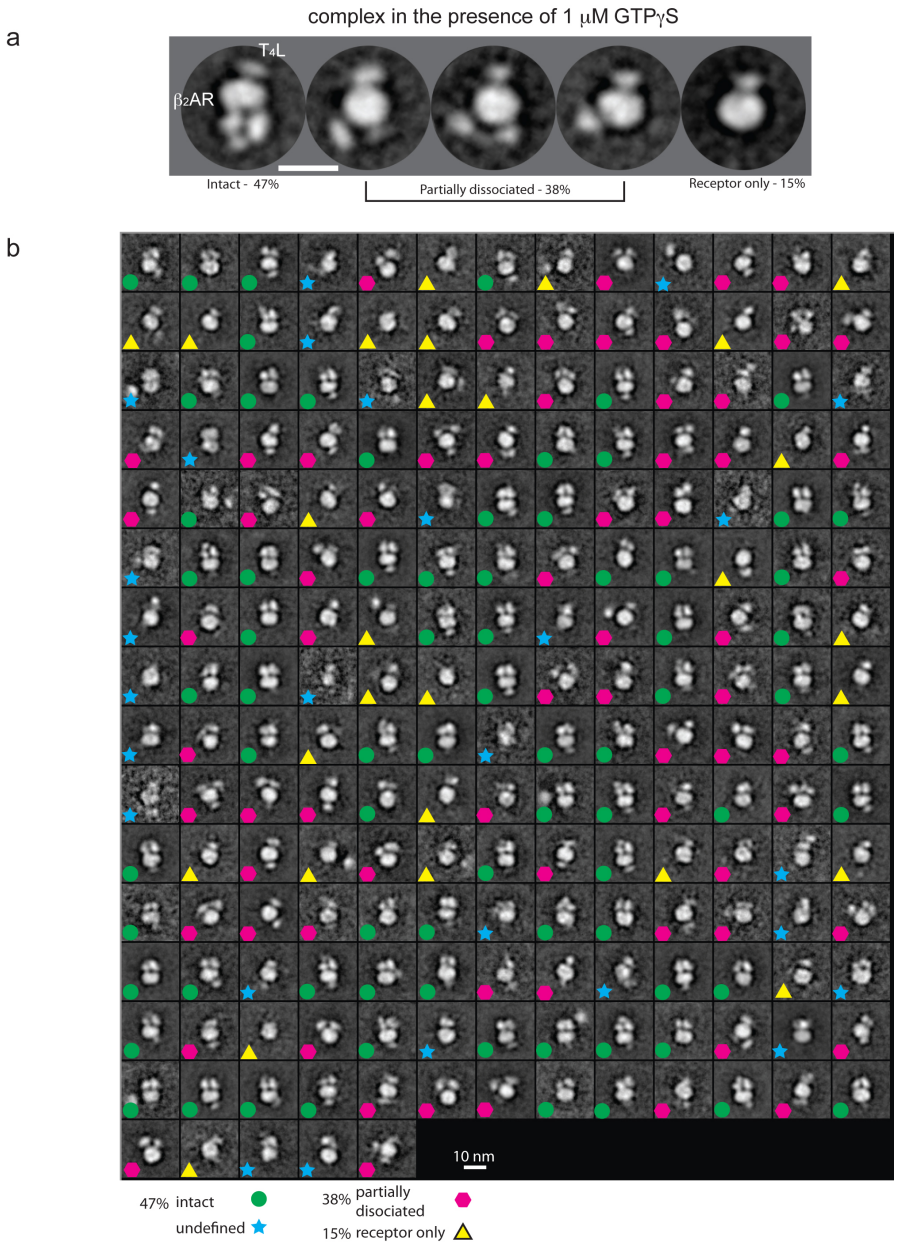


Figure 2-24. 2D classification of T4L-β_2AR-Gs complex in the presence of 1 μM GTPγS.

a) Representative 2D classifications of T4L-β_2AR-Gs complex in the presence of 1 μM GTPγS showing intact complexes, partially dissociated, and receptor only class averages. b) Full 2D classification. 10,140 particles were classified into 200 classes. Class averages showing intact T4L-β_2AR-Gs complex particles are marked with a green circle. Class averages where the T4L-β_2AR-Gs complex is partially dissociated are marked with a magenta hexagon. Class averages showing only density from the β_2AR receptor are marked with a yellow triangle. Class averages of insufficient quality to categorize are marked with a cyan star.

2.4 Discussion

In these studies we reveal the flexibility of the alpha helical (AH) domain of the β_2AR-Gs complex using negative stain electron microscopy. Interestingly, the position of the AH-domain of Gαs is observed as shown in the x-ray crystal structure behind the Ras domain (Rasmussen et al., 2011), below the Ras domain of Gαs, and in a number of other positions when labeled with nanobody. This, in addition to the ability to capture the β_2AR-Gs complex in transient states of dissociation when incubated with GTPγS shows why electron microscopy and single-particle analysis can be a useful tool to characterize the architecture of protein assemblies. Embedding the β_2AR-Gs complex in negative stain gave the necessary contrast to detect minor structural changes in the position of the alpha helical domain. EM, in comparison to x-ray crystallography that gives only a singular snap shot (while higher resolution), can reveal changes that are not detected by other structural techniques.

EM also played a very pivotal role in screening different purification conditions, including incubations with a number of different nanobody candidates, and the condition the eventually led to the crystal formation and x-ray structure was first characterized by EM. The combination of different structural techniques can productively contribute to a complete characterization of a given specimen. The different conformations of the β_2AR-Gs complex shown in these studies can now help to ask and answer additional questions in terms of the activation of this G protein-GPCR complex.

Support for the open conformation *in vivo* may come from studies on the action mechanism of the cholera toxin, the enterotoxin secreted by the pathogen *vibrio cholerae*. Cholera toxin, together with ADP ribosylation factor (ARF), ADP-ribosylate Gαs at R201, rendering the residue catalytically ineffective and the G protein GTPase-deficient(Freissmuth and Gilman, 1989). The ADP-ribosylated and constitutively active Gαs will continue to stimulate adenylyl cyclase, cAMP production and protein kinase A (PKA) activation. Activated PKA opens intestinal Cl^- channels and leads to increased water secretion that results in diarrhea (Gabriel et al., 1994). Crystallographic studies of G proteins in GDP and GTPγS-bound forms indicate that the catalytic arginine (R201 in Gs) is buried and likely inaccessible to cholera toxin and ARF (Sunahara et al., 1997).

However formation of the nucleotide-free form of Gαs and opening of the Ras and AH domains would facilitate accessibility to R201.

