα, β, and γ. Gα subunit (orange) has two domains, the Ras-like domain that has GTPase activity and the alpha helical domain. The β subunit has a WD40 seven-β propeller fold, and the γ subunit forms a constitutive heterodimer with the β subunit. The α and γ subunits also have posttranslational modifications to interact with the lipid bilayer. b, Crystal structure of G_t heterotrimer (Lambright et al. *Nature*. 1996; PDB: 1GOT). Subunits have same respective coloring as seen in (a) with GDP bound to the α subunit (purple).

The G protein heterotrimer consists of three subunits: a guanyl nucleotide binding α subunit (39 – 52 kDa), a β subunit (~35 kDa), and a γ chain (8 kDa) (Stryer and Bourne, 1986) (Figure 1-7). G proteins are divided into four families based on the sequence similarity in the Gα subunit: Gαs, Gαi, Gαq, and Gα12 (Oldham and Hamm, 2008). The α-subunit possesses two domains: the Ras domain that has GTPase activity, and the alpha helical domain. The Ras domain of Gα is the most highly conserved region. It possesses the GTPase activity, but also is involved in the binding interface to the βγ subunit to form the heterotrimeric complex. The β subunit is characterized by its WD40 repeat, seven β-propeller fold (Wall et al., 1995). The γ subunit forms a heterodimer with the β subunit by a coil-coil interaction on the γ N-terminus. Additionally, the C-terminus

of the γ subunit undergoes an isoprenyl post-translational modification that allows for membrane interaction.

β₂AR and the G protein cycle

β₂AR can activate two families of G proteins: Gs (stimulatory) and Gi (inhibitory), which differentially regulates one of its downstream effectors adenylyl cyclase (AC). In its basal, inactive state the G protein heterotrimer is in a GDP bound state. When agonist binds to the receptor, β₂AR goes from an inactive to active conformation. This conformational change on the intracellular portion of the receptor, now allows the G protein heterotrimer to associate with the receptor, forming a receptor/G protein complex. Upon complex formation, GDP is released from the α-subunit of the G protein. Due to significantly higher concentrations of GTP relative to GDP inside a cell, GTP then binds to the nucleotide-free α subunit leading to the dissociation of the G protein α and βγ heterodimer subunits from the receptor.
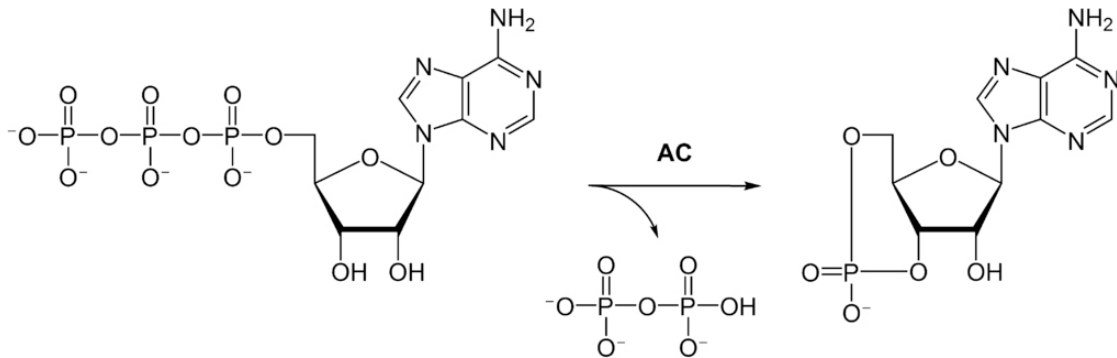
Figure 1-8. Conversion of ATP to 3',5'-cyclic AMP and diphosphate mediated by adenylyl cyclase (AC).

After the dissociation of the G protein heterotrimer, the separate Gα-GTP and Gβγ subunits go on to modulate the activity of different cellular effectors (ie. kinases, channels, and other enzymes). In the case of Gαs, it proceeds to interact and stimulate adenylyl cyclase (AC). AC catalyzes the conversion of ATP to 3',5'-cyclic AMP (cAMP) (Figure 1-8). cAMP is characterized as a second messenger, and it is used for intracellular signal transduction. It is involved in the regulation of ion channels, protein kinases, and other cyclic nucleotide binding proteins. Gαi conversely leads to the

inhibition of AC, and therefore lower levels of cAMP. The dissociated Gβγ heterodimer also has its own down stream effectors, for example, regulating the function of Ca^{2+} channels. GTP hydrolysis then occurs in the Gα subunit, and the G protein heterotrimer reassembles from the Gα and Gβγ subunits (Figure 1-9).
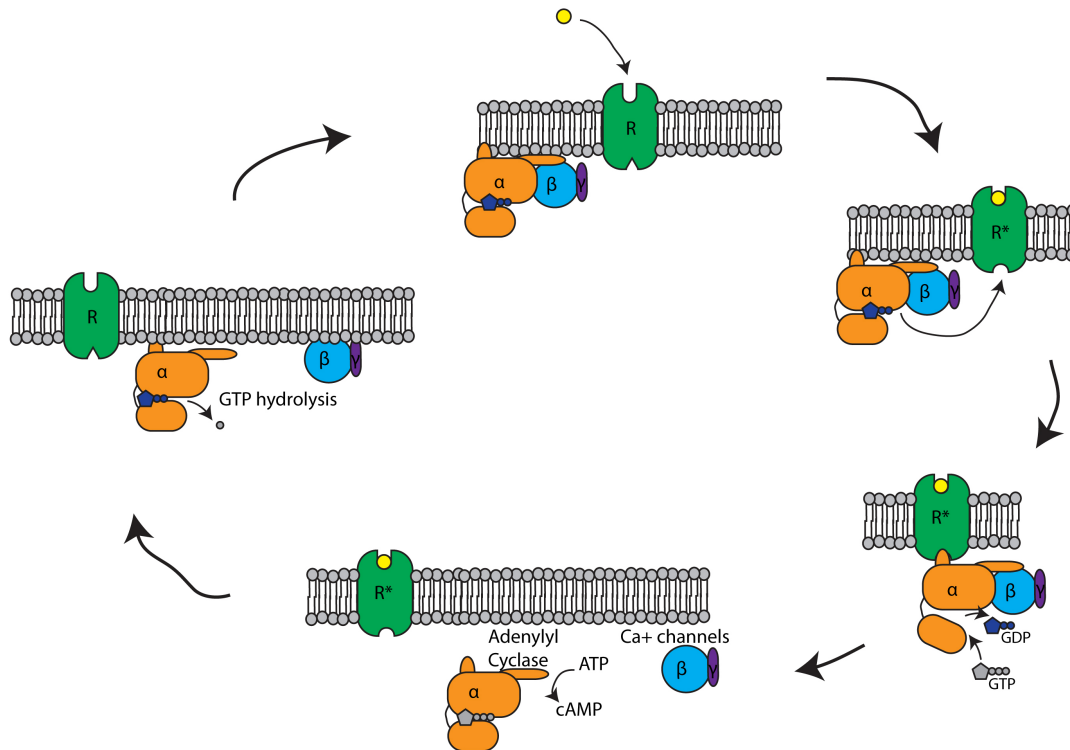
Figure 1-9. Activated receptor – G protein cycle
In their respective basal states, the G protein heterotrimer is in its GDP bound, trimeric form, and the receptor is in its unbound, inactive state. The G protein cannot interact with the inactive receptor. Agonist binding to the extracellular portion of the receptor causes a conformational shift of the transmembrane helices changing the intracellular conformation. This intracellular conformation in the receptor now allows the G protein to form a complex with the receptor. Upon complex formation GDP is released from the α- subunit and is replaced with GTP that is in relatively higher concentrations in the cell. GTP binding causes the dissociation of the G protein heterotrimer in the α-GTP monomer and βγ heterodimer. Gα interacts with its downstream effector Adenylyl cyclase that converts ATP to cyclic AMP (cAMP), while the βγ heterodimer regulates Ca^{2+} channels. Once the Gα subunit hydrolyzes GTP to GDP, the G protein heterotrimeric complex reforms.

GPCR silencing

After activation, GPCRs can become silenced or desensitized by arrestins. Arrestins are versatile adapter proteins that form complexes with GPCRs after agonist binding and phosphorylation by G protein receptor kinases (GRKs) (Luttrell and Lefkowitz, 2002). Arrestins consist of four subtypes: Arrestin-1, Arrestin-2 (β-Arrestin-1), Arrestin-3 (β-Arrestin-2), and Arrestin-4 (cone arrestin). Arrestin-1 and Arrestin-4 are considered 'visual arrestins' primarily being expressed in the retina, while the β-Arrestins (1 and 2) are expressed ubiquitously. β-Arrestin-1 (βarr1) can bind and desensitize different GPCRs including β_2AR and rhodopsin, but has a preference for β_2AR (Lohse et al., 1990). β-Arrestin leads to the termination of G protein activation by uncoupling the receptor from the G protein, and also is involved in targeting receptors for clathrin-mediated degradation (Figure 1-10).
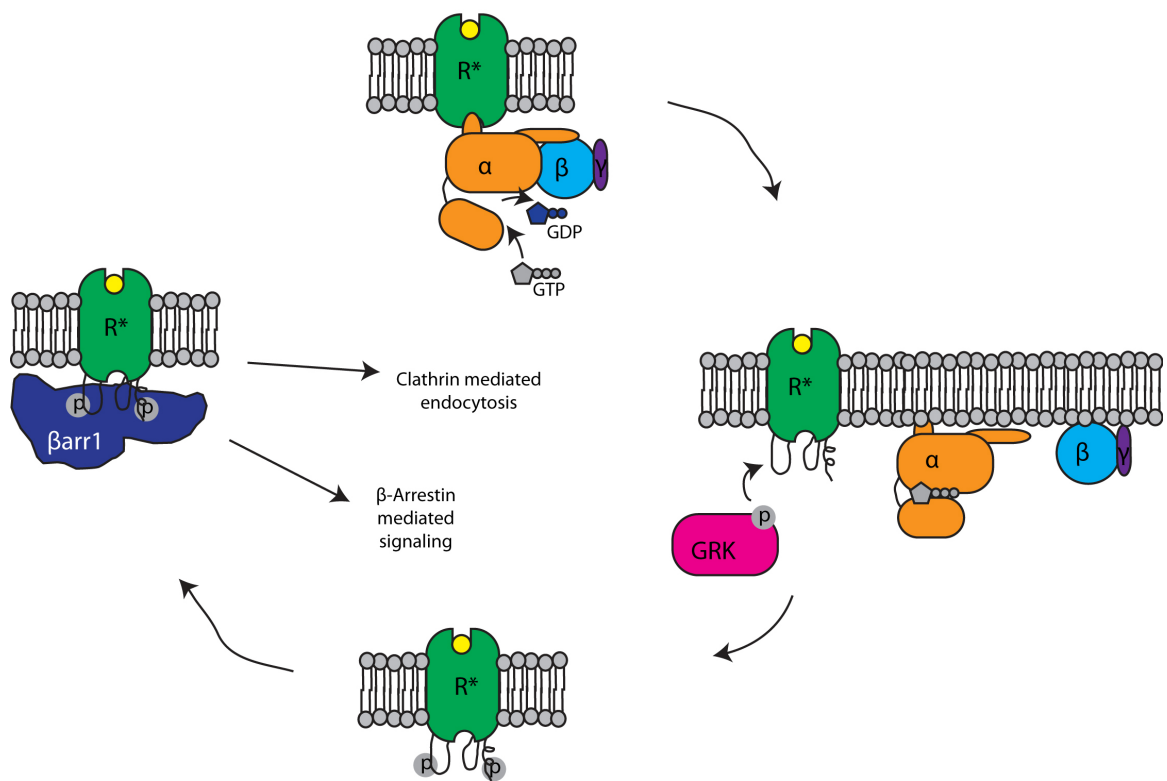
Figure 1-10. Receptor – arrestin cycle
As mentioned in the previous figure, once GDP is released and GTP binds the Gα-subunit, the G protein heterotrimer dissociates and moves to interact with downstream effectors. Subsequently the C-terminal tail and intracellular loops of the receptor become phosphorylated by G protein receptor kinases (GRKs). Intracellular receptor phosphorylation leads to the recruitment of arrestin. Receptor-arrestin complex

formation leads to subsequent clathrin mediated endocytosis, or Arrestin dependent signaling.

1.4 β-arrestins

Mechanism of β-arrestin / receptor interaction

The primary function of arrestin is to stop receptor signaling via G proteins. This was how the first member of its family, visual arrestin (Arrestin-1), was discovered (Kuhn, 1978). Before its function was fully known, it was understood that the activation and subsequent phosphorylation of rhodopsin led to the binding of arrestin. Later on, the ubiquitously expressed G protein receptor kinases (GRKs) were discovered based on the lack of coupling between β$_2$AR and G protein (Benovic et al., 1987). Arrestins have an elongated structure with two β-sheet domains, the C domain and the N domain.
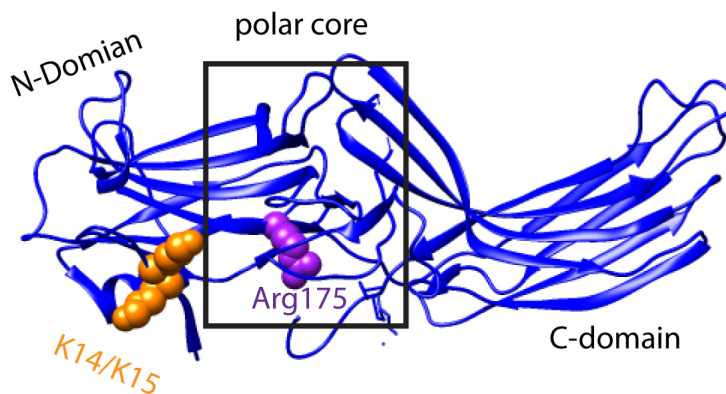
Figure 1-11. Crystal structure of β-Arrestin1.
The main phosphate sensor, Arg175 (highlighted in purple), shown in the polar core of arrestin is shielded, whereas other positively charged residues in the N-domain (highlighted in orange (K14/K15) are highly exposed.

A model of arrestin-receptor activation shows that there is a sequential multi-site process (Gurevich and Benovic, 1993). This shows arrestin has 2 "sensor" sites: a "phosphate sensor" that interacts with receptor-bound phosphates, and an "activation sensor" that binds the receptor after conformational changes upon activation. Mutagenesis assays have helped identify the residues that contribute to the "phosphate sensor," and are located in what is termed the polar core including residues from both the C and N domains of arrestin (Figure 1-11).

Non-canonical β-arrestin function

The canonical view of GPCR silencing where an activated receptor becomes phosphorylated by GRKs, forms a complex with arrestin, is desensitized and then is subsequently subjugated to clathrin-mediated endocytosis for degradation is only a part of β-arrestins function. β-arrestins also function as adaptor proteins that possess the ability to signal through various mediators such as phosphoinositide 3-kinase (PI3K), mitogen-activated protein kinases (MAPKs), SRC, and nuclear factor-κB (NF-κB) (Rajagopal et al., 2010) (Figure 1-12). One example of β-arrestin 1 mediated signaling is the recruitment of activated SRC, a non-receptor tyrosine kinase, which subsequently leads to the downstream activation of extracellular signal-regulated kinase (ERK) (Luttrell et al., 1999). The concept of biased agonism, where the binding of different ligands can preferentially dictate which signaling pathway to activate has been shown in the activation of different G proteins is hypothesized to apply to differential arrestin signaling as well.
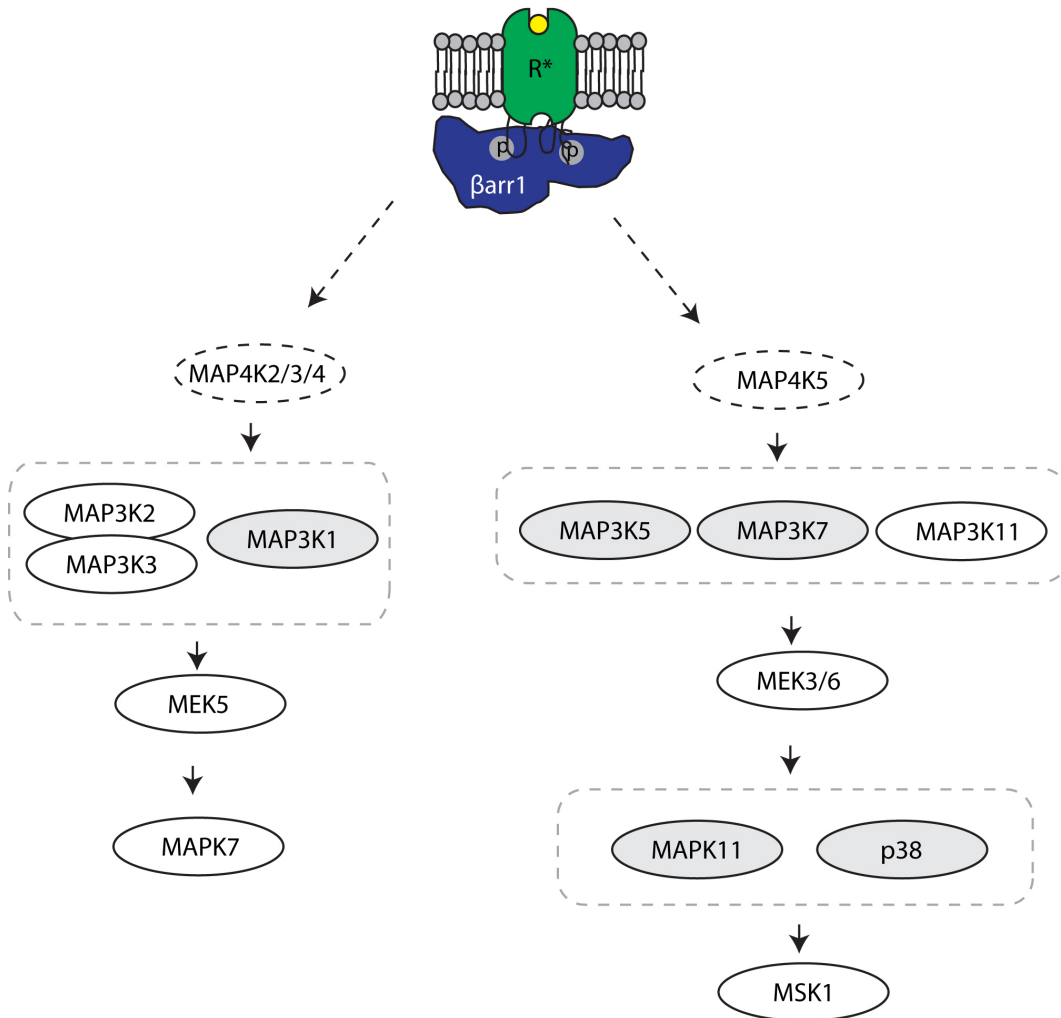
Figure 1-12. β-Arrestin signaling
Example of a β-Arrestin-dependent mitogen-activated protein (MAP) kinase signaling network downstream from the angiotensin receptor subtype 1a (AT1aR).

1.5   Epigenetic Regulation

The organization of genetic information is a well-orchestrated event. As cells go through different processes such as development and differentiation, how DNA is packed, moved, activated, silenced, and repaired is regulated by a number of different epigenetic cues. Histones pack DNA into nucleosomes, which the most basic form of packing and organizing DNA into the nucleus (Kornberg, 1974; Luger et al., 1997). The N-termini of histones consist of tails that protrude outward from the nucleosome and undergo different posttranslational modifications. These modifications regulate the accessibility of DNA to regulatory factors and affect events such at DNA repair, transcription, and replication.

Acetylation, methylation, sumoylation, ubiquitination, deimination, and ADP-ribosylation are the main posttranslational modifications that histones can undergo (Shilatifard, 2006; Zhang, 2003). Typically, histone lysine acetylation and serine/threonine phosphorylation upregulate gene expression, whereas the effects of histone lysine methylation are site-dependent (Shilatifard, 2006). Additionally, the extent of the histone lysine methylation (mono-, di-, or trimethylation) adds to the complexity of this type of regulation.