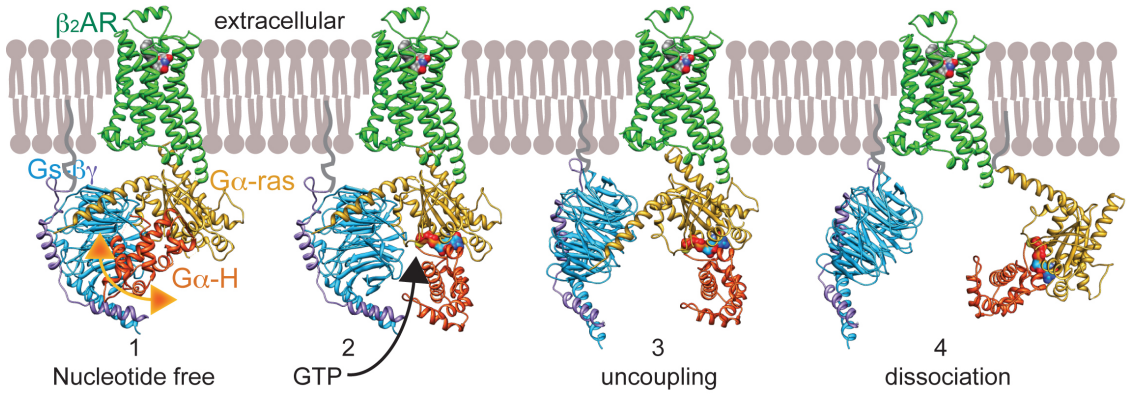


Figure 2-25. Model of conformational transitions in the β_2AR-Gs complex.
The nucleotide-free β_2AR-Gs complex is characterized by a highly flexible AH domain (model 1). GTP binding promotes stabilization of the AH domain on the Ras-like domain of Gαs (model 2). In this model, the AH domain has the same position as the one observed in the crystal structure of Gαs-GTPγS (Sunahara et at., 1997). GTP binding results in subsequent uncoupling of Gs and β_2AR (model 3), with eventual dissociation of Gαs and Gβγ (model 4).

In conclusion, single-particle EM analysis of the β_2AR-Gs complex has provided novel structural insights into the dynamic nature of the assembly (Figure 2-25). EM visualization of nanobody bound complexes allowed us to clearly reveal the variable positioning of the AH domain under nucleotide-free conditions and, additionally, to identify conditions that were key for the successful characterization of the complex by X-ray crystallography. Furthermore, single-particle EM examination of the β_2AR-Gs complexes in varying concentrations of GDP and GTPγS enabled us to capture transient intermediate dissociation states between β_2AR and Gs. This approach should prove useful for studying other signaling complexes involving GPCRs and other membrane proteins. The combination and integration of these technologies will be crucial for studying structural aspects of challenging macromolecular complexes at large.

2.5 Experimental Procedures

Specimen preparation and EM imaging of negative-stained samples

The T4L-β_2AR-Gs complex and nanobodies (Nb) were prepared as described in Rasmussen *et al* .(Rasmussen et al., 2011). Specimens were visualized by electron microscopy (EM) in the following conditions: a) T4L-β_2AR-Gs complex alone; b) T4L-β_2AR-Gs in presence of Nb35; c) T4L-β_2AR-Gs in presence of Nb37; d) T4L-β_2AR-Gs in presence of 1 mM PPi; e) T4L-β_2AR-Gs in presence of 10 mM PPi; f) T4L-β_2AR-Gs in presence of 10 mM Foscarnet and 10 mM $MgCl_2$; g) T4L-β_2AR-Gs in presence of 1 μM GDP and 10 mM $MgCl_2$; h) T4L-β_2AR-Gs in presence of 1 μM GTPγs and 10 mM $MgCl_2$. All samples were prepared for EM using the conventional negative staining protocol(Ohi et al., 2004). For nanobody labeling, T4L-β_2AR-Gs complex was incubated for 15 minutes at room temperature with approximately equimolar concentrations of Nb35 or Nb37, and subsequently prepared by negative staining. For nucleotide (GDP or GTPγS) and nucleotide fragment (PPi or foscarnet) incubations, these components were rapidly mixed with the complex and the sample was immediately fixed by negative stain embedding.

Specimens were imaged at room temperature with a Tecnai T12 electron microscope operated at 120 kV using low-dose procedures. Images were recorded at a magnification of 71,138x and a defocus value of ~1.5 μm on a Gatan US4000 CCD camera. All images were binned (2 x 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. Tilt pair particles from 60° and 0° images were selected using WEB(Frank et al., 1996). Particles for only 2D classification of 0° projections were excised using Boxer (part of the EMAN 1.9 software suite)(Ludtke et al., 1999). The number of particles or tilt-pair particle projections per condition is provided in Table 2-1.

Condition:	Number of Particles:	2D Classes:
T4L-β_2AR-Gs (nucleotide-free)	17,205 (TP)	200
+Nb35	6,514	75
+Nb37	14,704 (TP)	200
+1mM PPi	13,220	150
+10mM PPi	30,006 (TP)	200
+10mM Foscarnet, 10mM MgCl2	15,630 (TP)	200
+1μM GDP and 10mM MgCl2	17,438	200
+1μM GTPγs and 10mM MgCl2	10,140	200

Table 2-1. Particle numbers in each 2D classification.

2D classifications and 3D reconstructions of T4L-β_2AR-Gs

2D reference-free alignment and classification of particle projections were performed using SPIDER(Frank et al., 1996). For all conditions, the 0° particle projections were iteratively classified into multiple classes for 10 cycles. Table 2-1 provides the number of classes for each condition. For AH conformation assignments, we used the first classification to select only the particles from averages clearly displaying the profiles of Ras-like, Gβγ, β_2AR, and T4L domain densities in the same position, thereby restricting the range of particle projection orientations. These projections were pulled together and subjected to a second iterative classification (referred to as secondary classification). Table 2-2 provides the number of classes and particle projections for each condition in the secondary classification. For counting the numbers of particles with and without stabilized α-helical domain on the Ras-like domain, three different operators examined each secondary classification and assigned each class average according to the projection profile of the specific region. The assignment from the different operators was in good agreement, and the particle numbers belonging to individual classes were added to calculate percentages for each conformation. Assignments for each full individual data set were done in addition to the secondary classification, and the results were in

agreement. To test any bias, the particles from nucleotide-free, 1 mM PPi, 10 mM PPi, and foscarnet conditions were combined into a single data set of 15,753 particles and were classified into 200 classes. The individual class averages were assigned as before according to the visibility of the AH domain, and the percentage of projections from each condition was determined according to the number of projections contributing to the assigned class averages (Figure 2-26). The results of this "blind" test showed very good agreement with our assignments from individual classifications.

Condition:	Number of Particles:	2D Classes:
Nucleotide-free	4,378	50
+Nb35	1,151	30
+1mM PPi	3,196	50
+10mM PPi	4,464	50
+10mM Foscarnet, 10mM MgCl2	3,715	50
+1μM GDP and 10mM MgCl2	2,588	50

Table 2-2. Number of particles contributed to each 3D reconstruction.