Methylation of lysine 4 on histone 3 (H3K4) is a paradigm for the complexity of outcomes from histone tail posttranslational modifications due to the variation of process regulation relative to the level of methylation. Mono-, di-, and trimethylation of H3K4 have been shown to influence and regulate gene activation and transcription. H3K41me (monomethlylation), and to a greater level, H3K43me (trimethylation) have been shown to activate gene transcription. The first H3K4 methylase, Set1, was discovered in yeast and possess a catalytic SET domain (Miller et al., 2001). Set1 in yeast forms a macromolecular complex named COMPASS (Complex Proteins Associated with Set1), and has the capability to mono-, di-, and trimethylate H3K4 (Krogan et al., 2002; Miller et al., 2001).

1.6   COMPASS histone methyltransferase complex

The COMPASS complex was discovered in yeast as the first H3K4 methylase. It possesses the ability to mono-, di-, and trimethylate H3K4 via its catalytic subunit Set1. The other subunits in the yeast COMPASS complex are named Cps60 (Bre2), Cps50 (Swd1), Cps40 (Spp1), Cps35 (Swd2), Cps30 (Swd3), and Cps25 (Sdc1) according to their apparent molecular weight. How each subunit works in the methyltransferase activity of the complex has been the topic of intense investigation. The Cps30 and Cps50 subunits are characterized by their WD40 β-propeller fold. A crystal structure of WDR5, the human homolog of Cps30, was co-crystallized with peptides of RbBP5 (Cps50) and MLL1 (Set1) giving further evidence to the protein-protein interactions within the complex (Figure 1-13). Set1, Cps50, and Cps30 are necessary for mono-, di-, and trimethylation of H3K4, however, Cps25 and Cps40 are only necessary for the trimethylation of histone H3K4 (Schneider et al., 2005). There are homologous

COMPASS-like complexes in higher organisms, with at least seven complexes in mammalian cells that are capable of H3K4 methylation (Shilatifard, 2008). The MLL complexes (1-4) in mammalian cells have been shown to form COMPASS-like complexes and function as H3K4 methylases. Many subunits in these complexes are shared and have conserved function from yeast to human.
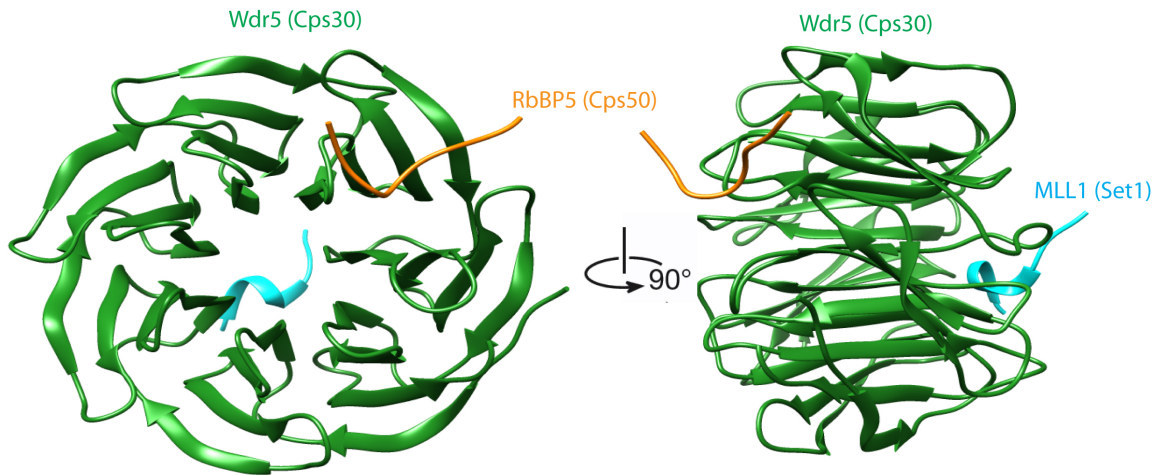
Figure 1-13. Crystal structure of human Cps30 (WDR5), co-crystallized with peptides.
Crystal structure of human WDR5 (green), yeast Cps30 homolog. Crystallized with RbBP5 (Cps50, orange), and MLL1 (Set1, cyan) peptides. PDB: 3P4F.

Set1 methyltransferase

Yeast Set1 methyltransferase is a 1,080 amino acid protein that contains two RNA recognition motifs (RRM1&2) near the C-terminus, an n-SET domain between amino acids (AAs) 762 and 938, the SET domain that contains the catalytic residues from AAs 938-1061, and a post-SET domain at the N-terminus. A recent crystal structure of the human MLL1 SET domain in complex with a histone H3 peptide revealed the binding characteristics of the peptide in the SET active site (Southall et al., 2009).

a

b
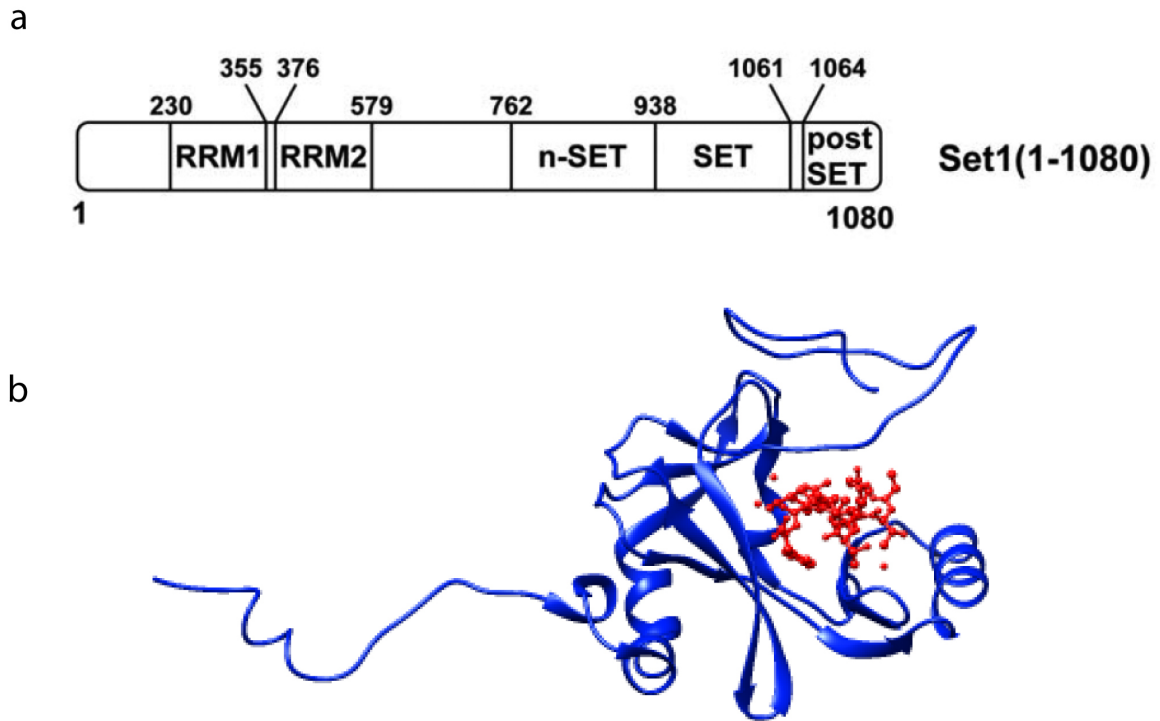
Figure 1-14. Set1 schematic and MLL1 crystal structure.
Schematic of the yeast Set1 protein. Set1 contains two RNA recognition motifs (RRM), an n-SET domain between residues 762 and 938, the catalytic SET domain from residues 938 to 1061, and a post-SET domain at the C-terminus. Bottom, crystal structure of the SET domain of human MLL1 (blue) co-crystalized with histone H3 peptide (red) (PDB: 2W5Z).

The essential active site residues (Phe3884, Tyr3942, Tyr3944, and Phe3946 and the main chain of the tetrapeptide, residues Cys3882 to Phe3885) have an arrangement similar to those of other SET domains. While Set1 and MLL(1-4) is the catalytic subunit of the COMPASS methyltransferase complex, they cannot methylate H3K4 alone, but require additional subunits of optimal activity. The additional subunits of COMPASS are necessary for mono-, di-, and trimethylation function, highlighting the role of the entire complex in methylation events.

COMPASS crosstalk in H3K4 methylation

Different epigenetic cues are the foundation for regulating the complexity of DNA organization, repair, replication, and gene expression. Recognition of different histone

posttranslational modifications for one purpose can lead to the occurrence of additional modifications to initiate an entirely different function. It has been shown that COMPASS purified from yeast strains lacking H2B monoubiquitination has several fold lower presence of the Cps35 subunit in the complex (Lee et al., 2007). COMPASS that has been purified from the same H2B monoubiquitin-deficient background lacks the enzymatic ability to trimethylate H3K4. Additionally, Cps35 has been shown to interact with chromatin in an H2B monoubiquitination-dependent manner (Lee et al., 2007; Takahashi et al., 2009). Wdr82, the human homolog of Cps35 also requires H2B monoubiquitination to interact with chromatin. These findings led to the hypothesis that COMPASS's trimethylation of H3K4 was partially mediated by crosstalk between Cps35 and H2B monoubiquitination. More recently, other evidence has suggested that the n-SET domain of Set1, and not Cps35, regulates H2B ubiquitylation-dependent H3K4 methylation (Kim et al., 2013). Further studies are needed to better understand what regulates this crosstalk.

1.7   Structural techniques in studying membrane proteins

Difficulties in structurally characterizing membrane proteins

The plasma membrane provides a crucial barrier that separates the extracellular environment from the intracellular space of a cell. The embedding of membrane proteins into the lipid bilayer provides a delicate interface that helps regulate their given function. Signal transduction, motility, respiration, and photosynthesis are some of the biological functions regulated by this lipid bilayer-membrane protein interaction. Lipids also serve an important role in helping to stabilize proteins embedded within them. Studying membranes proteins poses difficulties due to the fact that in order to characterize them structurally, you need to remove (or purify) them from the lipid bilayer. During protein purification, membrane proteins are typically solubilized with detergent and most of the lipids are lost in this process. In many cases, lipids are re-added during the purification process in order to restore the membrane protein stability as seen in the purification of the β_2AR (Cherezov et al., 2007; Raunser and Walz, 2009).

Methodologies for characterizing the structure of membrane proteins

There are a number of different techniques that are used to study the structural characteristics of membrane proteins and membrane protein complexes. Each method has its own benefits and drawbacks depending on the characteristics of the protein of interest. There are numerous factors that contribute to the complexity of studying membrane proteins including size, solubility, protein expression, flexibility, hydrophobicity, and quaternary organization. Below is an overview of some of the more commonly used techniques for studying the structure of membrane proteins.

X-Ray crystallography

X-ray crystallography is the most widely used technique for studying the structure of proteins, especially at high resolution. The majority of the structures in the protein data bank (PDB) are generated from X-ray crystallography. For this technique, after the protein of interest is purified, it is incubated with a series of different combinations of solvents and precipitants to find ideal conditions for crystal nucleation and growth. Crystal formation involves protein is packing within identical repeating units in the crystal lattice. Protein crystals are then subjected to an intense X-ray beam where a characteristic diffraction pattern is recorded, and information about the electron densities of the protein are calculated and mapped. X-ray crystallography is a powerful tool for studying nucleic acid and protein structure to atomic resolution. This technique is excellent for studying samples of a modest size that can nicely pack into a crystal lattice, and form ordered crystals. Larger proteins and protein complexes that have significant variability are more difficult to form highly ordered crystals. The inherent flexible nature of membrane proteins (ie. conformational changes of β_2AR transmembrane helices rearranging from inactive to active states) and protein solubility issues leads to difficulties in forming highly ordered crystal structures. Recent advances in protein purification and engineering, for example, insertion of a well-ordered globular domaing such as T4 lysozyme domain to the 3rd intracellular loop of the β_2AR for increased solubility and enhance protein-protein interactions and lattice formation (Cherezov et al., 2007), have helped advance using x-ray crystallography to study membrane proteins. Additionally, lipids and detergents are needed to help stabilize and order membrane

proteins. The use of lipidic cubic phase (LCP) where membrane proteins can nucleate crystals and grow in a 3-dimensional, continuous lipid bilayer matrix has aided tremendously in the study of 7-TM receptors (Landau and Rosenbusch, 1996). Finally the use of different antibodies and nanobodies (camelid antibodies) in combination with insertions of T4 lysozyme to the N-terminus of receptors have helped in advancing the 7-TM crystallization with other proteins, in the case of the β_2AR-Gs protein complex (Rasmussen et al., 2011).

Nuclear Magnetic Resonance (NMR)

Both solid state and solution NMR have been used to characterize membrane protein structure and membrane protein dynamics. NMR records the magnetic properties of different isotopes of atoms, and employs the collective resonance of the nearby atoms to calculate the location of each individual residue of a protein. The benefit to using NMR is that it can record a range of different protein states over a period of time and also study time related conformational changes in a protein sample, as opposed to x-ray crystallography that provides a rigid snapshot of a crystallized protein. One of the complications with using NMR as well as x-ray crystallography is that large amounts of purified proteins, typically milligrams, are needed to calculate structural properties. This can pose to be extremely difficult when attempting to purify large amounts of membrane proteins. What differentiates solution from solid-state NMR is in solution NMR, resonance is calculated by the rotational properties of the protein sample freely tumbling in solution NMR (Montaville and Jamin, 2010), whereas in solid state NMR when there is little to no mobility, the anisotropic interactions have a substantial effect on the behavior of nuclear spins. Larger proteins (size greater than 40 kDa), have a slower tumbling in solution than smaller proteins, and therefore makes it more difficult to study with solution NMR. The overlapping architecture of different hydrophobic residues in membrane proteins, forming primarily interacting α-helices and β-strands, can somewhat mask the signal from these residues and complicate the NMR spectra. This further adds to the difficulty of using NMR to study membrane proteins.

Electron Microscopy

Electron microscopy (EM) is a very useful tool in determining the architecture of proteins and protein complexes. There are various techniques used in electron microscopy to characterize the three dimensional shape of proteins including the formation of 2D crystals (2D crystallography) or averaging numerous individual particles in the case of single particle analysis. Both of these processes can be done using negative stain and cryo-EM.

2D electron crystallography is a technique that has been used to determine the structure of membrane proteins as early at 1975, as in the case of the 7-TM bacteriorhodopsin receptor (Henderson and Unwin, 1975). With this technique a purified, detergent solubilized membrane is slowly provided with lipids while the detergent is diluted away. This process drives protein packing in 2-dimensional crystals. The 2D crystals are then imaged by electron microscopy, which also allows for the recording of diffraction patterns. Merging phase information from the crystal image and amplitude information from diffraction patterns allows the calculation of three-dimensional maps.

The single-particle EM approach, where crystal formation is not necessary, depends on visualizing individual proteins or protein complexes at random orientations. Computational alignment and averaging of these individual projections allows the calculation of biological complexes. As supposed to x-ray crystallography and NMR where relatively large proteins (> ~100-150 kDa) become problematic for structural determination, protein complexes of larger size are usually more amenable to single particle EM. Additionally, symmetrical samples such as viruses that contain icosahedral symmetry, provide the benefit of signal redundancy that allows for improved averaging. The advantage from added symmetry has led to the calculation of virus 3D reconstructions to near atomic resolution (< 4 Å) using cryo electron microscopy (Grigorieff and Harrison, 2011). Since electron microscopy is the primary technique used in the studies subsequently covered, a more in depth overview will follow in the coming section.

1.8   Electron Microscopy

Sample preparation

Negative Stain Electron Microscopy

In negative stain electron microscopy, the sample is typically adsorbed on a carbon substrate that has been evaporated over a copper EM grid. The sample is fixed and embedded in a solution of heavy metal salt and subsequently dried on the EM grid. In areas where the protein sample is present, the staining solution is displaced (Figure 1-15, top). The intensity difference between stain-excluded areas containing the sample and the heavily stained background provide improved contrast of the sample under observation (Ohi et al., 2004). In addition, samples embedded in negative stain are more protected from radiation damage caused by the exposure to electrons. A major drawback to negative staining is that the samples will have a variable amount of collapse due to their adsorption on the carbon support and subsequent drying during staining. Also, the resolution of calculated 3D reconstructions is limited to the size of the dry stain grain (~20 Å). Nevertheless, negative staining is an excellent technique when looking at all proteins and particularly relatively small proteins (< ~200 kDa) due to the drastically improved contrast. In comparison to other structural techniques (x-ray crystallography and NMR), negative stain EM uses significantly lower concentrations of proteins (~200-500-fold less).

Negative Stain EM
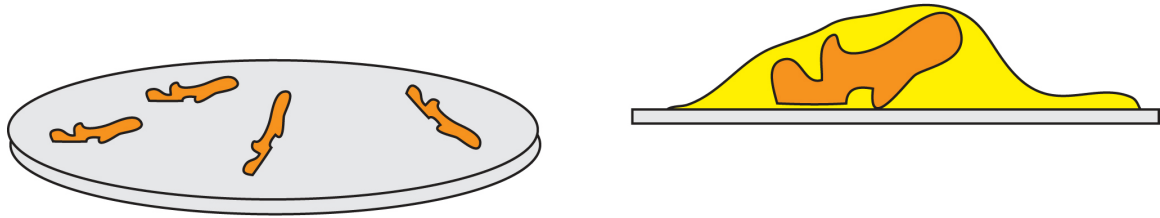
Cryo EM
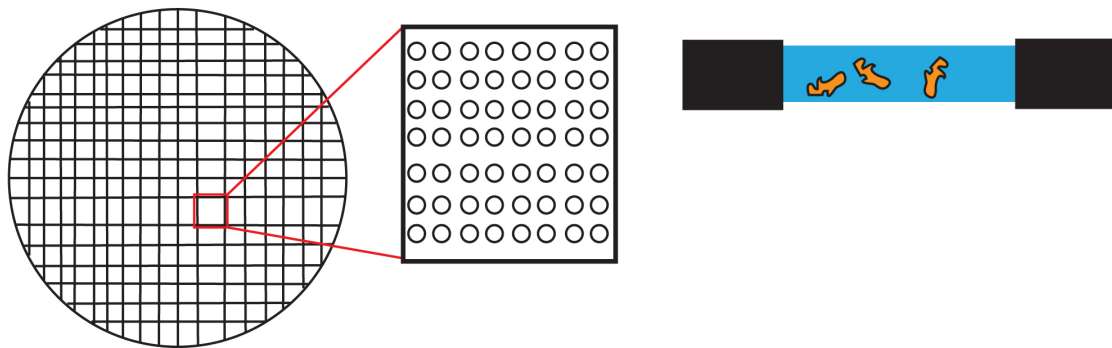
Figure 1-15. Negative stain and cryo-EM sample preparation.
Top, illustrates a continuous carbon grid. The sample is adsorbed to the carbon support.
The sample is subsequently coated in a heavy metal stain, typically uranyl formate or
uranyl acetate. Bottom, illustrates a holey carbon grid used for cryo-EM. The sample is
incubated on the holey carbon grid, excess sample is blotted away, and the grid is
plunged into liquid ethane that freezes the specimen in a later of vitreous ice.

Cryo Electron Microscopy

In cryo electron microscopy (cryo-EM), the sample is incubated on an EM grid composed
of perforated holes on the carbon support. Excess sample is blotted away and the grid is
plunged into liquid ethane at temperatures near -180º Celsius freezing the sample in a
layer of vitreous ice (Lepault et al., 1983) (Figure 1-15, bottom). In standard cryo-EM,
there is no addition of heavy metal salt solution to increase electron scattering and
improve contrast. Therefore, the contrast in cryo-EM is significantly reduced compared
to negative stain EM. However, in cryo-EM the sample is flash frozen in physiological
buffer and thus remains hydrated and devoid of artifacts such as flattening and collapse.
In addition, with cryo-EM preparation the sample mostly assumes many different

orientations in the vitreous ice, whereas in negative stain EM, the sample can take on a preferred orientation lying on a carbon support, and thus only provides a limited range of angular projections. On the other hand, in cryo-EM the sample is more susceptible to radiation damage, and therefore, lower electron doses have to be used during image acquisition. Due to the low signal to noise ratio and reduced contrast of cryo-EM, relatively larger (> ~200 kDa) proteins are suitable, which can limit the use of this technique on smaller proteins. A combination of negative stain and cryo-EM sample preparation, termed cryo-negative stain EM, can be used to improve this but is very tedious in practice (Bhella et al., 2004). Cryo-EM can inherently provide atomic resolution information, and in fact near atomic resolution (~4 Å) reconstructions have been achieved in cases of samples with high levels of symmetry (ie. viruses).