For 3D reconstructions, in a first step we employed the random conical tilt technique(Radermacher et al., 1987) to determine initial 3D maps by back-projection of tilted particle images belonging to individual classes. After a first round of angular refinement, corresponding particles from the images of the untilted specimen were added, and the images were subjected to another cycle of refinement. We thus generated reliable initial models for complexes with variability in the positioning of the α-helical domain of Gαs (Figure 2-8). After contrast transfer function (CTF) correction according to local defocus values obtained by CTFTILT(Mindell and Grigorieff, 2003), the full dataset from each condition was subjected to multiple reference-supervised alignment(Brink et al., 2004; Menetret et al., 2005) with the *multirefine* routine in EMAN (1.9) by using our initial models as reference maps. This approach allowed us to separate particles from the entire dataset (of each condition) according to the positioning of the α-helical domain of Gαs. For final maps we used the separated datasets, as provided by the multiple reference-supervised alignment, and employed FREALIGN(Grigorieff, 2007) for further

refinement of the orientation parameters and reconstruction. The resolution for each map was determined at FSC=0.5.

Molecular Modeling

The crystal structure of T4L-β_2AR-Gs (Rasmussen et al., 2011) was fit in the EM density as a rigid body. Due to the presence of the detergent micelle, which accounted for significant density surrounding β_2AR, all docking operations were performed manually with visual inspection of the best fit. The docking of the T4L-β_2AR-Gs complex revealed that the EM density corresponding to Gβγ was shifted further away from the receptor in all 3D maps compared to the crystal structure. This difference might be attributed to the crystal packing, or perhaps a limited deformation of the complex due to the presence of the carbon support on the EM grid, or both. Accordingly, Gβγ was translated manually by 9 Å to best fit its density while retaining all interactions with Gαs. This final model is shown unmodified for all fittings in 3D reconstructions. For maps showing the α-helical domain of Gαs on the Ras-like domain, we manually modeled the α-helical domain in its corresponding position by taking into account steric constraints. In this conformation, the position of the α-helical domain is very similar to the one observed in the crystal structure of Gαs-GTPγS(Sunahara et al., 1997) (Figure 2-3e).

Deuterium Exchange Mass Spectrometry

1.5 ml of R:G complex or 1.5 ml of R:G:NB was mixed with 4.5 ml of D_2O buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM BI-167107, 100 mM TCEP, 0.0015% MNG-3 in D_2O) and incubated for 10, 100, 1000, and 10000 seconds on ice. At the indicated times, the sample was quenched by 15 ml of ice-cold Quench solution (0.1 M NaH_2PO_4, 20 mM TCEP, 16.6% glycerol, pH 2.4), immediately frozen on dry ice, and stored at -80 °C. Nondeuterated control was prepared in H_2O buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 μM BI-167107, 100 μM TCEP, 0.0015% MNG-3 in H_2O), mixed with Quench solution, and snap-frozen on dry ice. Samples were thawed and immediately passed through an immobilized porcine pepsin column (16 ml bed volume) at a flow rate of 20 ml/min of 0.05% trifluoroacetic acid. Peptide fragments were collected contemporaneously on a C18 trap column for desalting and separated by a Magic C18AQ

column (Michrom BioResources Inc., Auburn, CA) using a linear gradient of acetonitrile from 6.4% to 38.4% over 30min. Mass spectrometric analysis was performed using LCQ Classic mass spectrometer from Thermo Finnigan, with capillary temperature of 20 °C. Deuterium quantification data were collected in MS1 profile mode, and peptide identification data were collected in data dependent MS/MS mode. Recovered peptide identification and analysis were carried out using DXMS Explorer (Sierra Analytics Inc, Modesto CA), a software specialized in processing DXMS data (Figure 2-6).

Bodipy-GTPγS binding

The effect of foscarnet and PPi was measured using 100 nM bodipy-GTPγS-FL (Invitrogen, CA). Fluorescence intensity of bodipy-GTPγS-FL (l_{ex}~470nm) increases upon G protein binding as demonstrated by McEwan *et al.* (2001) (McEwen et al., 2001). 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) was determined to assess optimal spectroscopy conditions. The capacity of pyrophosphate (PPi) and the chemically stable pyrophosphate analogue of PPi, phosphonoformate (foscarnet) to inhibit bodipy-GTPγS-FL (l_{ex}~470 nm, l_{em}~515 nm) was measured as described in Figure 2-14 b and c in both Gαs (b & c) and heterotrimeric Gαsβγ (c). The fluorescence of 100 nM bodipy-GTPγS-FL was measured in the presence of 1 mM G protein. PPi or foscarnet were added together with bodipy-GTPγS-FL and initiated by the addition of G protein (1 mM) in 20 mM Tris-HCl, pH 8.0, 3 mM $MgCl_2$, 1 mM DTT in a final volume of 200 mL. Bodipy-GTPγS-FL binding to heterotrimeric G protein included 0.1% dodecylmaltoside (final). Fluorescence was measured on a short time scale (600 s) to minimize the accumulation of hydrolysis product bodipy-phosphate (Jameson et al., 2005). Hydrolysis, which also appears as an increase in fluorescence, was determined simply by chelating Mg^{2+} with 10 mM EDTA following the 600s second incubation (not shown). Inhibition of bodipy-GTPγS-FL by PPi and foscarnet can be reversed with the subsequent addition of high Mg^{2+} (25 mM) which enhances bodipy-GTPγS-FL binding, indicating that PPi and foscarnet are not irreversibly binding or denaturing the G protein (not shown). Gαs was purified as described in Sunahara *et al.* (Sunahara et al., 1997). Gαsβγ was purified as

described by Rasmussen *et al.* (Rasmussen et al., 2011). Fluorescence was measured in a 96-well microtiter plate format on a M5 fluorescence plate reader (Molecular Precision).

Nucleotide state:	Stable:
Nuc.-Free	36%
1mM PPi	35%
10mM PPi	40%
Foscarnet	57%

Figure 2-26. Combined 2D classification of T4L-β2AR-Gs complex.
Combined stacks of nucleotide-free, 1 mM PPi, 10 mM PPi, and Foscarnet T4L-β2AR-Gs complex (15,753 total particles) were classified into 200 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.

2.6 Acknowledgements

Data from this chapter is contained in the manuscript "Structural flexibility of the the Gαs α-helical domain in the β_2-adrenoceptor Gs complex" which was published in the *Proceedings of the National Academy of Sciences*, November 2011. Many members in the Sunahara and Kobilka lab were tirelessly involved in the purification of the β_2AR-Gs complex, nucleotide binding and deuterium exchange assays. Min Su helped in processing the data and ensured the proper instrument alignment. Somnath Dutta assisted in data collection as well as processing. Austin Oleskie helped in the imaging, and the tremendous amount of particle picking that went into this study.

2.7 References

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.

Chung, K.Y., Rasmussen, S.G., Liu, T., Li, S., DeVree, B.T., Chae, P.S., Calinski, D., Kobilka, B.K., Woods, V.L., Jr., and Sunahara, R.K. (2011). Conformational changes in the G protein Gs induced by the beta2 adrenergic receptor. Nature *477*, 611-615.

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.

Freissmuth, M., and Gilman, A.G. (1989). Mutations of GS alpha designed to alter the reactivity of the protein with bacterial toxins. Substitutions at ARG187 result in loss of GTPase activity. J Biol Chem *264*, 21907-21914.

Gabriel, S.E., Brigman, K.N., Koller, B.H., Boucher, R.C., and Stutts, M.J. (1994). Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. Science *266*, 107-109.

Grigorieff, N. (2007). FREALIGN: high-resolution refinement of single particle structures. J Struct Biol *157*, 117-125.

Jameson, E.E., Roof, R.A., Whorton, M.R., Mosberg, H.I., Sunahara, R.K., Neubig, R.R., and Kennedy, R.T. (2005). Real-time detection of basal and stimulated G protein GTPase activity using fluorescent GTP analogues. J Biol Chem *280*, 7712-7719.

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.

McEwen, D.P., Gee, K.R., Kang, H.C., and Neubig, R.R. (2001). Fluorescent BODIPY-GTP analogs: real-time measurement of nucleotide binding to G proteins. Anal Biochem *291*, 109-117.

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.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol *142*, 334-347.

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.

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.

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.

Rubinstein, J.L. (2007). Structural analysis of membrane protein complexes by single particle electron microscopy. Methods *41*, 409-416.

Sprang, S.R. (1997). G protein mechanisms: insights from structural analysis. Annu Rev Biochem *66*, 639-678.

Sunahara, R.K., Tesmer, J.J., Gilman, A.G., and Sprang, S.R. (1997). Crystal structure of the adenylyl cyclase activator Gsalpha. Science *278*, 1943-1947.

Sundquist, B., and Oberg, B. (1979). Phosphonoformate inhibits reverse transcriptase. J Gen Virol *45*, 273-281.

Van Eps, N., Preininger, A.M., Alexander, N., Kaya, A.I., Meier, S., Meiler, J., Hamm, H.E., and Hubbell, W.L. (2011). Interaction of a G protein with an activated receptor

opens the interdomain interface in the alpha subunit. Proc Natl Acad Sci U S A *108*, 9420-9424.

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.

Zhang, B., Zhang, Y., Shacter, E., and Zheng, Y. (2005). Mechanism of the guanine nucleotide exchange reaction of Ras GTPase--evidence for a GTP/GDP displacement model. Biochemistry *44*, 2566-2576.

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

3.1 Abstract

Seven transmembrane receptors (7TMRs), also known as G Protein Coupled Receptors (GPCRs) are critically regulated by β-arrestins (βarr) that not only desensitize G protein signaling through these receptors but also initiate a G protein independent wave of signaling. Recent surge of structural data on a number of GPCRs, including landmark structural studies on β_2 adrenergic receptor (β_2AR)-G protein complex, has provided novel insights into the structural basis of receptor activation. However, in order to obtain a complete picture of GPCR activation and regulation, a complementary set of structural studies are absolutely required on a GPCR-βarr complex. Here, we describe a novel approach involving a conformationally selective antibody fragment against activated βarr1, to stabilize and purify a stable, biochemically and pharmacologically functional β_2AR-βarr1 complex. We apply electron microscopy with 2D single particle analysis to characterize the architecture of this important signaling complex. The results described here present for the first time, a three-dimensional architecture of a GPCR-βarr signaling complex and should provide a structural framework to fully understand the structural basis of GPCR activation, signaling and regulation.

3.2 Introduction

GPCRs participate in virtually every physiological process through their innumerable signaling networks in cells. Obviously, an incredibly tight regulation of these receptors is absolutely required to prevent cellular signaling from going haywire. Cells have devised desensitization mechanisms to precisely control the spatial and temporal signaling patterns of GPCR signaling that primarily involve phosphorylation of agonist activated

receptors by GRKs followed by binding of multifunctional scaffold proteins, βarrs (Lohse et al., 1990). GPCR-βarr interaction not only terminates further G protein coupling through steric hindrance but in addition, it also facilitates receptor internalization through clathrin coated machinery and initiates a G protein independent wave of signaling downstream the activated GPCRs (Luttrell and Lefkowitz, 2002). However, detail studies to decipher the structural aspects of GPCR-βarr interaction are very sparse. Therefore, we set out purify a stable β_2AR-β-arr complex and obtain detail structural insights in to this highly important signaling complex. However, our extensive attempts to form a stable β_2AR-β-arr complex from purified components *in vitro* were not successful, presumably due to transient nature of the complex or detergent sensitivity of purified β-arrestins.

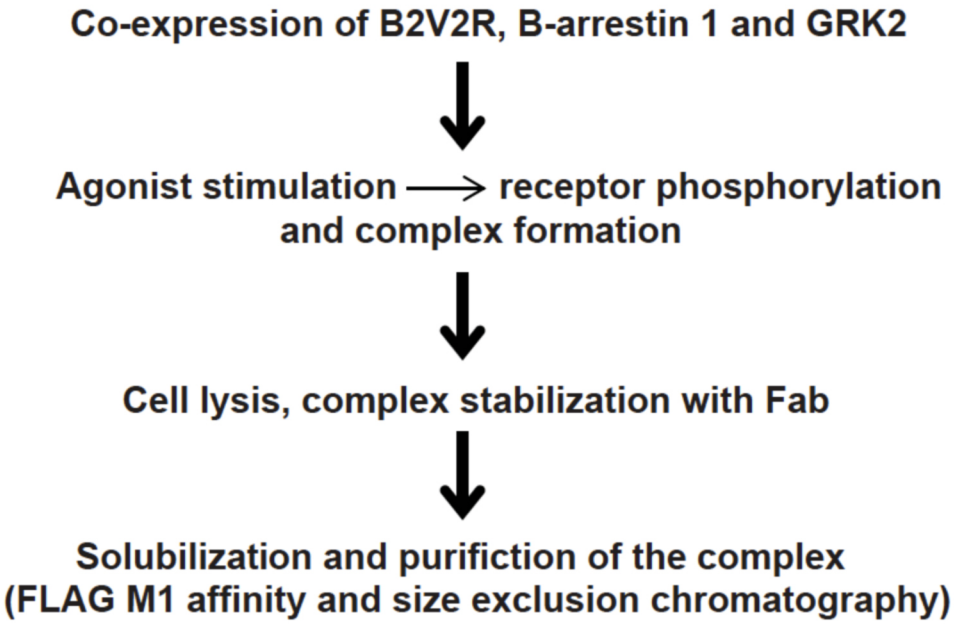


Figure 3-1. Purification scheme for β_2V_2-βarr-Fab(ScFv)30 complex.
(Figure courtesy of Arun Shukla)