Image formation

Image formation in the electron microscope begins with the electrons emitted from an electron source, typically either tungsten or lanthanum hexaboride (LaB_6) filament that are either heated or combined with a field emission gun (FEG) to extract electrons. The electron beam is deflected through a single or set of condenser lenses to assure the beam is parallel (Figure 1-16A), and subsequently passes through the specimen, where the electrons are either scattered elastically, inelastically, or left unscattered. After passing through the specimen the electrons pass through the magnetic field of an objective lens, followed by an intermediate and projector lens that forms the final image. This image can then be viewed on a fluorescent screen, and is usually recorded using conventional film or a charge-coupled device (CCD) camera. On the EM image, the portion of the electrons that are scattered appear in a range of grey whereas the unscattered electrons appear as bright spots. The inelastically scattered electrons represent the background noise of an image (Figure 1-16B). In regions where the electrons are unable to pass through the sample, the image appears dark (Frank, 2006).
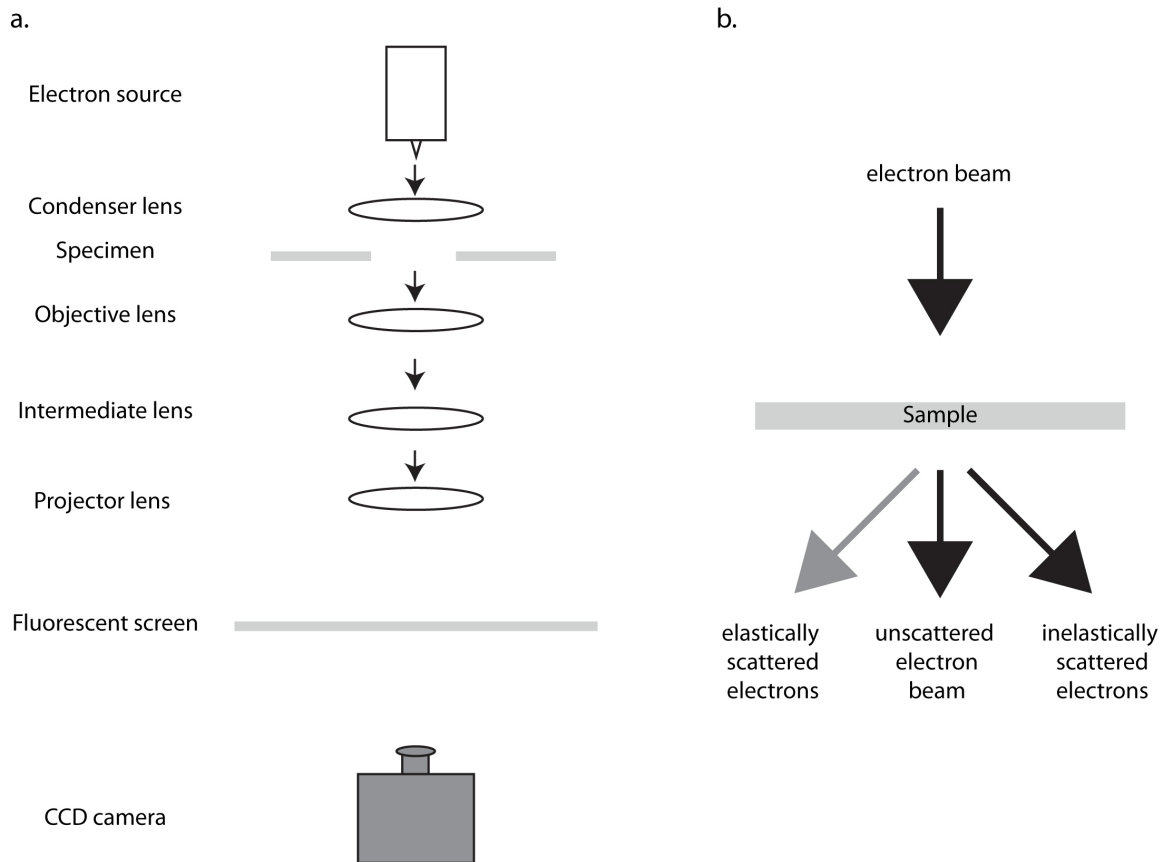
Figure 1-16. Microscope schematic and electron diffraction.
A) Illustrates the electron path through the microscope. B) Electron diffraction pattern through a given specimen.

The achievable resolution in transmission electron microscopy (TEM) depends on a number of factors including the lens aberrations. There are three primary sources of aberration: chromatic aberration, spherical aberration, and astigmatism. Chromatic aberration comes from the relationship between the energy of the electrons from the electron source being bent by the objective lens, with electrons of different energy being bent according to their respective energy. For example, electrons of lower energy will be bent more by the objective lens than electrons of higher energy. Spherical aberration is a result of the peripheral electrons being deflected more than electrons closer to the center of the specimen. Thinner specimens can improve on spherical aberration. Astigmatism is caused by the inability for the electromagnetic lens to have perfect cylindrical symmetry. Therefore astigmatism affects the capability to focus an image, but this can be

corrected easily by stigmators, a collection of small electromagnetic coils within the objective lens that apply additional magnetic fields to the electron beam (Frank, 2006).

Single Particle Analysis

The aim of single particle analysis is to characterize and determine the structure of macromolecules from images of individual particles. Images from TEM have a relatively low signal-to-noise ratio, whether by cryo-EM or negative stain EM. One of the goals for using single particle analysis is to average similar particle projections to increase the signal-to-noise ratio.   Single particle analysis uses a combination of different computational techniques to characterize different individual particle projections (hundreds to hundred of thousands), and allows the combination of these different views in order to reconstruct a 3D model of a macromolecule. The number of particles, the number of different orientations of projections, and the inherent homogeneity of a given specimen all affect the resolution and overall quality of the reconstruction.

Reference-free alignment and classification

The goal of alignment and classification is to separate different particle views and increase the signal-to-noise ratio of the individual particle projections. From each image collected on the EM, the individual particles are boxed out. These particles are then subjected to translational and rotational alignment using reference-free procedures.   In this reference-free approach, particle projections from the experimental data set are randomly selected as initial references and the remaining particles are aligned and grouped against the references using cross correlation criteria. Particle images within each group are then averaged, and these averages are used as new references for the next cycle. The process continues iteratively until there are no changes in the class averages (Ohi et al., 2004). A sample's heterogeneity can obviously affect the number of different conformations or projections a specimen can have. Therefore, the number of classes needed to separate a particular data can vary based on sample variation and the number of particle projections. EMAN (Ludtke et al., 1999), SPIDER (Frank et al., 1996), and IMAGIC (van Heel et al., 1996) are some of the different available software packages for applying this reference-free alignment and classification.

Contrast transfer function (CTF)

The transmission of the electron beam through the electron microscope distorts the structural information from the specimen and changes the phases and amplitude of the recorded electron waves. This electron-optical shortcoming depends on the microscope's operating voltage, objective lens spherical aberration coefficient, the defocus value used, and the spatial frequency. The contrast transfer theory explains how the Fourier transform of an object's image is related to the Fourier transform of an object's Coulombic potential. To correct for the modulation of the electron beam, the Fourier transform of an image must be multiplied by the contrast transfer function (CTF) (Wade, 1992).
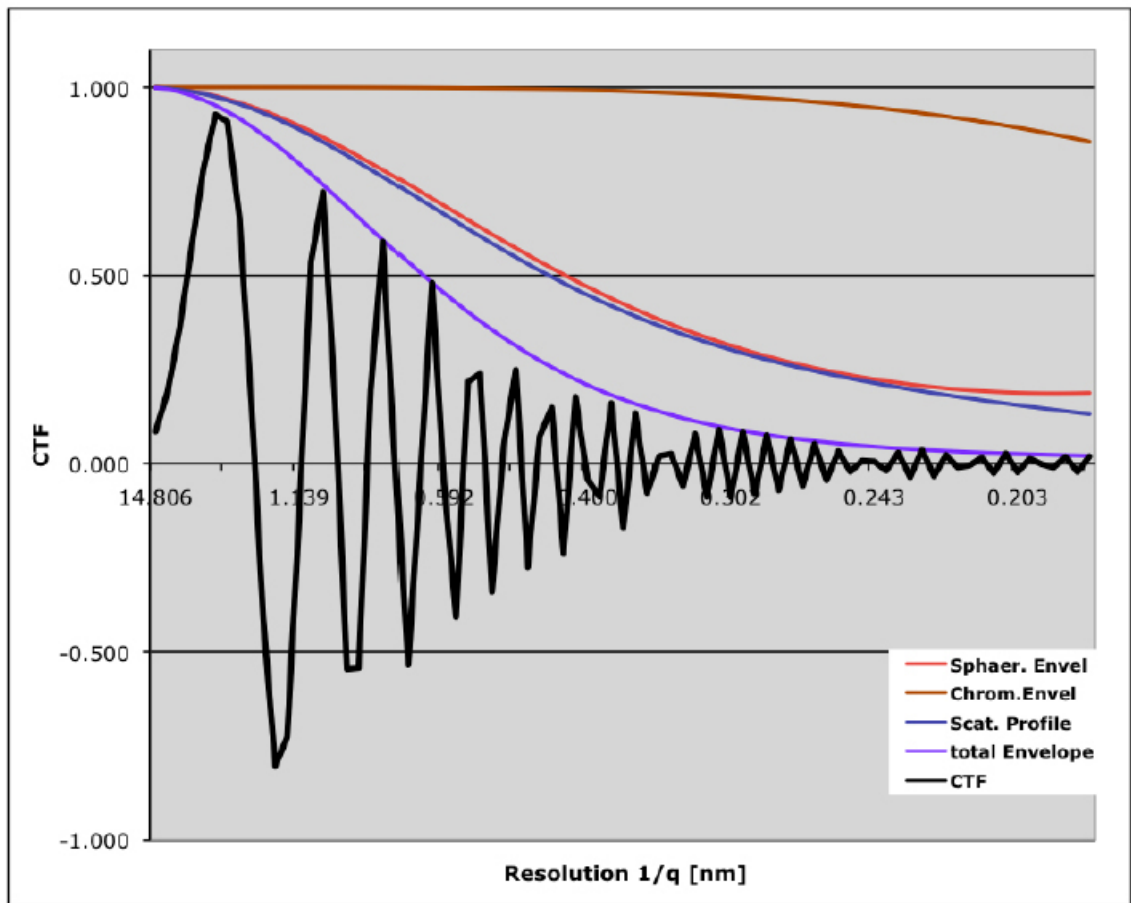
Cs=2 mm, Df=1000 nm   **V= 200 kV**

Figure 1-17. Contrast transfer function (CTF) curve.
A representative contrast transfer function curve for a transmission electron microscope (TEM) operating at 200 kV, with Cs=2, and a defocus value of 1 μm. The signal

amplitude is plotted against the resolution in inverse angstroms (x-axis). The experimental CTF curve oscillates from positive to negative values, while the signal gradually dampens with increasing resolution.

CTF is a sinusoidal oscillating function that becomes attenuated with increased spacial frequency. Figure 1-15 shows the CTF as a function of spatial frequency (resolution) for a 200 kV TEM and 1000nm defocus. The CTF transitions from positive to negative values (contrast) and in the regions it passes through zero, no contrast or transmittance is transferred. As a result, the resolution limit is the frequency where the CTF crosses the first zero. To overcome this limitation the EM images can be corrected by inverting the negative values of the CTF to positive, thereby extending the information to higher resolution ranges. 'Phase flipping' is the simplest way to computationally correct the CTF. This is done by determining the parameter values contributing to the CTF of each micrograph, and using those values to flip the negative regions to the positive phase (Zhu et al., 1997). Furthermore, the lack of information in spatial frequencies where CTF is zero can be compensated by collecting images with a range of different defocus values, thereby shifting the curve of the CTF sinusoidal function.
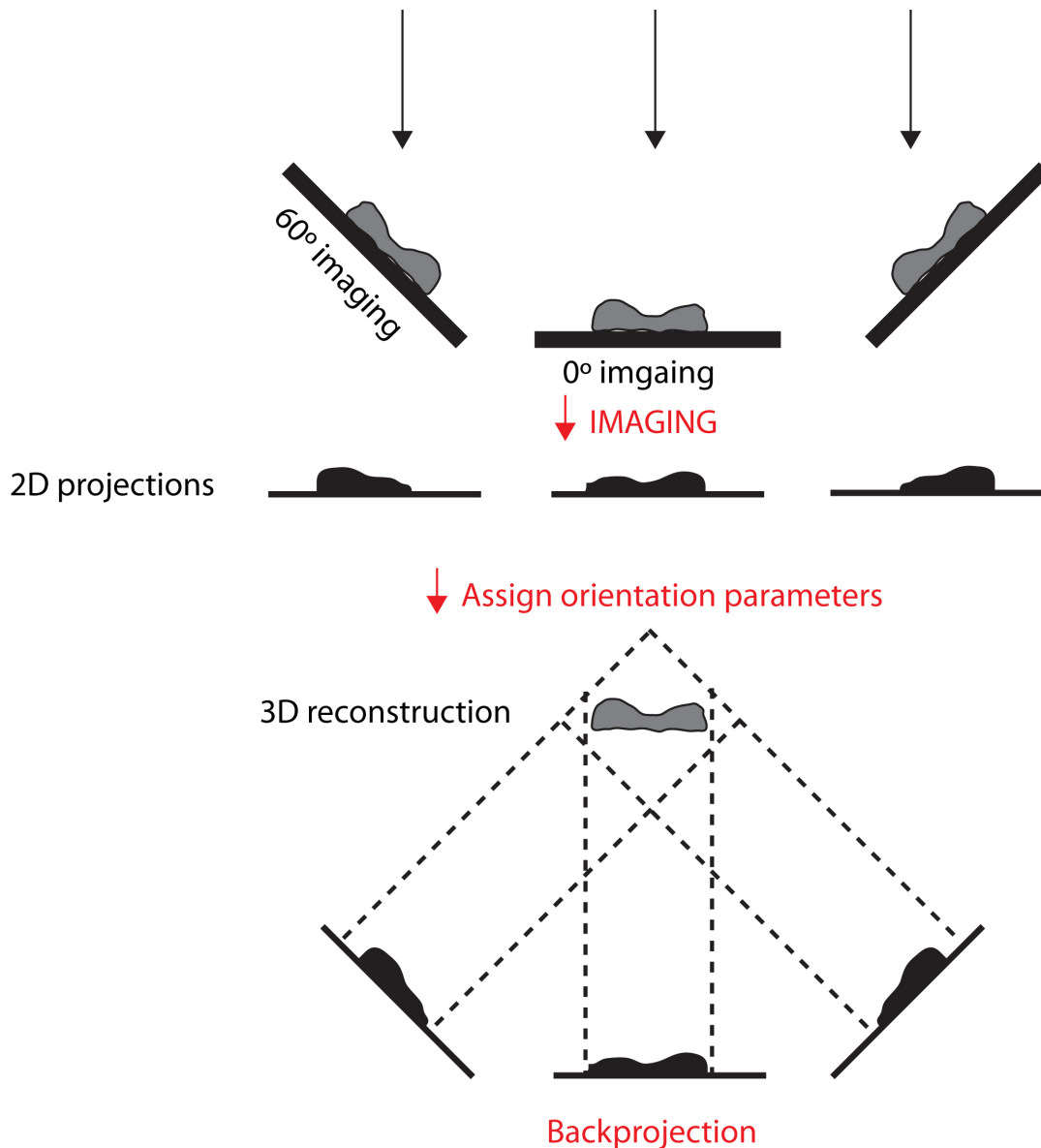
Figure 1-18. Tilt-pair imaging and backprojection.
The black arrows show the electron path through a specimen creating a 2-dimensional projection. Identical particles are imaged at 0° and 60°. The 0° particle projections are subjected to classification, and the orientation parameters are assigned to both the 0° and 60° particle projections. Using the known orientation parameters the projections can be backprojected to recreate the 3-dimensional features of the specimen.

Random conical tilt reconstruction

The random conical tilt (RCT) approach takes advantage of the preferred orientation of a specimen on the carbon support to generate a three dimensional object from two

dimensional projections (Radermacher et al., 1987). In this method, an image is taken at 0º of the goniometer and then the sample is tilted to an angle of ~60º (can be any angle from 50º - 70º), and an image is taken of the same area. The two images are digitized and put next to each other so that the same particles can be paired from both the 0º and 60º images. The untilted particles are subsequently subjected to alignment and classification, whereby the particles are grouped into different classes based on their orientation. In this step, the x-y shifts and in-plane rotation of the projections belonging to each class are defined and are also applicable to the tilted particle projections. Since the high tilt angle is consistent, the different projections from the tilted particles will form a cone, with the azimuthal angle of each tilted particle known from the in plane rotation of the untilted particles of its class. The tilted particle projections can then be back-projected, and a 3D reconstruction can be generated. This approach is advantageous because no initial model is necessary to generate a 3D reconstruction, and the volume generated can be used in subsequent negative stain and cryo-EM reconstructions.

Angular refinement using model-based projection matching

In angular refinement technique, 2D projections are generated from an initial model, also know as reprojections, that evenly cover all the angular space (Fuller et al., 1996; Zhu et al., 1997). The individual particle projections from the experimental dataset are directly compared to the theoretical reprojections, and assume the Euler angles of the reprojection they matched best by cross correlation (CC) value. Next, a new model is reconstructed from the experimental projections based on the obtained angles, and this new model is employed as the reference for the subsequent iteration. This process continues iteratively until the Fourier Shell correlation (FSC) curve converges. FSC measures the corresponding cross correlation values between the densities of two 3D volumes in Fourier space at each spatial frequency. The calculation to determine the final overall resolution of a 3D reconstruction is done by randomly splitting the experimental dataset to two half datasets and calculate two 3D reconstructions with each dataset. Comparison of the two reconstructions by FSC provides the resolution, typically obtained at FSC=0.5 (van Heel et al., 1996) (Figure 1-19).
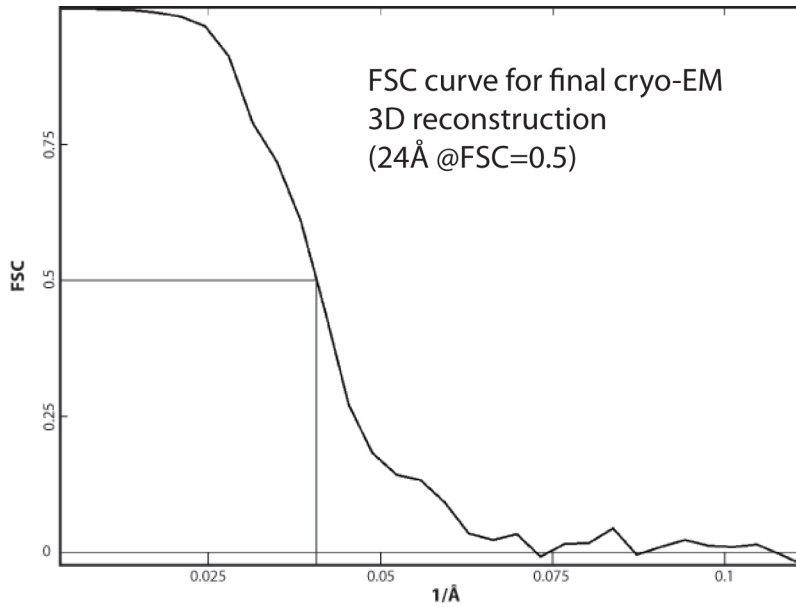
Figure 1-19. Fourier shell correlation (FSC).
Representative Fourier shell correlation (FSC) curve of the cryo-EM structure of the COMPASS methyltransferase complex. Indicated resolution of the map is indicated to be at an FSC = 0.5.