3.3 Results

In celltro purification of β-arrestin/β_2V_2/Fab30 complex

Due to the transient nature and detergent sensitivities of β-arrestins and GPCRs in complex, we set out purify a stable β_2AR-β-arr complex and obtain detail structural insights in to this highly important signaling complex. However, our extensive attempts to form a stable β_2AR-β-arr complex from purified components *in vitro* were not successful, presumably due to transient nature of the complex or detergent sensitivity of purified β-arrestins. Therefore, we explored an alternative novel strategy to purify a pre-formed complex from the cells co-expressing the receptor and the βarr (Figure 3-1). We utilize a modified β_2AR construct (β_2V_2R) where the carboxyl terminus of β_2AR is replaced by the carboxyl terminus of vasopressin receptor, $AVPR_2$. This construct maintains the pharmacological properties identical to the β_2AR, however, it displays a "class B" profile of βarr interaction i.e. it binds to βarrs in a stable fashion compared to the β_2AR. We co-expressed the β_2V_2R, βarr 1 and $GRK2^{CAAX}$ in insect cells followed by agonist stimulation and purification of the complex through affinity purification of the tagged receptor. However, we were not able to isolate any detectable amount of β_2V_2R-βarr1 complex (Figure 3-2).

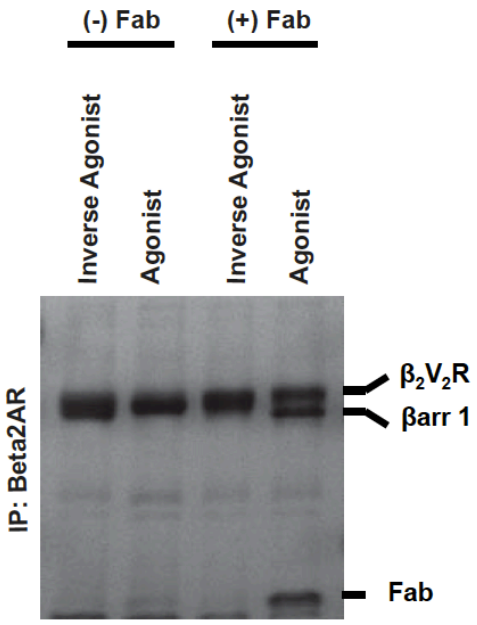


Figure 3-2. Co-immunoprecipitation revealing β_2V_2-βarr-Fab(ScFv)30 complex formation.

Stable complex formation of β_2V_2-βarr-Fab(ScFv)30 complex in the presence of receptor, βarr, Fab and agonist. No complex is formed in the presence of inverse agonist. (Figure courtesy of Arun Shukla)

We reasoned that the β_2V_2R-βarr complex might be dissociating during detergent solubilization and extensive washing during the purification steps and hypothesized that presence of a "stabilizing chaperone" during solubilization and purification steps might be able to secure the complex formation. We have recently reported an antibody fragment that selectively recognizes and stabilizes an active conformation of βarr 1 (Shukla et al., 2013). We reasoned that this Fab can stabilize the β_2V_2R-βarr complex in membrane and therefore, can maintain the integrity of the complex during solubilization and purification steps. Indeed, pre-incubation of this Fab with the pre-formed complex in the membrane resulted in robust yield of complex after purification only in response to agonist (BI-167107) but not an inverse agonist (ICI 118551) (Figure 3-3b). As one of the primary goals of this exercise is to purify the complex for structural studies, we tested in fact a number of Fabs, which are similar to Fab 30 in terms of stabilizing the active β-arrestin1 conformation, yield the stabilization of the complex, albeit to different levels. The degree of stabilization of the complex by different Fabs might reflect their relative affinities for the active β-arr1 conformation or their conformational epitope on β-arr1.

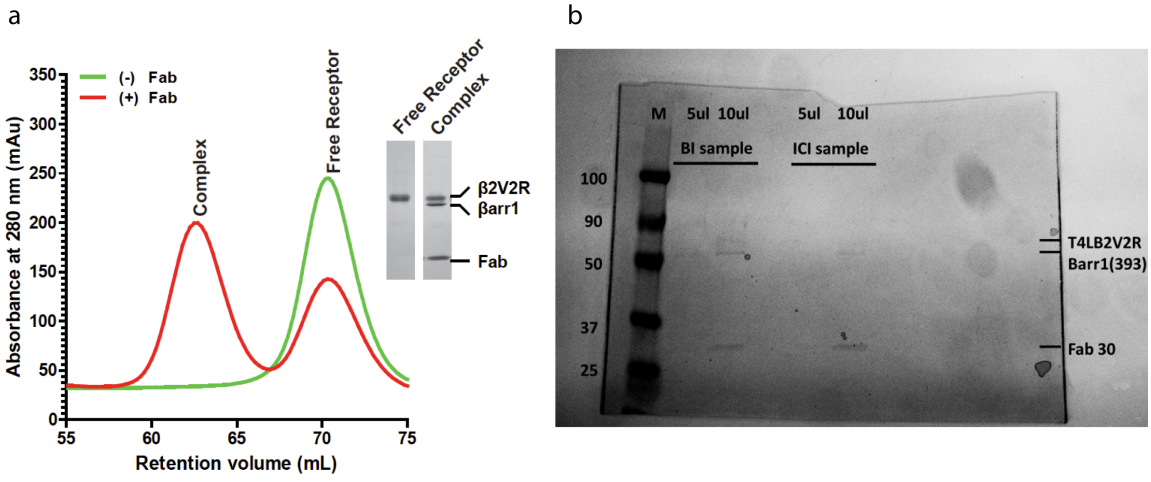


Figure 3-3. Stable β_2V_2-βarr-Fab(ScFv)30 complex formation.
a) Size exclusion chromatography (SEC) profile of β_2V_2-βarr-Fab(ScFv)30 complex formation. Red, the elution profile in presence of Fab pre-incubation. Green, elution

profile in the absence of Fab. b) Commassie blue stain of gel showing increased complex formation in presence of agonist (BI) versus inverse agonist (ICI). (Figure courtesy of Arun Shukla)

Strikingly, the efficiency of complex purification using this approach directly mirrors the pharmacological efficacy of the ligand used to stimulate the cells and that it also directly corresponds to the ligand occupancy of the receptor (Figure 3-4). This direct correlation of ligand efficacy and occupancy with complex purification efficiency reflects that the approach presented here yields a complex that involves an activated receptor conformation as well as receptor phosphorylation. Activation of βarrs upon interaction with 7TMR also leads to recruitment of clathrin to receptor-βarr complex and thereby facilitates clathrin-mediated endocytosis of receptors. Purified β_2V_2R-βarr-Fab complex also exhibited robust interaction with purified clathrin compared to βarr1 alone suggesting that βarr1 in this complex is in a physiologically relevant and functionally competent conformation. Interestingly, this purified β_2V_2R-βarr-Fab complex was extremely stable in solution, which is required for structural studies (Figure 3-5). As we were able to routinely and reproducibly prepare this biochemically stable, pharmacologically relevant and functionally competent signaling complex, we utilized electron microscopy based single particle analysis to obtain structural insights in to this complex.

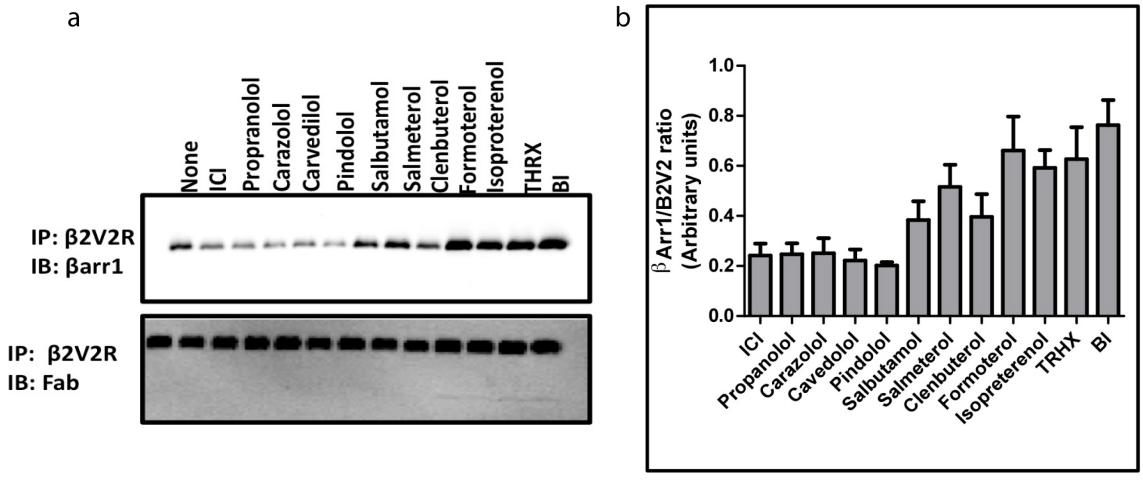


Figure 3-4. β_2V_2R-βarr-Fab complex formation relative to ligand used to stimulate cells.
a, Co-immunoprecipitation (Co-IP) for β_2V_2R and Western blotting for activated βarr1 (top) and Fab (bottom) using different ligands to stimulate cells. b, Graph illustrating ratios from β_2V_2R Co-IP of βarr1 alone to β_2V_2R with their respective ligand. (Figure courtesy of Arun Shukla)

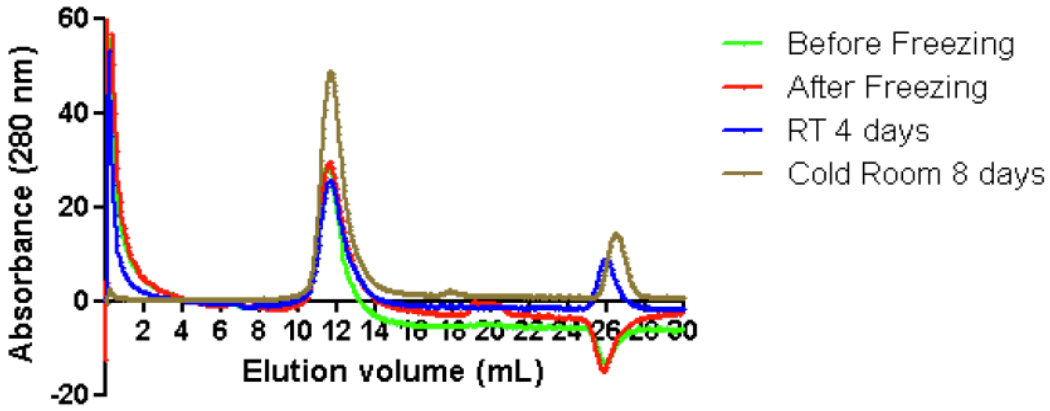


Figure 3-5. Size exclusion chromatograms of β_2V_2-βarr-Fab(ScFv)30 complex illustrating stability over time.
(Figure courtesy of Arun Shukla)

2D negative stain classification of native β-arrestin/β_2V_2/Fab30 complex

We sought to examine the architecture of the β_2V_2R-βarr-Fab30 complex. In a first step, we employed electron microscopy (EM) and single-particle analysis. Due to the limited size of the particle (~120 kDa) we visualized specimens embedded in negative stain, which provides sufficient contrast from relatively small protein assemblies. This approach enabled us to obtain 2D projection averages and 3D reconstructions that provided new insights into the composition of the β_2V_2R-βarr-Fab30 complex. EM visualization of the complex showed a monodisperse particle population (Figure 3-6).

Reference-free alignment and classification of ~16,500 particles revealed class averages that had a range of different conformations of the particle.