Multiple-reference angular refinement

Multiple-reference supervised classification is a process that quantitatively separates an experimental dataset that might involve two or more particle species or particle conformations in the relevant subsets (Brink et al., 2004; Menetret et al., 2005). The process can be followed by 3D reconstruction based on projection matching and run iteratively to improve the dataset partitioning and reconstructions. In the application of this method, two or more similar initial models with experimentally expected differences are used. Particle projections are compared with the reprojections of the reference models, and are assigned to an angular projection of a single model with which it obtains the highest cross correlation value. If the initial models are accurate, a single iteration is typically sufficient to properly sort the particle projections. The separate subset of particle projections can then be used to further refine each individual model. At this point differences in each respective model can be compared to determine a conformational shift, or presence or absence of a domain.

The majority of EM reconstructions are at low to intermediate resolution, therefore docking known high-resolution structures (typically from x-ray crystallography) can help interpret the EM 3D map. On the other hand, larger macromolecules can be problematic for high-resolution determination via crystallography or NMR, and these techniques often resort to characterizing smaller individual parts of a large protein or complex. Thus, the combination of electron microscopy to determine the architecture of a macromolecule at lower resolution, with X-ray or NMR to obtain high-resolution structures and dock them into the EM maps can give critical insight into the structure and function of large assemblies.

1.9   References

Azzi, M., Charest, P.G., Angers, S., Rousseau, G., Kohout, T., Bouvier, M., and Pineyro, G. (2003). Beta-arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. Proc Natl Acad Sci U S A *100*, 11406-11411.

Benovic, J.L., Kuhn, H., Weyand, I., Codina, J., Caron, M.G., and Lefkowitz, R.J. (1987). Functional desensitization of the isolated beta-adrenergic receptor by the beta-adrenergic receptor kinase: potential role of an analog of the retinal protein arrestin (48-kDa protein). Proc Natl Acad Sci U S A *84*, 8879-8882.

Bhella, D., Ralph, A., and Yeo, R.P. (2004). Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. J Mol Biol *340*, 319-331.

Brink, J., Ludtke, S.J., Kong, Y., Wakil, S.J., Ma, J., and Chiu, W. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. Structure *12*, 185-191.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K.*, et al.* (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science *318*, 1258-1265.

Frank, J. (2006). Three-dimensional electron microscopy of macromolecular assemblies visualization of biological molecules in their native state. Oxford ; New York:

Oxford University Press *2nd. edition*.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol *116*, 190-199.

Fredriksson, R., Lagerstrom, M.C., Lundin, L.G., and Schioth, H.B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol *63*, 1256-1272.

Freedman, N.J., and Lefkowitz, R.J. (1996). Desensitization of G protein-coupled receptors. Recent Prog Horm Res *51*, 319-351; discussion 352-313.

Fuller, S.D., Butcher, S.J., Cheng, R.H., and Baker, T.S. (1996). Three-dimensional reconstruction of icosahedral particles--the uncommon line. J Struct Biol *116*, 48-55.

Ghanouni, P., Gryczynski, Z., Steenhuis, J.J., Lee, T.W., Farrens, D.L., Lakowicz, J.R., and Kobilka, B.K. (2001a). Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor. J Biol Chem *276*, 24433-24436.

Ghanouni, P., Steenhuis, J.J., Farrens, D.L., and Kobilka, B.K. (2001b). Agonist-induced conformational changes in the G-protein-coupling domain of the beta 2 adrenergic receptor. Proc Natl Acad Sci U S A *98*, 5997-6002.

Grigorieff, N., and Harrison, S.C. (2011). Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr Opin Struct Biol *21*, 265-273.

Gurevich, V.V., and Benovic, J.L. (1993). Visual arrestin interaction with rhodopsin. Sequential multisite binding ensures strict selectivity toward light-activated phosphorylated rhodopsin. J Biol Chem *268*, 11628-11638.

Henderson, R., and Unwin, P.N. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. Nature *257*, 28-32.

Kim, J., Kim, J.A., McGinty, R.K., Nguyen, U.T., Muir, T.W., Allis, C.D., and Roeder, R.G. (2013). The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. Mol Cell *49*, 1121-1133.

Kobilka, B.K. (2007). G protein coupled receptor structure and activation. Biochim Biophys Acta *1768*, 794-807.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Krogan, N.J., Dover, J., Khorrami, S., Greenblatt, J.F., Schneider, J., Johnston, M., and Shilatifard, A. (2002). COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J Biol Chem *277*, 10753-10755.

Kuhn, H. (1978). Light-regulated binding of rhodopsin kinase and other proteins to cattle photoreceptor membranes. Biochemistry *17*, 4389-4395.

Landau, E.M., and Rosenbusch, J.P. (1996). Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. Proc Natl Acad Sci U S A *93*, 14532-14535.

Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084-1096.

Lepault, J., Booy, F.P., and Dubochet, J. (1983). Electron microscopy of frozen biological suspensions. J Microsc *129*, 89-102.

Lohse, M.J., Benovic, J.L., Codina, J., Caron, M.G., and Lefkowitz, R.J. (1990). beta-Arrestin: a protein that regulates beta-adrenergic receptor function. Science *248*, 1547-1550.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol *128*, 82-97.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luttrell, L.M., Ferguson, S.S., Daaka, Y., Miller, W.E., Maudsley, S., Della Rocca, G.J., Lin, F., Kawakatsu, H., Owada, K., Luttrell, D.K.*, et al.* (1999). Beta-arrestin-dependent formation of beta2 adrenergic receptor-Src protein kinase complexes. Science *283*, 655-661.

Luttrell, L.M., and Lefkowitz, R.J. (2002). The role of beta-arrestins in the termination and transduction of G-protein-coupled receptor signals. J Cell Sci *115*, 455-465.

Menetret, J.F., Hegde, R.S., Heinrich, S.U., Chandramouli, P., Ludtke, S.J., Rapoport, T.A., and Akey, C.W. (2005). Architecture of the ribosome-channel complex derived from native membranes. J Mol Biol *348*, 445-457.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A *98*, 12902-12907.

Montaville, P., and Jamin, N. (2010). Determination of membrane protein structures using solution and solid-state NMR. Methods Mol Biol *654*, 261-282.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol Proced Online *6*, 23-34.

Oldham, W.M., and Hamm, H.E. (2008). Heterotrimeric G protein activation by G-protein-coupled receptors. Nat Rev Mol Cell Biol *9*, 60-71.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J Microsc *146*, 113-136.

Rajagopal, S., Rajagopal, K., and Lefkowitz, R.J. (2010). Teaching old receptors new tricks: biasing seven-transmembrane receptors. Nat Rev Drug Discov *9*, 373-386.

Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F.*, et al.* (2007). Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature *450*, 383-387.

Rasmussen, S.G., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D.*, et al.* (2011). Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature *477*, 549-555.

Raunser, S., and Walz, T. (2009). Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. Annu Rev Biophys *38*, 89-105.

Rosenbaum, D.M., Zhang, C., Lyons, J.A., Holl, R., Aragao, D., Arlow, D.H., Rasmussen, S.G., Choi, H.J., Devree, B.T., Sunahara, R.K.*, et al.* (2011). Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. Nature *469*, 236-240.

Schneider, J., Wood, A., Lee, J.S., Schuster, R., Dueker, J., Maguire, C., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2005). Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. Mol Cell *19*, 849-856.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem *75*, 243-269.

Shilatifard, A. (2008). Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr Opin Cell Biol *20*, 341-348.

Southall, S.M., Wong, P.S., Odho, Z., Roe, S.M., and Wilson, J.R. (2009). Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol Cell *33*, 181-191.

Strader, C.D., Fong, T.M., Tota, M.R., Underwood, D., and Dixon, R.A. (1994). Structure and function of G protein-coupled receptors. Annu Rev Biochem *63*, 101-132.

Strader, C.D., Sigal, I.S., and Dixon, R.A. (1989). Genetic approaches to the determination of structure-function relationships of G protein-coupled receptors. Trends Pharmacol Sci *Suppl*, 26-30.

Stryer, L., and Bourne, H.R. (1986). G proteins: a family of signal transducers. Annu Rev Cell Biol *2*, 391-419.

Takahashi, Y.H., Lee, J.S., Swanson, S.K., Saraf, A., Florens, L., Washburn, M.P., Trievel, R.C., and Shilatifard, A. (2009). Regulation of H3K4 trimethylation via Cps40 (Spp1) of COMPASS is monoubiquitination independent: implication for a Phe/Tyr switch by the catalytic domain of Set1. Mol Cell Biol *29*, 3478-3486.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J Struct Biol *116*, 17-24.

Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., and Babu, M.M. (2013). Molecular signatures of G-protein-coupled receptors. Nature *494*, 185-194.

Wade, R.H. (1992). A Brief Look at Imaging and Contrast Transfer. Ultramicroscopy *46*, 145-146.

Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., and Sprang, S.R. (1995). The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. Cell *83*, 1047-1058.

Wieland, K., Zuurmond, H.M., Krasel, C., Ijzerman, A.P., and Lohse, M.J. (1996). Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. Proc Natl Acad Sci U S A *93*, 9276-9281.

Xiao, R.P., Cheng, H., Zhou, Y.Y., Kuschel, M., and Lakatta, E.G. (1999). Recent advances in cardiac beta(2)-adrenergic signal transduction. Circ Res *85*, 1092-1100.

Yao, X., Parnot, C., Deupi, X., Ratnala, V.R., Swaminath, G., Farrens, D., and Kobilka, B. (2006). Coupling ligand structure to specific conformational switches in the beta2-adrenoceptor. Nat Chem Biol *2*, 417-422.

Zhang, Y. (2003). Transcriptional regulation by histone ubiquitination and deubiquitination. Genes Dev *17*, 2733-2740.

Zhu, J., Penczek, P.A., Schroder, R., and Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. J Struct Biol *118*, 197-219.

Chapter 2
Structural flexibility of the Gαs
α-helical domain in the β₂-adrenoceptor Gs complex

2.1   Abstract

The active state complex between an agonist bound receptor and a guanine nucleotide-free G protein represents the fundamental signaling assembly for the majority of hormone and neurotransmitter signaling. We applied single particle electron microscopy (EM) analysis to examine the architecture of agonist occupied β₂-adrenoceptor (β₂AR) in complex with the heterotrimeric G protein Gs (Gαsβγ). EM 2D averages and 3D reconstructions of the detergent-solubilized complex reveal an overall architecture that is in very good agreement with the crystal structure of the active state ternary complex. Strikingly however, the α-helical (AH) domain of Gαs appears highly flexible in the absence of nucleotide. In contrast, the presence of the pyrophosphate mimic foscarnet (phosphonoformate), and also the presence of GDP favor the stabilization of the AH domain on the Ras-like domain of Gαs. Molecular modeling of the AH domain in the 3D EM maps suggests that in its stabilized form it assumes a conformation reminiscent to the one observed in the crystal structure of Gαs-GTPγS. These data argue that the AH domain undergoes a nucleotide-dependent transition from a flexible to a conformationally stabilized state.

2.2   Introduction

The majority of hormones and neurotransmitters communicate information to cells via G protein coupled receptors (GPCRs), which instigate intracellular signaling by activating their cognate heterotrimeric G proteins on the cytoplasmic side. GPCRs constitute the largest family of membrane proteins and play essential roles in regulating every aspect of normal physiology, thereby representing major pharmacological targets. Despite a wealth of biochemical and biophysical studies on inactive and active conformations of several

heterotrimeric G proteins, the molecular underpinnings of G protein activation remain elusive. The β₂-adrenergic receptor (β₂AR) and its complex with heterotrimeric stimulatory G protein Gs (Gαsβγ) represent an ideal model system for the large family of GPCRs activated by diffusible ligands. Agonist binding to the β₂AR promotes interactions with GDP-bound Gsαβγ heterotrimer leading to the exchange of GDP for GTP, and the functional dissociation of Gs into Gα-GTP and Gβγ subunits. To examine the architecture of agonist occupied β₂AR in complex with Gαsβγ under different conditions, we employed electron microscopy (EM) and single particle analysis. Due to the limited size of the protein complex (~148 kDa) we visualized specimens embedded in negative stain, which provides sufficient contrast from relatively small protein assemblies (Ohi et al., 2004). This approach allowed us to obtain 2D projection averages and 3D reconstructions that provided new insights into dynamic features of the β₂AR-Gs complex, and helped guide a successful approach to crystallize the complex enabling a high-resolution structure (Rasmussen et al., 2011).
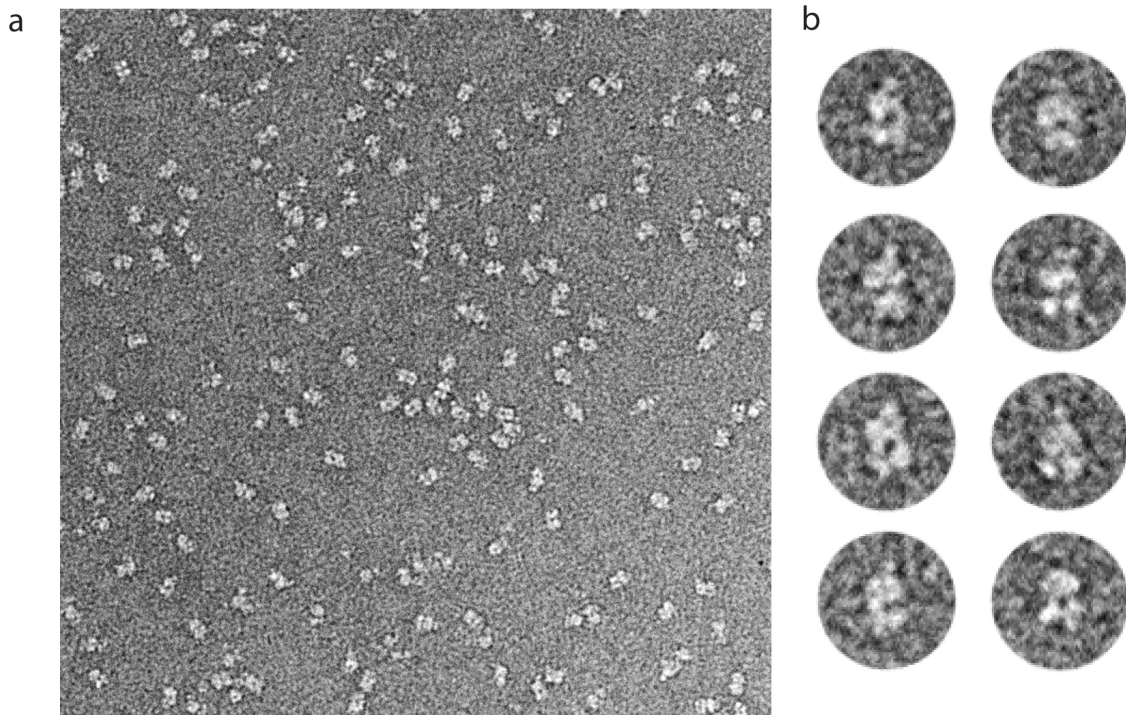
Figure 2-1. | Raw EM image of negative stained nucleotide-free T4L-β₂AR-Gs complex.
a, Raw image of negative stained nucleotide-free T4L-β₂AR-Gs complex. b, Excised particles of nucleotide-free T4L-β₂AR-Gs complex.

Figure 2-2. Initial and secondary 2D classification of nucleotide-free T4L-β₂AR-Gs complex.

a) 17,205 particles classified into 200 classes. b) 4,378 particles of nucleotide-free T4L-β₂AR-Gs complex were classified into 50 classes. Class averages with the α-helical (AH) domain observed to be stabilized on the Ras-like domain of Gαs are marked with an

orange square. Class averages where the AH domain projection profile on the Ras-like domain is not visible are marked with a green circle. Scale bar = 10nm.

2.3 Results

Negative Stain EM of nucleotide-free β₂AR-Gs

In a first step, we sought to examine the architecture of complexes in the nucleotide-free state of Gαs. Prior to coupling with an agonist-bound receptor, the nucleotide binding pocket of the α subunit of the Gαsβγ heterotrimer is occupied by GDP. Upon forming a complex with the β₂AR, GDP dissociates, and the resulting nucleotide-free β₂AR-Gs complex is highly stable (Rasmussen et al., 2011). EM visualization of the nucleotide-free complex showed a monodisperse particle population (Figure 2-1). Reference-free alignment and classification of ~17,000 particle projections revealed characteristic class averages with an overall density that is in very good agreement with the crystal structure of the complex (Rasmussen et al., 2011). Due to its shape, the complex adsorbs on the carbon support with small variations (+/-20º) of mainly two diametrically opposite preferred orientations that generate practically identical, mirror-related 2D projections (Figure 2-2, 2-3a).

The distinct features of the class averages in these preferred orientations allowed us to assign the negative stain projection profiles from specific components of the complex (Figure 2-3a & b). A central oval density represents the β₂AR in a detergent micelle, with a small protruding density corresponding to T4 lysozyme (T4L) that replaces the unstructured extracellular N-terminus of the receptor and serves as an orienting landmark. This interpretation was confirmed by EM analysis of complexes lacking T4L (Figure 2-3c & d). Some class averages of the T4L-β₂AR-Gs complex do not reveal a density corresponding to the T4L. Besides the presence of a relatively flexible linker connecting T4L and the β₂AR, this effect is mostly because T4L lies at an angle to the longitudinal axis of the complex, as shown in the X-ray structure (Rasmussen et al., 2011). Due to this geometry, even a 10-degree variation in the way the particle adsorbs on the carbon support drastically reduces the visibility of the T4L projection profile, as demonstrated by projection simulation experiments (Figure 2-4). Thus, the visibility of the T4L projection

profile is very sensitive to even limited out of plane particle tilting, e.g. due to particle "rock" and "roll" or because of variations in the flatness of the carbon support.
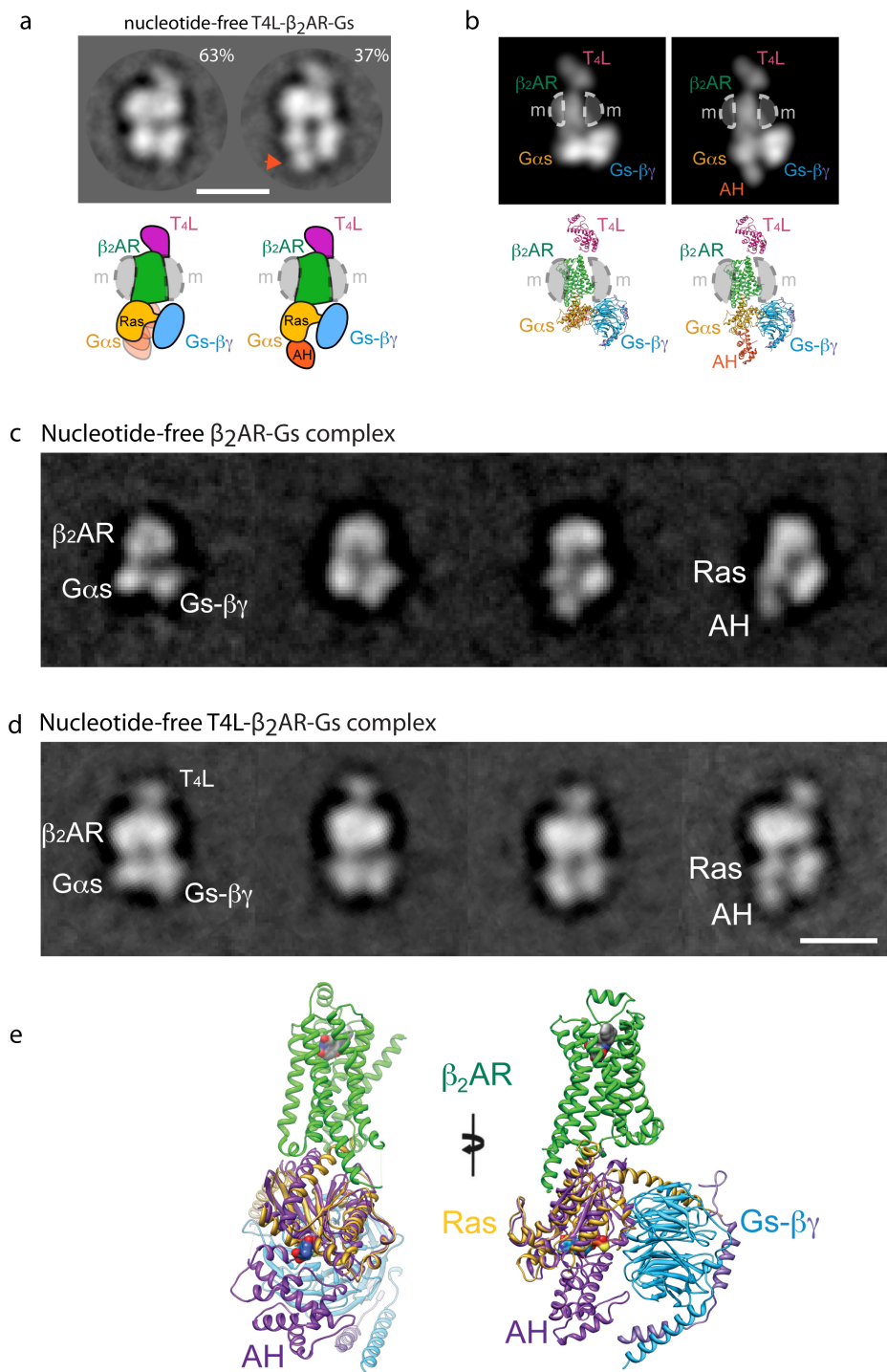
Figure 2-3. 2D class averages and models for alpha helical domain flexibility and T4L comparison.