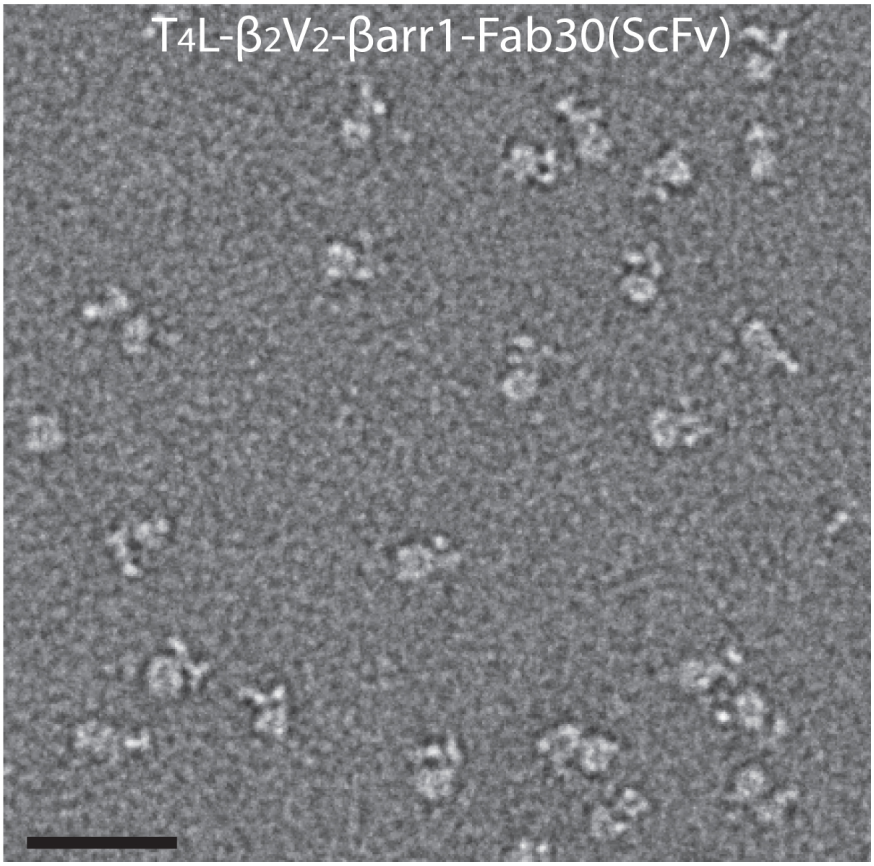


Figure 3-6. Negative stain image of β_2V_2-βarr-Fab(ScFv)30 complex.
Scale bar = 25 nm.

While the 2D averages revealed different conformations, the distinct features of the class averages allow us to assign the negative stain projection profiles from specific components of the complex. Based on previous EM data of receptor complexes, we can assign the central oval density to represent β_2V_2R in detergent micelle. There is no density from the T4-lysozyme (T4L) engineered to the C-terminus of the receptor in most of the class averages. This is due to the variable position of this flexible domain and its density being averaged out in the 2D classification. A T-like structure, containing 2 distinct domains, is attached to the receptor density with connectivity to one side. This single connection seems to lead to the inherent flexibility of the T-like structure (Figure

3-7, top left). Other recent structural studies have shown the end of Fab30 will co-crystalize to the middle of β-Arrestin1 (βarr1) (Shukla et al., 2013).

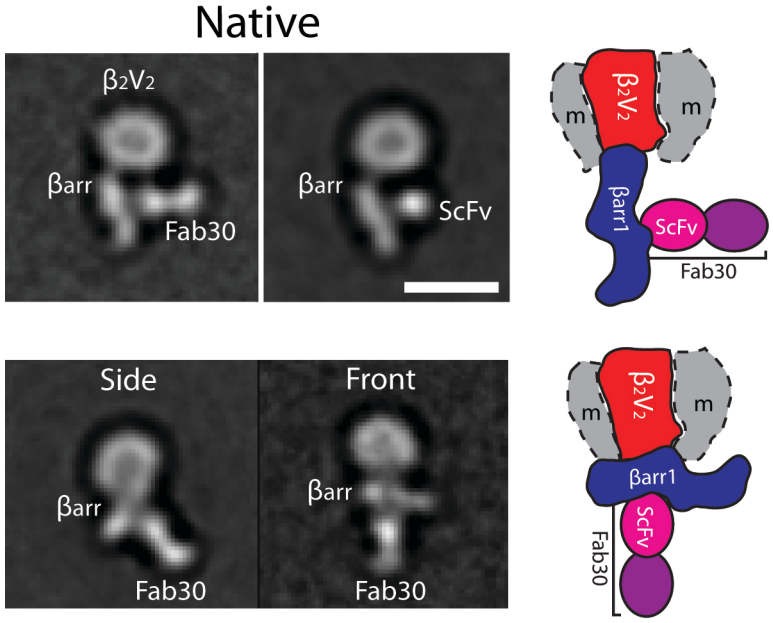


Figure 3-7. Representative 2D class averages of native β_2V_2-βarr-Fab(ScFv)30 complex.
Top, β_2V_2-βarr-Fab(ScFv)30 complex (left, Fab & right, ScFv) showing loosely associated βarr-Fab(ScFv) associated with the receptor. Bottom, side and front views of the β_2V_2-βarr-Fab30 complex with a more 'snuggly' association with the receptor. Scale bar = 10 nm.

This information, along with biochemical information that βarr1 should mediate the interaction with the receptor allows us to designate the domain connecting the T-like structure to the receptor to be βarr1, and the domain attaching perpendicular to it, to be Fab30. To further validate this domain interpretation, we next did 2D single particle analysis on the β_2V_2R-βarr-ScFv30 complex. Directly comparing similar conformations of the ScFv30 complex to the Fab30 complex, we see nearly identical 2D projections with the exception of the density assigned to be Fab30 halved in the ScFv30 complex, which is to be expected (Figure 3-7, top right). Comparing class averages with the βarr-Fab30 domains illustrating the T-like structure interacting more snuggly with the receptor to class averages that seem to have a smaller density for βarr, this can be attributed to a ~90º *z-axis* rotation of the particle's orientation on the carbon support (Figure 3-7,

bottom). The perspectives of these representative class averages will be termed front and side views, respectively.

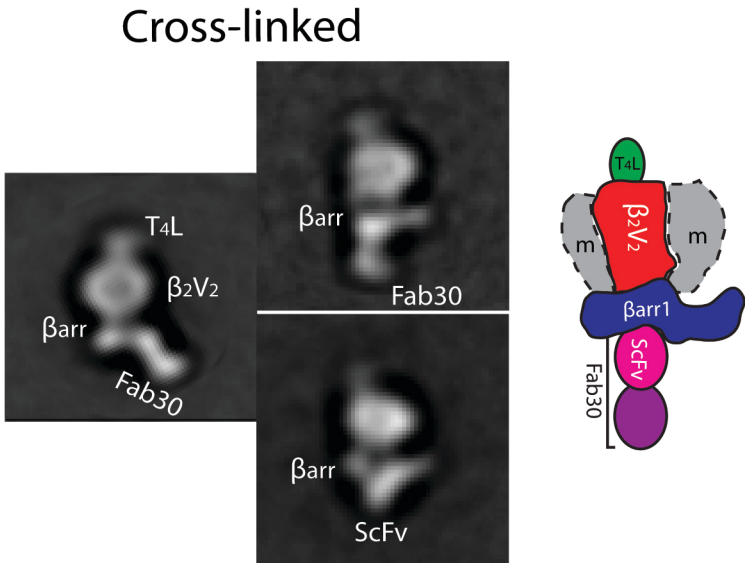


Figure 3-8. Representative 2D class averages of cross-linked β_2V_2-βarr-Fab(ScFv)30 complex.
Left, representative side view of cross-linked complex. Top right, representative front view of β_2V_2-βarr-Fab30 complex. Bottom right, representative front view of β_2V_2-βarr-ScFv30 complex

We next visualized the Fab30 and ScFv30 complexes after cross-linking to obtain a more uniform perspective of the complex. After characterizing ~15,000 particles of each respective complex with single-particle analysis, class averages revealed a more singular population. The class averages of the cross-linked complexes resembled projections that were very similar to the native complex (Figures 3-7 & 3-8). The cross-linking stabilizes the orientation of the receptor, evidenced by the presence of the T4L in most of the class averages relative to the native complex. To obtain a more detailed view of the complex architecture, we used the random conical-tilt approach to calculate initial 3D reconstructions of the cross-linked complex in its different orientations. 0° and 60° particle projections from individual class averages were used to calculate and refine 3D reconstructions for each of the Fab30 and ScFv complexes. This approach allowed us to obtain quality 3D reconstruction(s) (~30 Å @ FSC = 0.5) of the complex, particularly in the front view (Figure 3-9).