a) Representative EM class averages of the nucleotide-free complex with the projection profile of the a-helical (AH) domain not visible (left panel), or visible on the Ras domain (right panel, AH indicated by arrow). The cartoon models represent the conformations reflected by the EM averages, with the one on the left depicting the variable positioning of the AH domain, suggesting flexibility or multiple conformations (the position of the detergent micelle is indicated by gray shaded arcs and labeled with "m"). b) Reprojections (top) of the crystal structure (Rasmussen et al., 2011) (bottom) in the same orientation as (a) reveal the identity of each EM density component. The crystal structure on the right panel shows the AH domain in the same position (relative to the Ras-like domain) as the one determined in the crystal structure of Gαs-GTPγS alone (Sunahara et al., 1997). c) Representative class averages of nucleotide-free β$_2$AR-Gs complex with no T4 lysozyme fusion at the N-terminus of the β$_2$AR receptor. The detergent used in this preparation is a variant (MNG 28 analogue-1) of maltose neopentyl glycol [NG-310; Chae et al. Nature Methods 2010 Dec;7(12):1003-8]. The class average on the right shows the α-helical (AH) domain stabilized on the Ras-like domain of Gαs. d) Representative class averages of nucleotide-free T4L-β$_2$AR-Gs complex with T4 lysozyme fusion on the N-terminus of the β$_2$AR receptor. The detergent used in this preparation is maltose neopentyl glycol (NG-310). The class average on the right shows the AH domain stabilized on the Ras-like domain of Gαs. Note that the difference in the size of the detergent micelle around β$_2$AR between the two complexes is due to the different variants of detergents. The scale bar corresponds to 10 nm. e) Overlay of the crystal structures of the β$_2$AR-Gs complex and Gαs-GTPγS (1AZT). The Ras-like domain of the β$_2$AR-Gs complex is shown in orange, and the Gαs-GTPγS structure is shown in purple. The α-helical (AH) domain of the Gαs-GTPγS structure is stabilized on the Ras-like domain in the presence of GTPγS, similarly to the conformation observed by single-particle EM.
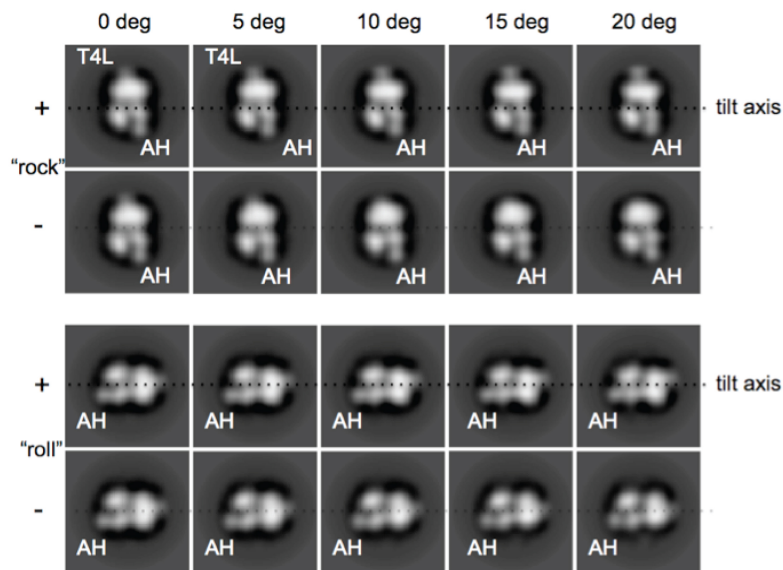
Figure 2-4. Simulated projections of our 3D reconstruction displaying an ordered AH domain.

Simulated projections of our 3D reconsturction diplaying an ordered AH domain. The reprojections are up to +/- 20 degrees around the preferred particle orientation (40 degrees total) and show that the projection profile of the ordered AH domain remains fully visible. In contrast, the separate projection profile of the T4L domain disappears at just 10 degree particle "rocking".

Since we observe a single density corresponding to T4L, the detergent micelle contains only a single copy of the β_2AR, in agreement with the crystal structure. Therefore, the significant additional density around the receptor stems from the large micelle formed by the detergent (Rubinstein, 2007). Diametrically opposite to the T4L domain, two main interacting densities representing the Gs trimer appear in close proximity to the receptor on its intracellular surface. One of the two domains appears to extensively interact with the receptor density, suggesting it corresponds to the Ras-like domain of Gαs, while its neighboring domain has a profile consistent with the side view of G$\beta\gamma$. Interestingly however, several class averages revealed an additional small globular density bound on the Ras-like domain of Gαs (Figure 2-3a, right panel). In this location, the additional density could only be attributed to the α-helical (AH) domain of Gαs, occupying a position expected from the crystal structure of Gαs-GTPγS alone (Sunahara et al., 1997) and the structure of the Gi heterotrimer (Wall et al., 1995) (Figure 2-3b, 2-3e), but in entirely different location from that observed in the crystal structure of the β_2AR-Gs complex (Rasmussen et al., 2011).

Conformations of alpha helical domain

To assess the fraction of particles displaying the AH domain in this location we selected and classified only projections clearly displaying the profiles of Ras-like, G$\beta\gamma$, β_2AR, and T4L domain densities in the same position, thereby restricting the range of particle projection orientations. The classification revealed that the AH domain was ordered on the Ras-like domain in ~35% of the particles, while in most other particle projections this density was absent (Figure 2-3a). It should also be noted that in contrast to the T4L domain, the projection profile of the AH domain in this position is not sensitive to the relatively limited out of plane tilts (+/- 20º) of the preferred particle orientation on the carbon support (Figure 2-4). This EM analysis provided the first evidence for a high degree of mobility of the AH domain relative to the Ras domain in the nucleotide-free

β$_2$AR-Gs complex. Furthermore, the structural heterogeneity observed provided insights to the challenges in obtaining three-dimensional crystals of the complex.

Nanobody binding to β$_2$AR-Gs

To promote complex stabilization for high-resolution structural studies we generated and screened llama antibodies (nanobodies) to the purified complex (Rasmussen et al., 2011). Nanobodies are small (approximately 15 kDa), clonable variable domains of a heavy chain only antibody, obtained by immunizing a llama with purified detergent-solubilized β$_2$AR-Gs complex stabilized with a short homobifunctional crosslinker (Rasmussen et al., 2011). By screening samples with negative stain EM we identified two nanobodies (Nb35 and Nb37) that bound to the complex, but not the receptor alone. Class averages of particles incubated with Nb35 indicated a homogeneous protein complex displaying increased density between Gαs and Gβγ. Even though the nanobody projection profile was not clearly distinguished in the preferred particle orientations, its presence appeared to enhance the uniformity in the disposition of Gαs-ras and Gβγ domains. The use of Nb35 indeed allowed us to obtain the crystal structure of the T4L-β$_2$AR-Gs complex, which showed that the nanobody binds at the interface of the Gαs-Ras and Gβγ. In this location, Nb35 would not be predicted to interact with or stabilize the AH domain (Rasmussen et al., 2011). Accordingly, the classification of Nb35-bound complexes revealed a similar distribution of particles with an ordered AH domain on the Ras-like domain as in the absence of Nb35 (Figure 2-5).

c

d

34% ■ stabilized AH domain on the Ras-like domain of Gαs
66% ● non-stabilized AH domain
★ undefined
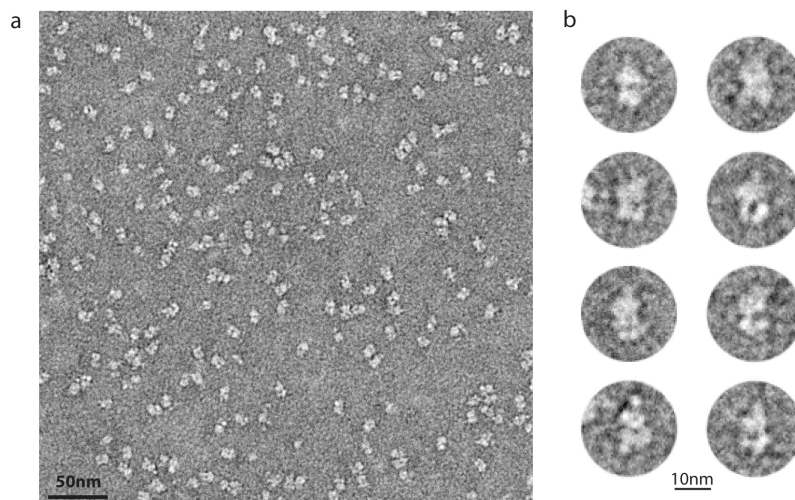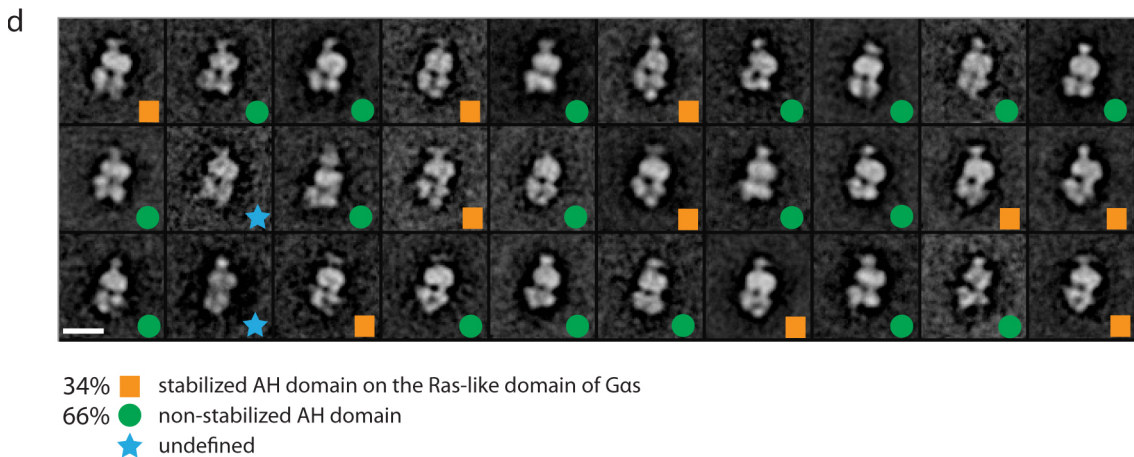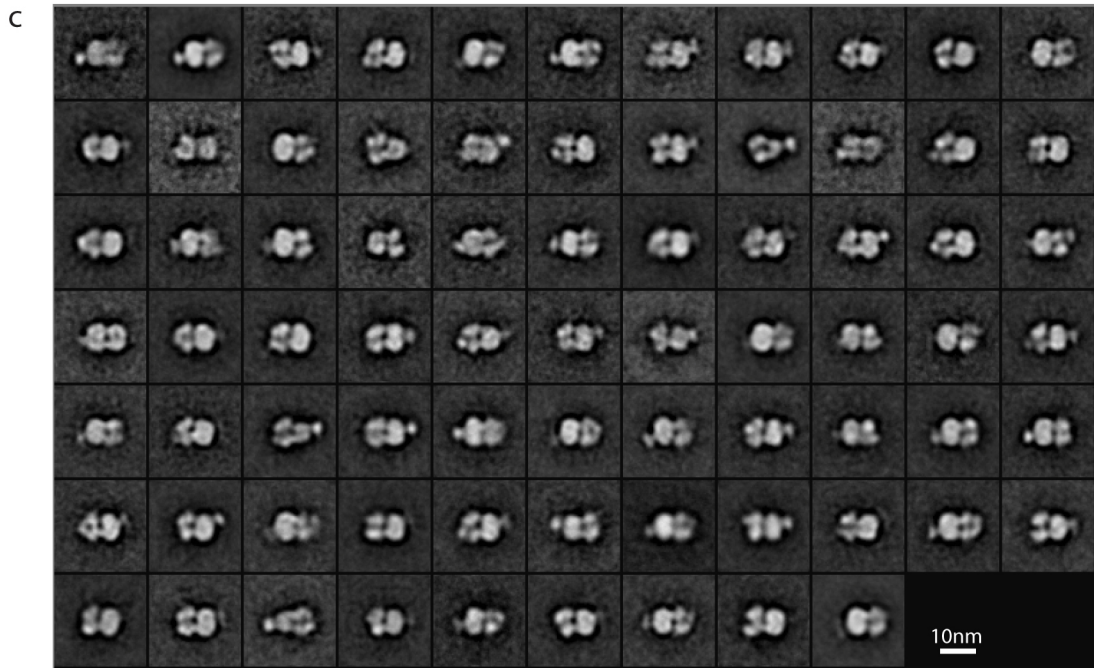
Figure 2-5. Raw image and 2D classification of negative stained T4L-β₂AR-Gs complex with bound camelid antibody Nb35.

a) Raw image of negative stained T4L-β_2AR-Gs complex with bound camelid antibody (nanobody; Nb35). b) Excised particles of T4L-β_2AR-Gs complex with bound Nb35. c) 2D classification 6,514 particles classified into 75 classes. d) Secondary 2D classification 1,151 particles of nucleotide-free T4L-β_2AR-Gs complex with bound camelid antibody Nb35 were classified into 30 classes. Class averages with the α-helical (AH) domain observed to be stabilized on the Ras-like domain of Gαs are marked with an orange square. Class averages where the AH domain projection profile on the Ras-like domain is not visible are marked with a green circle. Class averages of insufficient quality to categorize are marked with a cyan star. Scale bar = 10nm.

In contrast to Nb35, Nb37 appears bound directly to the AH domain (as also determined by DXMS, Figure 2-6) and could be distinguished in single particle EM class averages of the β2AR-Gs complex as an extension of the AH domain. Using Nb37 as a domain marker, allowed us to track the variable positioning of the small α-helical region of Gαs (Figure 2-7). 2D class averages from this preparation reveal an enhanced and elongated density adopting different orientations around the Ras-like domain, ranging from close proximity to the Gβγ module, to extending much further out of the complex in the opposite direction (Figure 2-7b). Collectively, these findings suggest that in the absence of nucleotide, the AH domain is flexible, thereby sampling different positions around the Ras-like domain. Deuterium exchange studies are consistent with a dynamic interface between the Gαs Ras and AH domains (Chung et al., 2011). Therefore, the unexpected position of the "open" AH domain in the crystal structure (Rasmussen et al., 2011) represents just one of the possible conformations.
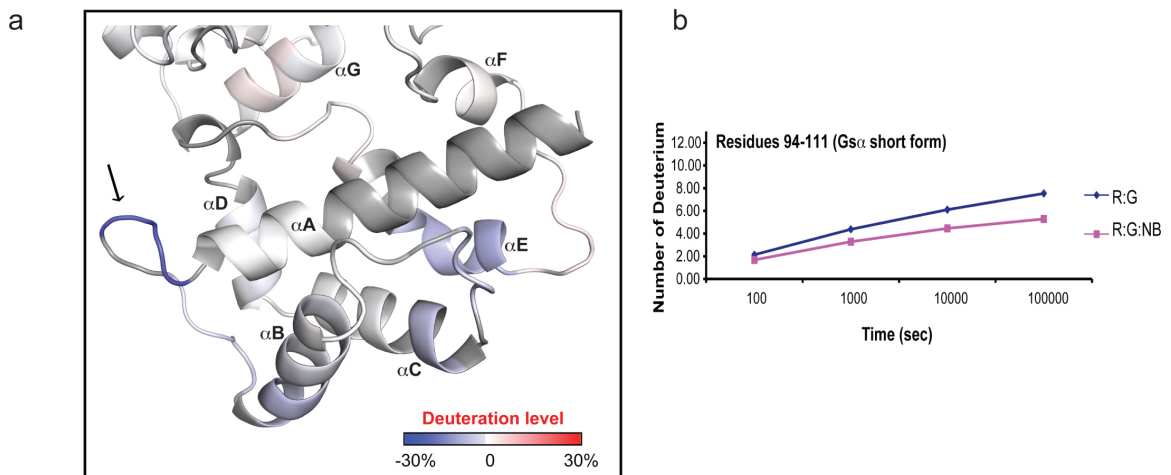
Figure 2-6. Effect of Nb37 binding to deuterium exchange profile of helical domain of Gαs unit.
a) The difference of deuterium exchange percentage of Gαs subunit after Nb37 binding. Decreased deuterium exchange upon Nb37 binding was marked in blue, and increased deuterium binding exchange was marked in red. b) the number of deuterium incorporated in the region indicated with arrow (a) was plotted in the absence (blue) and in the presence (pink) of Nb37. The scale of y-axis is the maximum exchangeable amides. (Figure courtesy of Sunahara lab)
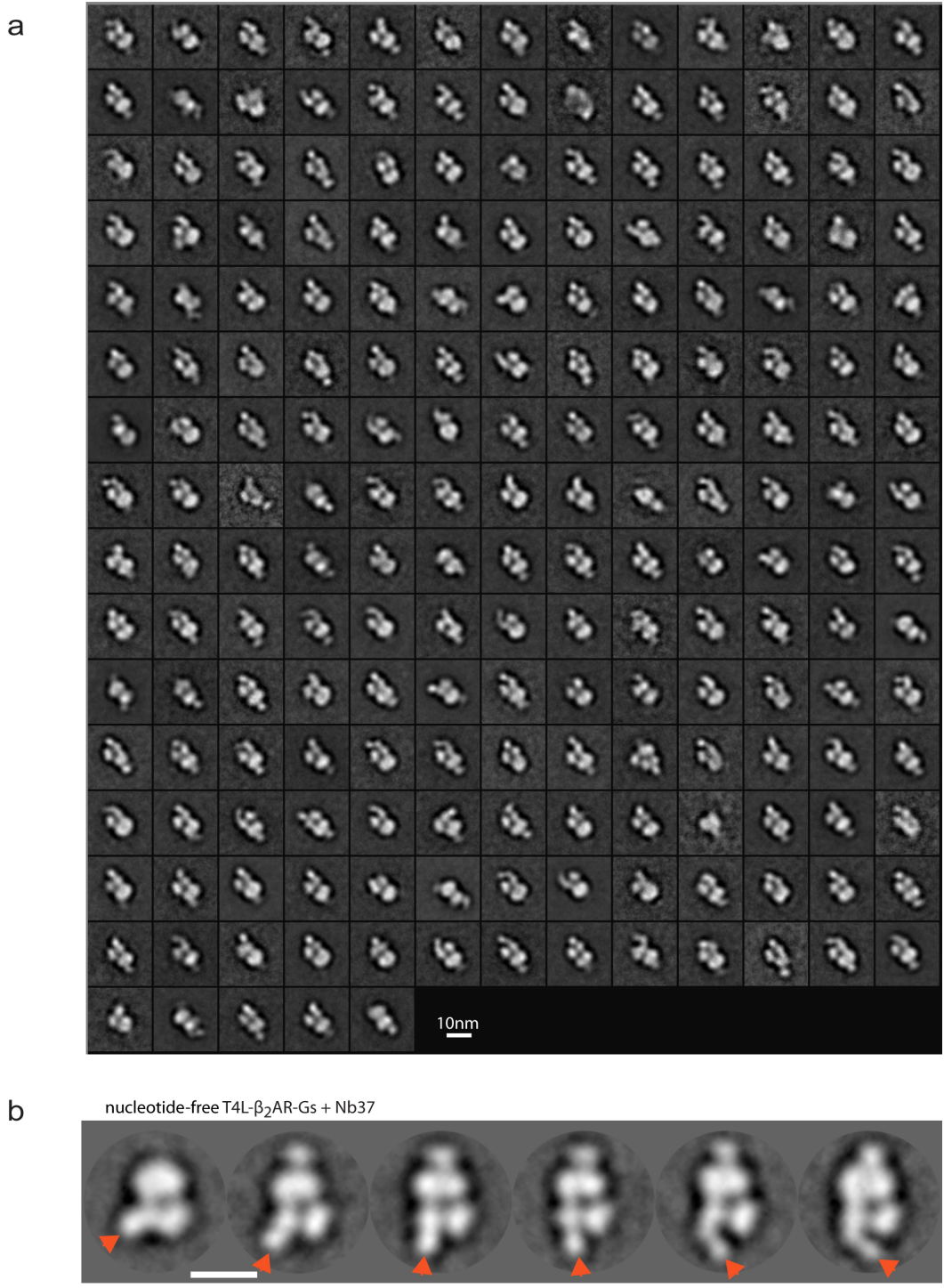
Figure 2-7. 2D classification of T4L-β₂AR-Gs complex with bound camelid antibody Nb37.

a) 14,704 particles classified into 200 classes. b) Representative class averages of nucleotide-free complex with nanobody Nb37 bound on the AH domain (arrows) reveal its flexibility. Scale bar = 10 nm.

To obtain a more detailed view of the complex architecture, we used the random conical-tilt approach (Radermacher et al., 1987) to calculate initial 3D reconstructions of complexes with and without ordered AH domain on the Ras-like domain (Figure 2-8). These initial 3D models were subsequently employed for multireference supervised alignment (Brink et al., 2004; Menetret et al., 2005) to separate particle projections from our entire dataset according to the AH positioning. This approach allowed us to obtain quality 3D reconstructions from particle projections with and without density corresponding to AH domain on the Ras-like domain.
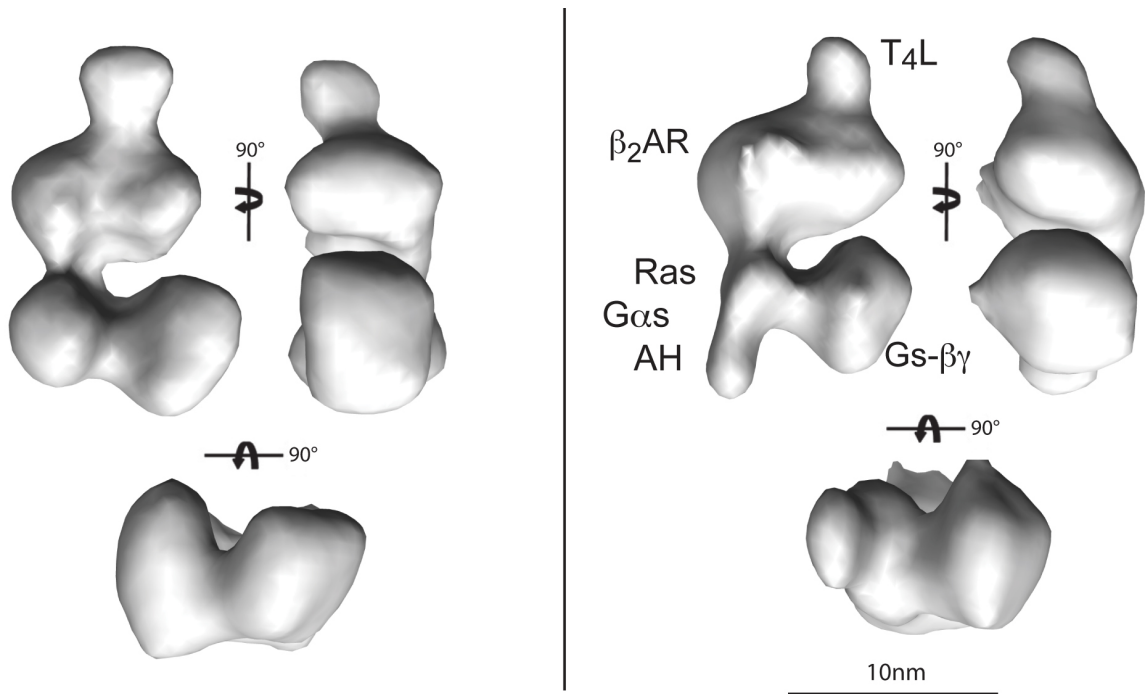
Figure 2-8. Initial models of T4L-β₂AR-Gs complex.
Initial models calculated from individual particle classes based on the random conical-tilt method. Left, an initial model of T4L-β₂AR-Gs complex with no α-helical (AH) domain visible on the Ras-like domain of Gαs. Right, an initial model with the AH domain visible on the Ras-like domain of Gαs.

The reconstructions are in excellent agreement with the corresponding 2D averages (Figure 2-9a, 2-10). In 3D reconstructions of particles where density for AH domain is observed, its orientation relative to the Ras-like domain appears to be similar to that found in the crystal structure of Gαs-GTPγS (Sunahara et al., 1997) (Figure 2-9a, Figure

2-3e). In addition, we obtained 3D reconstructions of nucleotide-free complexes with bound Nb37 marking the positioning of the AH domain. The 3D maps clearly reveal that the Nb37-enhanced density of the AH domain can adopt different conformations around the Ras-like domain, indicative of a relative flexibility in the interaction between the two domains (Figure 2-9b, Figures 2-11 & 2-12). This variability in the 3D conformation of the AH domain or the AH/Nb37 module around the Ras-like domain was further confirmed by cross-validating 3D reconstructions (Figure 2-13).
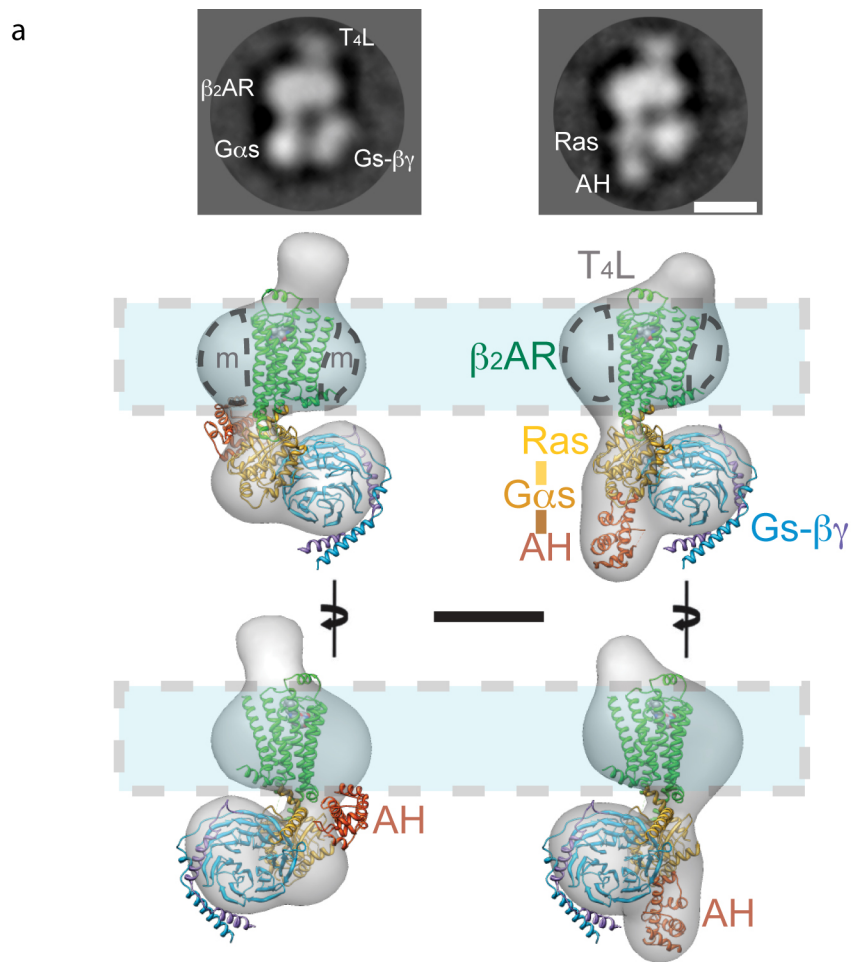
b
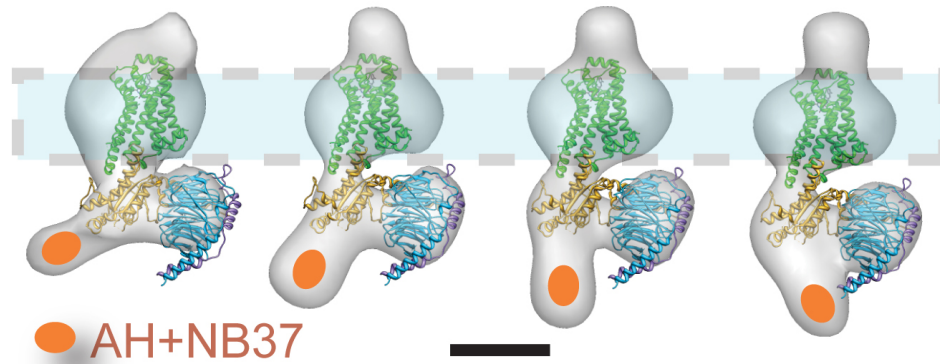
Figure 2-9. 3D reconstructions of the T4L-β₂AR-Gs complex in the nucleotide-free state.
Representative class averages and corresponding 3D reconstructions of particles in each category show the variability in the positioning of the AH domain in the nucleotide-free complex. In the reconstruction to the left, the AH domain (orange ribbon) is shown in the same position as found in the docked crystal structure (Rasmussen et al., 2011). Absence of sufficient density to accommodate this domain indicates that its position is highly variable in this particle population. In the reconstruction to the right, the AH domain is modeled within the available EM density right below the Ras-like domain of Gαs, as also suggested by the 2D averages. b) 3D reconstructions of distinct conformations of nucleotide-free T4L-β₂AR-Gs complex with bound nanobody Nb37. The Nb37-enhanced density of the AH domain (marked with an oval) shows variable positioning around the Ras-like domain of Gαs. The scale bars correspond to 5 nm.
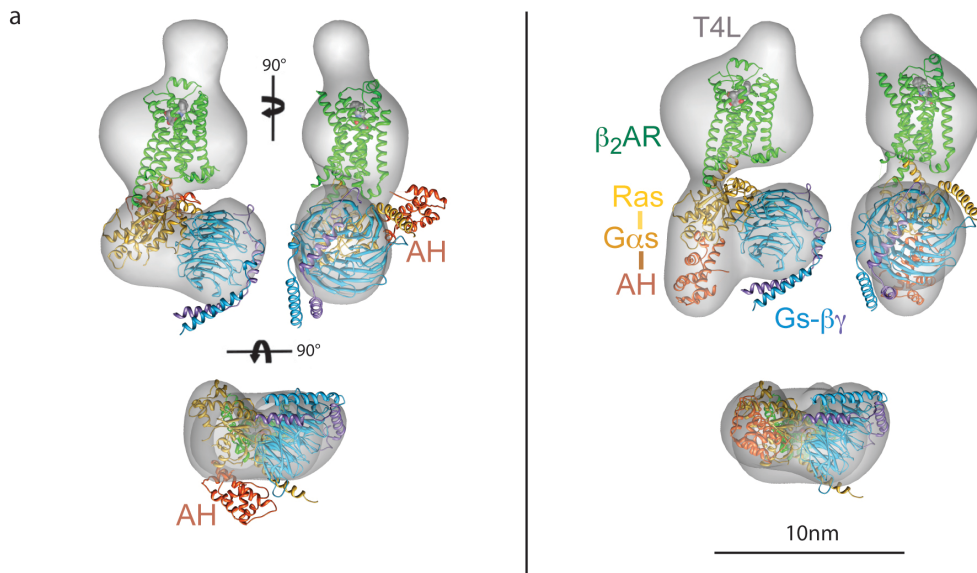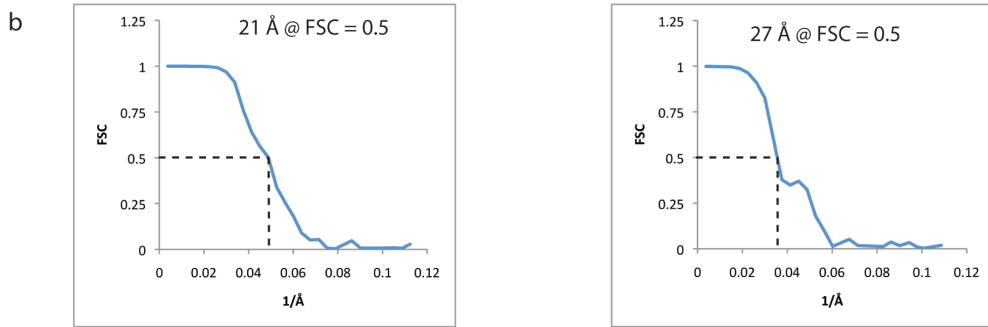
Figure 2-10. 3D reconstructions of nucleotide-free T4L-β₂AR-Gs complex.
a, 3D reconstructions of T4L-β₂AR-Gs complex in two different observed conformations of the α-helical (AH) domain in nucleotide-free conditions. Shown are front, side, and top views of the T4L-β₂AR-Gs complex. Left, model with no density sufficient to accommodate the AH domain. The AH domain is shown in the same position shown in crystal structure (Rasmussen et al., 2011). Right, a model is shown with the AH domain observed on the Ras-like domain of Gαs. The AH domain structure (orange) has been docked in this position. b, Fourier shell correlation (FSC) curves for each of the respective 3D reconstructions.
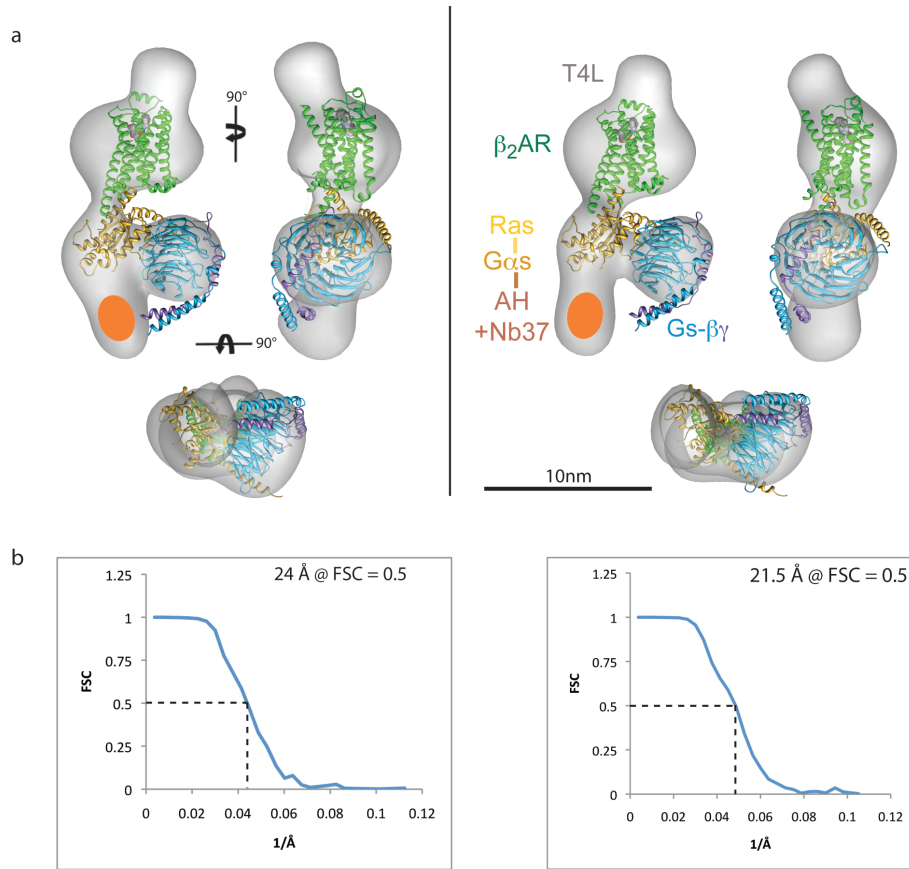
Figure 2-11. 3D reconstructions of T4L-β₂AR-Gs complex with bound camelid antibody Nb37.

a, 3D reconstructions of different conformations of T4L-β₂AR-Gs complex with bound camelid antibody Nb37 targeted to the α-helical (AH) domain of Gαs. Models of front, side, and top views of the T4L-β₂AR-Gs complex with bound Nb37. The density marked by the orange oval corresponds to AH domain with bound Nb37. The structure of the AH domain is not shown. b, Fourier shell correlation (FSC) curves for each of the respective 3D reconstructions.
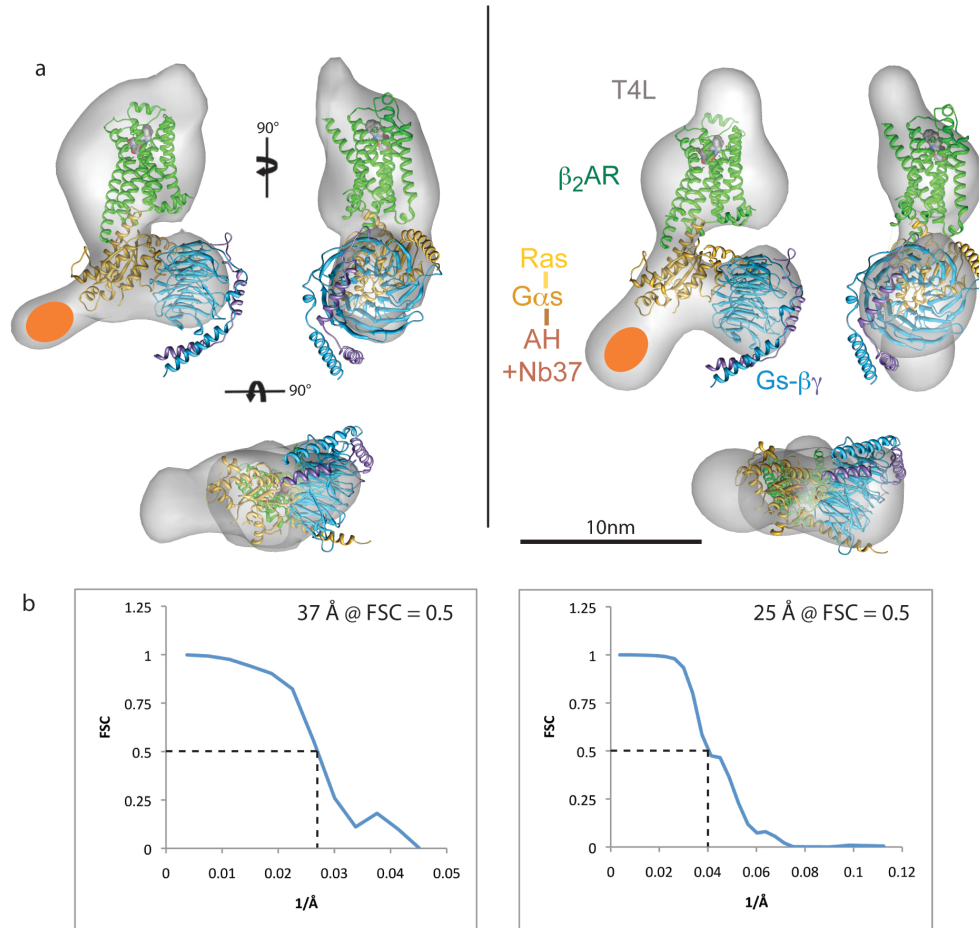
Figure 2-12. 3D reconstructions of T4L-β₂AR-Gs complex with bound camelid antibody Nb37.
a, 3D reconstructions of different conformations of T4L-β₂AR-Gs complex with bound camelid antibody Nb37 targeted to the α-helical (AH) domain of Gαs. Models of front, side, and top views of the T4L-β₂AR-Gs complex with bound Nb37. The density marked by the orange oval corresponds to AH domain with bound Nb37. The structure of the AH domain is not shown. b, Fourier shell correlation (FSC) curves for each of the respective 3D reconstructions.
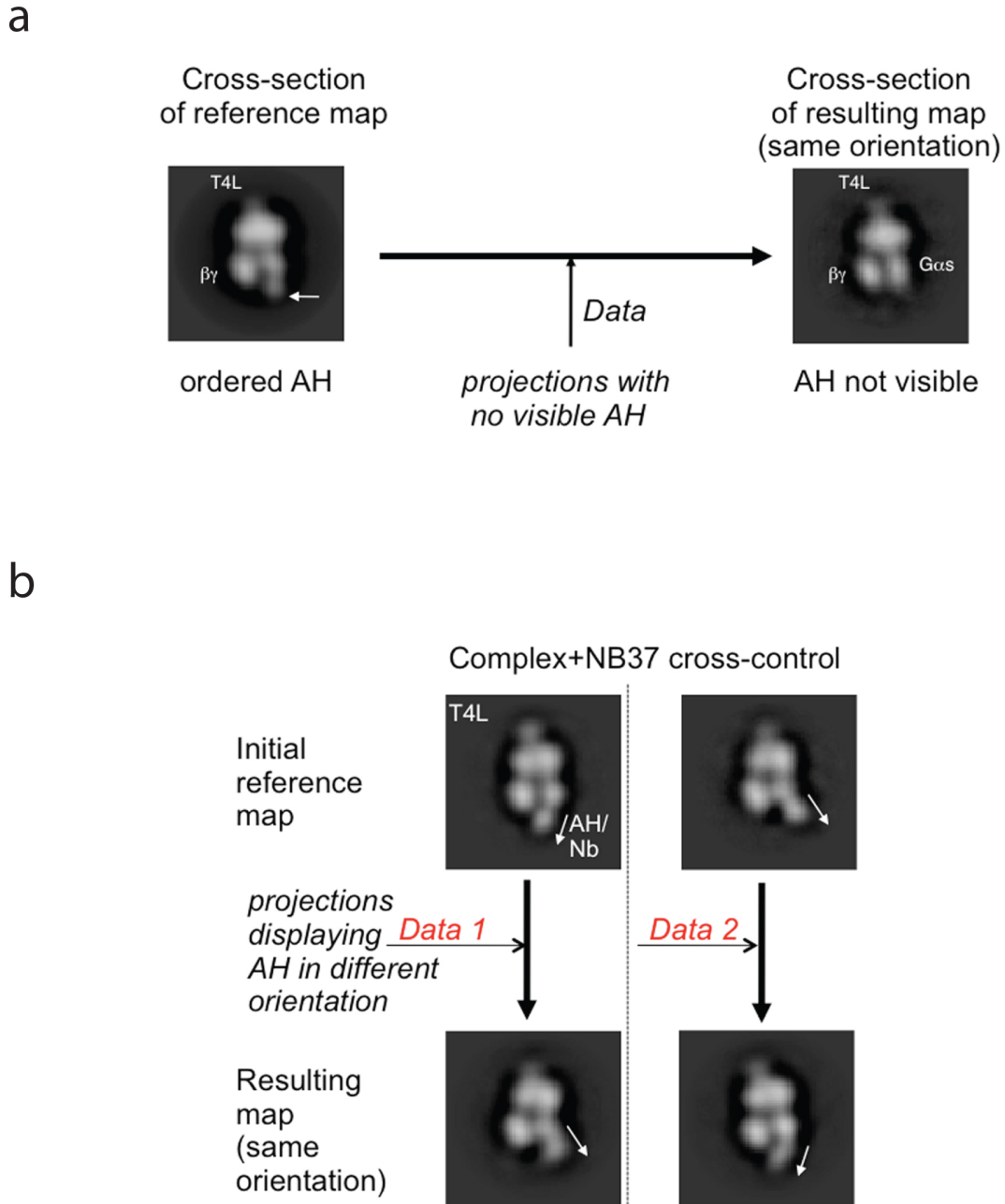
Figure 2-13. Cross-validating 3D reconstructions.
a, Projections classified as not having an ordered AH domain were aligned against a 3D reference volume with ordered AH domain. The resulting 3D map shows no ordered AH domain, as its density disappears. b, Cross sections of starting and final maps in the same orientation after alignment and reconstruction of projections with AH/Nb37. The 2D input data have a different conformation of the AH/Nb37 module compared to the starting reference map. The final 3D map reflects the AH/Nb37 conformation of the input data, which is entirely different from the initial reference 3D map.

53

Incubation of different nucleotides or mimics with β₂AR-Gs

As noted above, previous crystal structures of G proteins show that bound nucleotides contribute to the stability of interactions between the Ras and AH domains. We therefore investigated the positioning of the AH domain in the presence of guanine nucleotides and nucleotide fragments. Pyrophosphate (PPi), representing two phosphates in GTP or GDP, has been shown to bind Ras in a Mg^{2+}-dependent manner, presumably at the β- and γ-phosphate positions (Zhang et al., 2005). PPi and its chemically more stable analogue foscarnet, more known for its antiviral properties (Sundquist and Oberg, 1979), also bind to heterotrimeric Gαs with an apparent affinity of ~0.5 and 1.6 mM, respectively, as determined by competition binding with a fluorescent GTPγS probe (Bodipy-GTPγS) (Figure 2-14). Binding of PPi and foscarnet most likely substitutes for the α- and β-phosphates of GDP rather than the β- and γ-phosphates of GTP. Binding of β- and γ-phosphates would result in modification of the switch II domain with subsequent Gβγ dissociation and dissolution of the complex (Sprang, 1997). However, PPi (with our without Mg^{2+}) does not disrupt the receptor-G protein complex. This is in contrast to dissociation of the complex observed with the GTP mimetic GTPγS (Rasmussen et al., 2011).
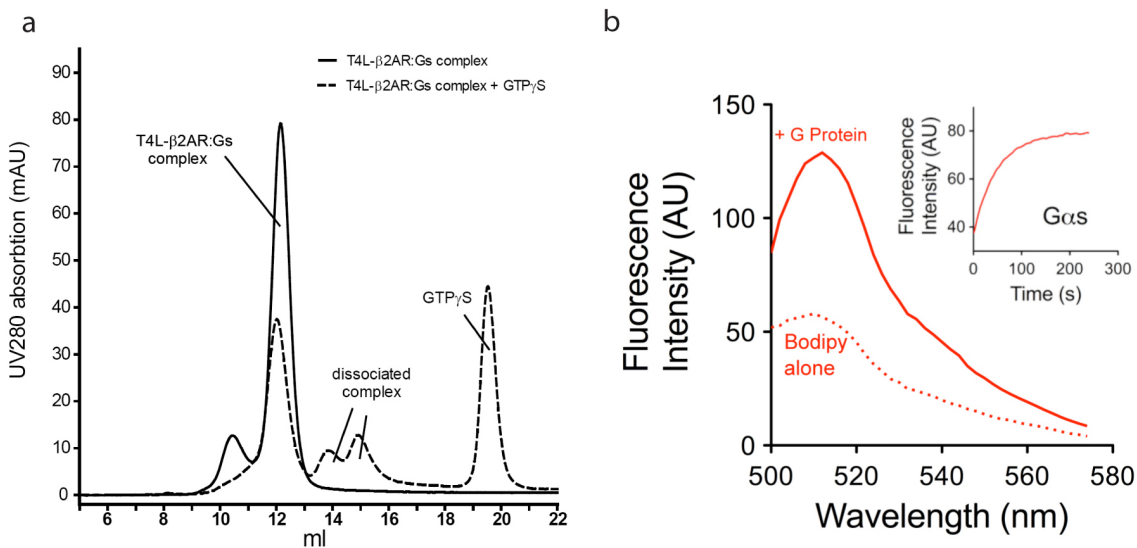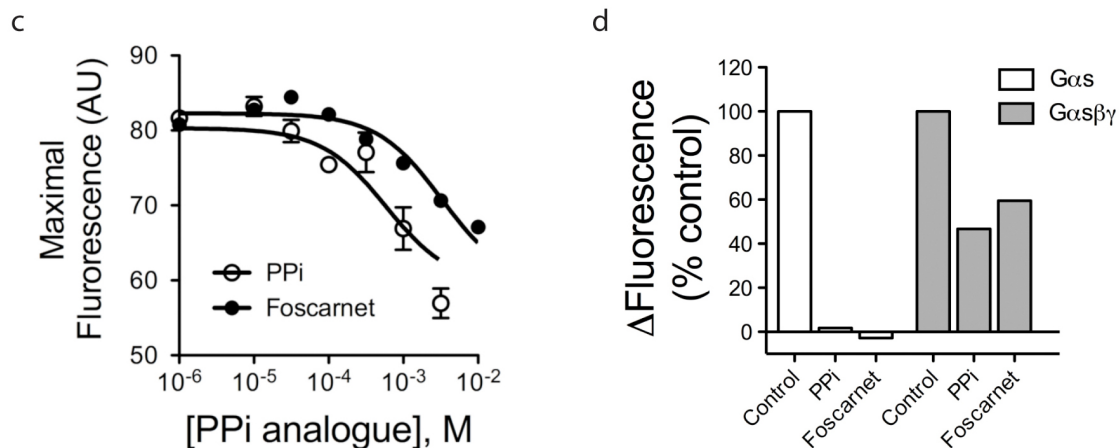
Figure 2-14. GTPγS binding
a, Analytical gel filtration of the T4L-β₂AR:Gs complex (120 ug) incubated in the absence (solid line) or presence of 100 μM GTPγS (broken line) for 5 min at room temperature. b, The effect of foscarnet and PPi was measured using 100 nM bodipy-GTPγS-FL (Invitrogen, CA). Fluorescence intensity of bodipy-GTPγS-FL ($\lambda ex\sim470$ nm) increases upon G protein binding as demonstrated by McEwan et al. 2001. Fig b shows a wavelength scan of bodipy-GTPγS-FL (100 nM) in the absence (dotted) or presence (solid) of a molar excess of purified Gαs (1 mM). (inset), Bodipy-GTPγS (100 nM) was incubated with 1 μM purified Gαs and fluorescence measured in realtime. c, Pyrophosphate (PPi) and its chemically stable pyrophosphate analogue phosphonoformate (foscarnet), inhibits bodipy-GTPγS-FL ($\lambda ex\sim470nm$, $\lambda em\sim515$) binding to purified Gαs as in b) in a concentration-dependent manner (IC50 ~ 0.5 and 1.6 mM, respectively). d, inhibition of bodipy-GTPγS-FL by both PPi (3 mM) and foscarnet (10 mM) in both Gαs alone or in its trimeric form with Gβγ (Gαsβγ). Figure courtesy of the Sunahara Lab.

Even though the presence of PPi does not appear to affect the AH domain positioning, in the presence of Mg²⁺•foscarnet we observe a significantly higher proportion (~70%) of complexes with an ordered AH region on the Ras-like domain (Figures 2-15 thru 2-20). The observations in 2D class averages were also reproduced by individual 3D reconstructions for the different conformers in each state (Figures 2-19 and 2-20). These results further confirm that the variability in the visibility of the AH domain is indeed due to its variable positioning and not due negative stain artifacts such as incomplete embedding. The ability of foscarnet to stabilize the AH domain on the Ras domain suggests it is acting as a ligand fragment that binds to the nucleotide binding pocket. Given its low affinity, it is not surprising that the stabilization is incomplete.
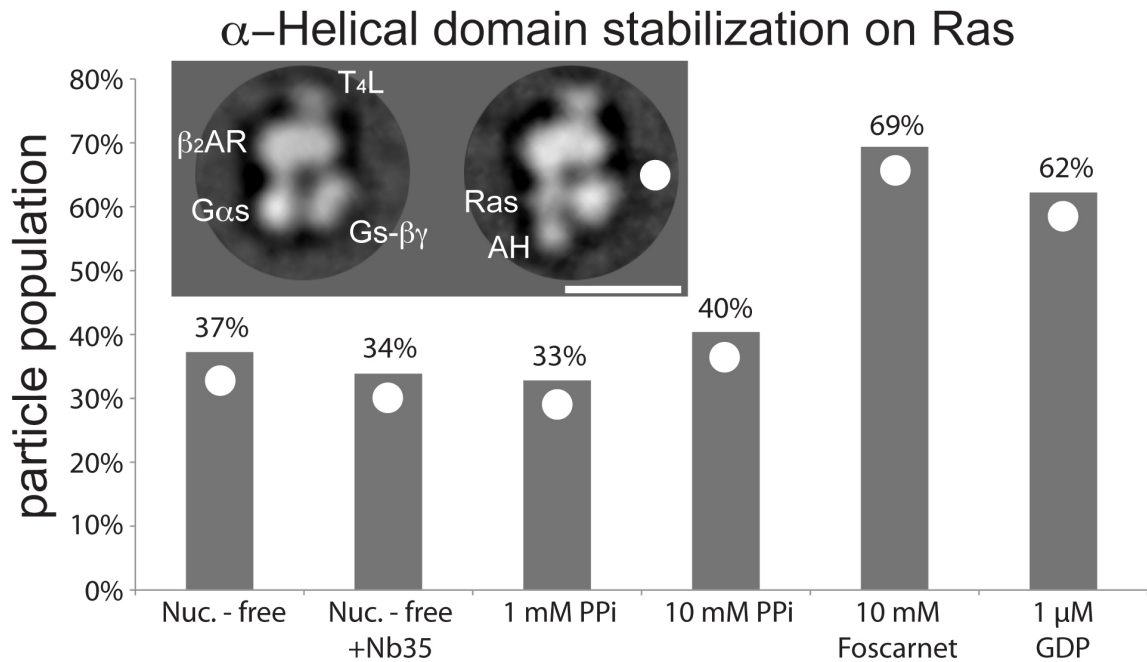
Figure 2-15. Nucleotide dependent positioning of the Gαs AH domain

Distribution of particles with a distinct projection profile of the AH domain stabilized on the Ras-like domain across different conditions (inset, right panel, and marked with a white dot). A class average of a particle with a non-visible AH domain is shown for comparison (inset, left panel). The presence of foscarnet and GDP significantly increases the number of particles with stabilized AH domain.

Figure 2-16. Initial and secondary 2D classification of T4L-β₂AR-Gs complex in the presence of 1 mM Pyrophosphate (PPi).
a) Initial 2D classification. 13,220 particles classified into 150 classes. b) Secondary 2D classification. 3,196 particles of T4L-β₂AR-Gs complex in the presence of 1 mM PPi were classified into 50 classes. Class averages with the α-helical (AH) domain observed to be stabilized on the Ras-like domain of Gαs are marked with an orange square. Class averages where the AH domain projection profile on the Ras-like domain is not visible are marked with a green circle. Scale bar = 10 nm.
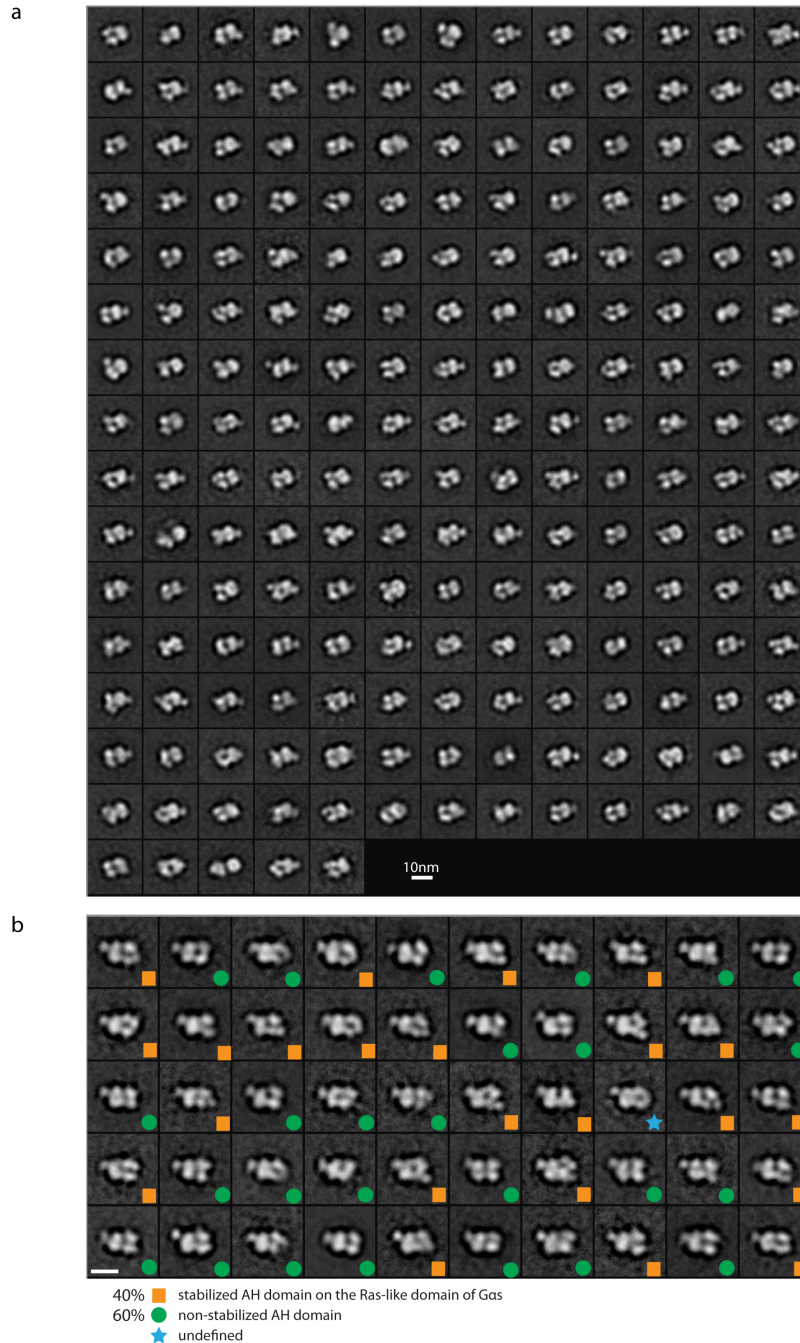
Figure 2-17. Initial and secondary 2D classification of T4L-β₂AR-Gs complex in the presence of 10 mM Pyrophosphate (PPi).
a) Initial 2D classification. 30,006 particles classified into 200 classes. b) Secondary 2D classification. 4,464 particles of T4L-β₂AR-Gs complex in the presence of 10 mM PPi were classified into 50 classes. Class averages with the α-helical (AH) domain observed to be stabilized on the Ras-like domain of Gαs are marked with an orange square. Class averages where the AH domain projection profile on the Ras-like domain is not visible are marked with a green circle. Class averages of insufficient quality to categorize are marked with a cyan star. Scale bar = 10 nm

Figure 2-18. Initial and secondary 2D classification of T4L-β₂AR-Gs complex in the presence of 10 mM Foscarnet.

a) Initial 2D classification. 15,630 particles were classified into 200 classes. b) Secondary 2D classification. 3,715 particles of T4L-β₂AR-Gs complex in the presence of 10 mM Foscarnet were classified into 50 classes. Class averages with the α-helical (AH) domain observed to be stabilized on the Ras-like domain of Gαs are marked with an orange square. Class averages where the AH domain projection profile on the Ras-like domain is not visible are marked with a green circle. Scale bar = 10 nm.
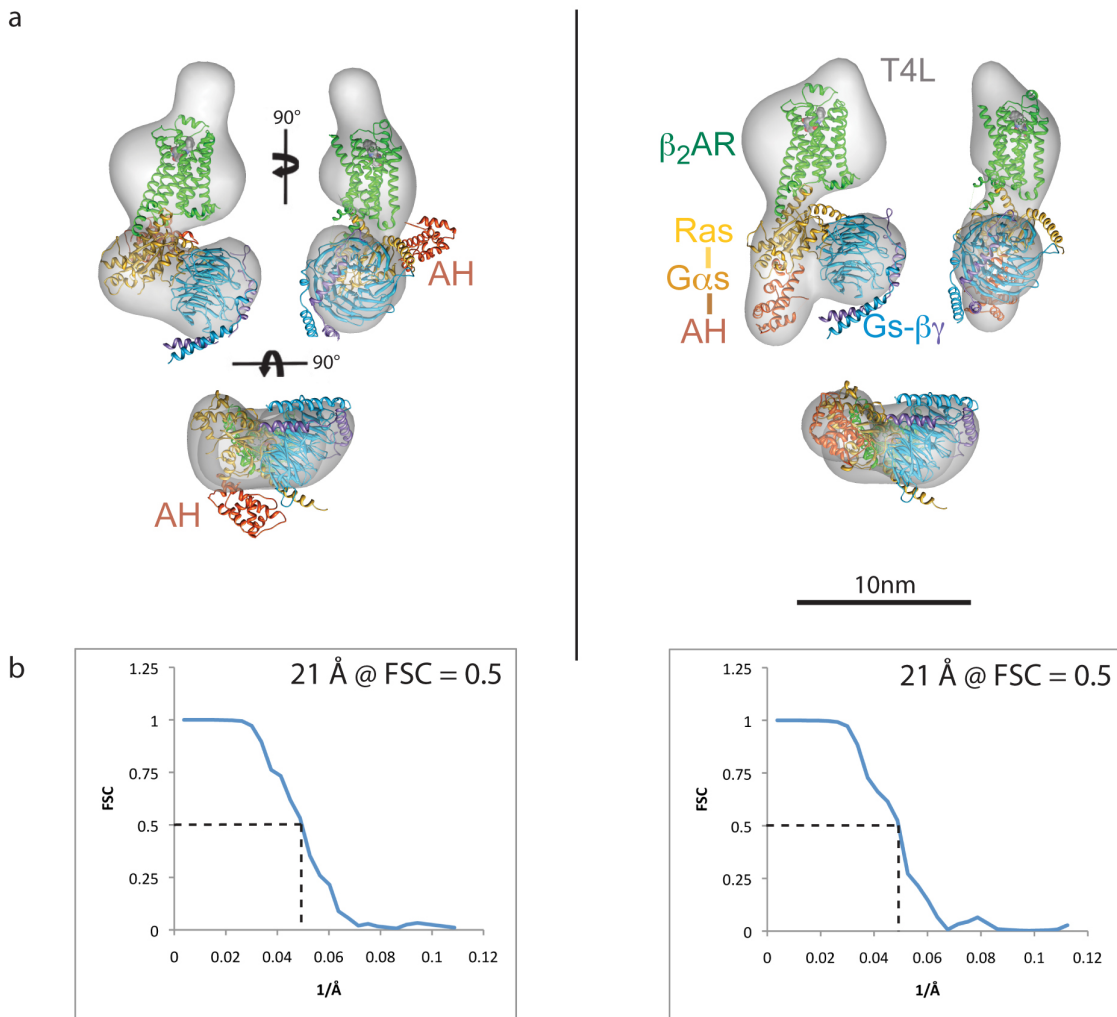
Figure 2-19. 3D reconstructions of T4L-β₂AR-Gs complex in the presence of 10 mM Pyrophosphate (PPi).

a, 3D reconstructions of T4L-β₂AR-Gs complex in two different observed conformations of the α-helical (AH) domain in the presence of 10 mM PPi. Shown are front, side, and top views of the T4L-β₂AR-Gs complex. Left, model with no density sufficient to accommodate the AH domain. The AH domain is shown in the same position shown in crystal structure (Rasmussen et al., 2011). Right, a model is shown with the AH domain observed on the Ras-like domain of Gαs. The AH domain structure (orange) has been docked in this position. b, Angular distributions of particle projections for each of the respective 3D reconstructions. c, Fourier shell correlation (FSC) curves for each of the respective 3D reconstructions.
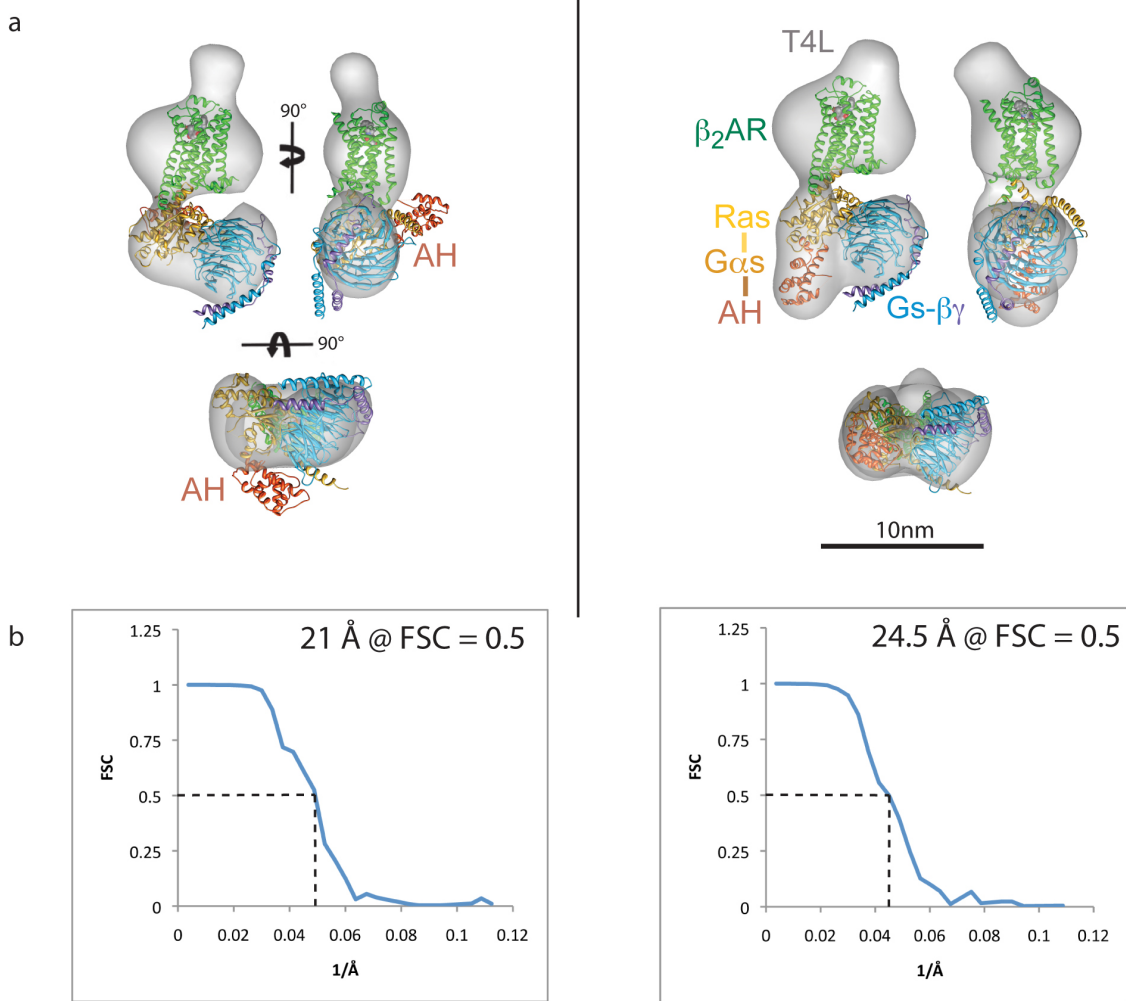
Figure 2-20. 3D reconstructions of T4L-β₂AR-Gs complex in the presence of 10 mM Foscarnet.

a, 3D reconstructions of T4L-β₂AR-Gs complex in two different observed conformations of the α-helical (AH) domain in the presence of 10 mM foscarnet. Shown are front, side, and top views of the T4L-β₂AR-Gs complex. Left, model with no density sufficient to accommodate the AH domain. The AH domain is shown in the same position shown in crystal structure (Rasmussen et al., 2011). Right, a model is shown with the AH domain observed on the Ras-like domain of Gαs. The AH domain structure (orange) has been docked in this position. b, Angular distributions of particle projections for each of the respective 3D reconstructions. c, Fourier shell correlation (FSC) curves for each of the respective 3D reconstructions.

In contrast to PPi and foscarnet, addition of GDP or the non-hydrolyzable GTP analogue GTPγS leads to dissociation of the β$_2$AR-Gs complex (Rasmussen et al., 2011). To examine the effect of GDP and GTPγS, we rapidly mixed the complex with nucleotide and immediately fixed the sample by negative stain embedding. Addition of either of these nucleotides at concentrations above 10 μM resulted in significant amounts of partially dissociated complexes (Figure 2-21). This result is expected since a large excess of either of these nucleotides would uncouple the G protein from the receptor. However, short incubation with lower GDP concentrations (1 μM) and immediate sample fixation for EM allowed to us to examine intact complexes, revealing that the AH region was ordered in ~60% of the intact particles (Figures 2-22 and 2-23). In contrast, even low concentrations of GTPγS (1 μM) showed a significant amount of destabilized complexes, and we were able to capture an array of intermediate dissociation states (Figure 2-24a and 2-22c & 2-22d). Collectively, these data strongly suggest that the presence of nucleotide, or nucleotide fragments such as foscarnet, results in AH domain stabilization against the Ras-like domain of Gαs. In the absence of nucleotide, the position of the AH domain is highly variable (Figures 2-15 and 2-25).

Confirming alpha helical flexibility with other techniques

Our results are in agreement with a recent study of the complex formed by Gi and rhodopsin. Using double electron-electron resonance (DEER) spectroscopy, Hamm, Hubbell and colleagues documented large (up to 20 Å) changes in distance between nitroxide probes positioned on the Ras and AH domains of Gi upon formation of a complex with light-activated rhodopsin (Van Eps et al., 2011). The broad distance distributions observed for several labeling pairs are compatible with multiple conformations in dynamic equilibrium. Our findings are also consistent with results from DXMS that show increased deuterium exchange at both the nucleotide binding pocket and at sites of interaction between the Ras and AH domains upon formation of the β$_2$AR-Gs complex (Chung et al., 2011).
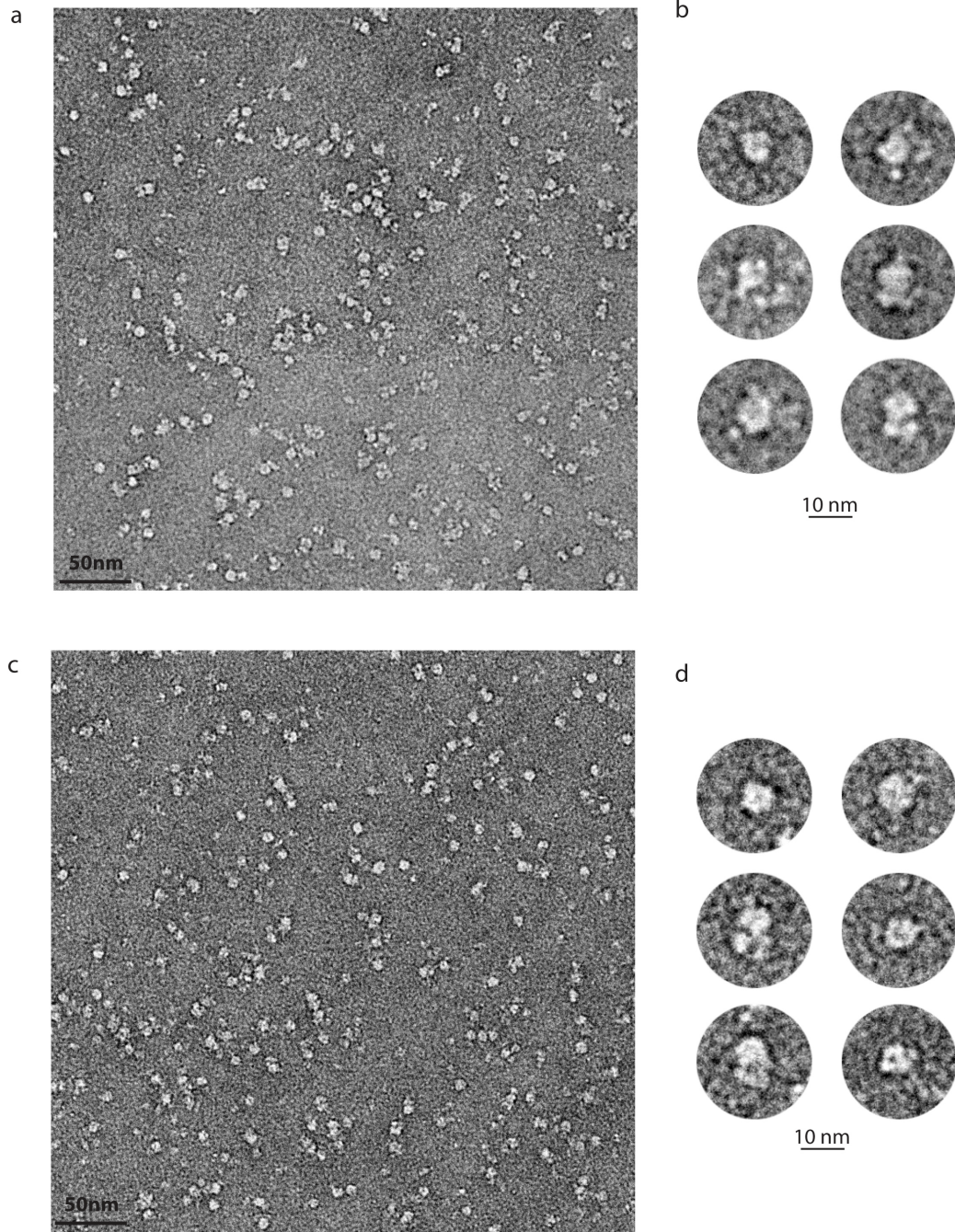
Figure 2-21. Raw EM image of negative stained T4L-β₂AR-Gs complex in the presence of 10 μM GTPγS and GDP.
a, Raw image of negative stained T4L-β₂AR-Gs complex in the presence of 10 μM GTPγS. b, Excised particles of T4L-β₂AR-Gs complex in the presence of 10 μM GTPγS show complex dissociation. c, Raw image of negative stained T4L-β₂AR-Gs complex in the presence of 10 μM GDP. d, Excised particles of T4L-β₂AR-Gs complex in the presence of 10 μM GDP show complex dissociation.
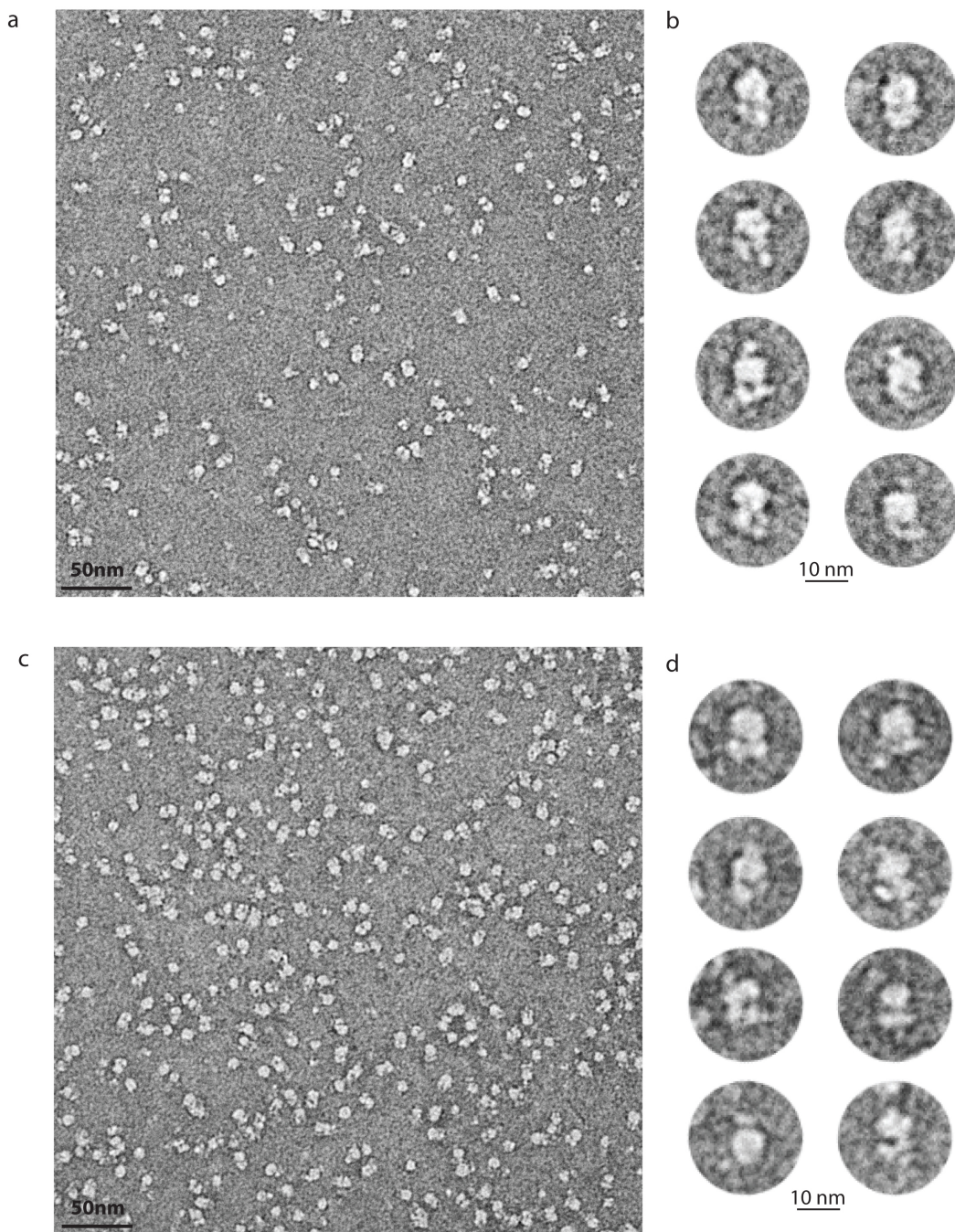
Figure 2-22. Raw EM image of negative stained T4L-β₂AR-Gs complex in the presence of 1 μM GDP and 1 μM GTPγS.
a, Raw image of negative stained T4L-β₂AR-Gs complex in the presence of 1 μM GDP. b, Excised particles of T4L-β₂AR-Gs complex in the presence of 1 μM GDP. c, Raw image of negative stained T4L-β₂AR-Gs complex in the presence of 1 μM GTPγS. d, Excised particles of T4L-β₂AR-Gs complex in the presence of 1 μM GTPγS.

a) partially dissociated

62% ■ stabilized AH domain on the Ras-like domain of Gαs
38% ● non-stabilized AH domain
★ undefined

Figure 2-23. Initial and secondary 2D classification of T4L-β₂AR-Gs complex in the presence of 1 μM GDP.

a) Initial 2D classification. 17,438 particles were classified into 200 classes. Class averages where the T4L-β₂AR-Gs complex is partially dissociated are marked with a magenta hexagon. b) Secondary 2D classification. 2,588 particles of T4L-β₂AR-Gs complex in the presence of 1 μM GDP were classified into 50 classes. Class averages with the α-helical (AH) domain observed to be stabilized on the Ras-like domain of Gαs are marked with an orange square. Class averages where the AH domain projection profile on the Ras-like domain is not visible are marked with a green circle. Class averages of insufficient quality to categorize are marked with a cyan star. Scale bar = 10 nm.