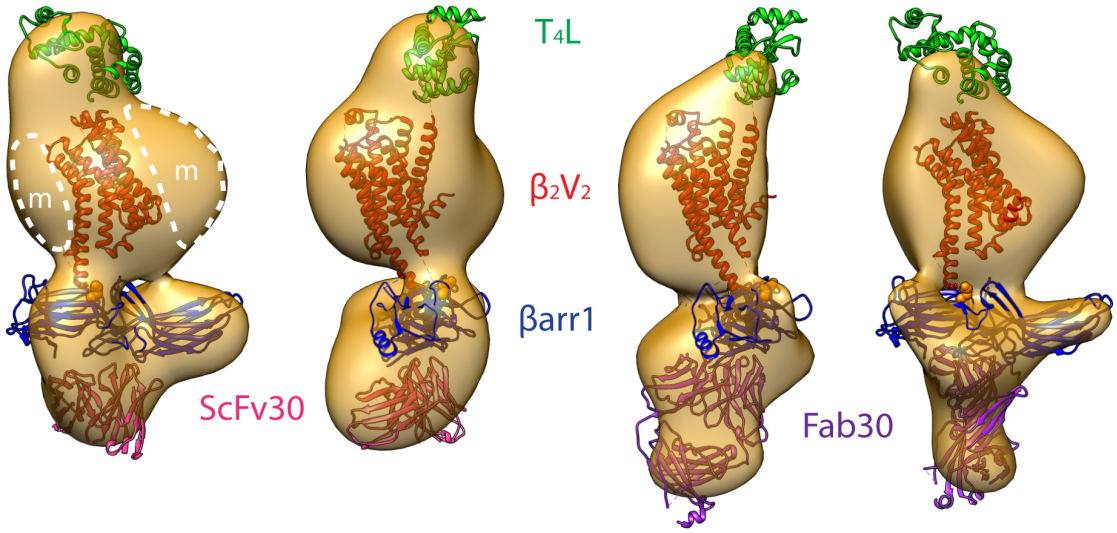


Figure 3-9. 3D reconstructions of β_2V_2-βarr-Fab(ScFv)30 complexes.
3D reconstructions of β_2V_2-βarr-Fab(ScFv)30 complexes, ScFv-right and Fab-left. Docked crystal structures ScFv30 in pink, Fab30 in purple, βarr in blue, β_2V_2 in red, and T4L in green. K77 of βarr is shown in orange as one of the residues cross-linked to β_2V_2 on the TM5 helix.

Molecular modeling of β-arrestin/β_2V_2/Fab30 complex

In May 2013, a crystal structure of activated βarr, phosphorylated vasopresssion C-terminal tail, and Fab30 was published. Using this recent structure of Fab-βarr complex, we tried to model these high-resolution structures into the EM maps. For both the Fab and ScFv complexes, there seems to be a reasonable fit of these two densities within each of the domains assigned from 2D negative stain analysis for each of the 3D maps. The cross-linked β_2V_2R-βarr-Fab30 complex was submitted to mass spectroscopy to identify which specific residues were being cross-linked. Lysine 77 (K77) of β-Arrestin1 (shown in orange, Figure 3-9) was identified as a cross-linking partner with residues K232 and K235 in the β_2V_2 receptor located near the intracellular end of transmembrane helix 5 (TM5). Additionally, deuterium exchange mass spectroscopy (HDXMS) revealed that upon receptor binding to arrestin regions consisting of the "finger loop," "middle loop," and "lariat loop," all had reduced deuterium exchange (Figure 3-10). This suggests that these three loops are most likely involved in the core interaction of forming this complex.

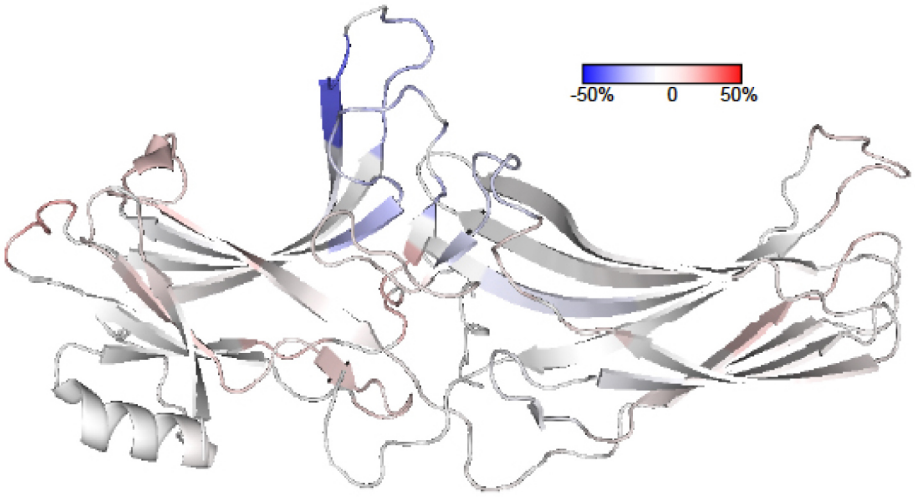


Figure 3-10. Hydrogen deuterium exchange mass spectroscopy (HDXMS) difference map of arrestin in complex with or without bound receptor.
HDXMS difference map between arrestin complex with or without bound receptor. Blue highlighted areas shows reduced deuterium exchange, and red highlighted areas show increased deuterium exchange after receptor binding to βarr.

Using this information, in addition to the extra density in the EM volume of the receptor from the T4-lysozyme, we can orient the receptor inside the detergent micelle. Keeping within these biochemical and steric restraints, we find a reasonable fit for all docked structures within the 3D volumes for both the Fab30 and ScFv30 complexes. From these findings, we generated a model of basic interactions between a GPCR and βarr. There is an initial interaction between the GPCR and the N-domain of βarr, and then there is a subsequent second interaction between the GPCR and the central polar core domain of βarr that allows Arrestin to have a more snuggly conformation on the receptor (Figure 3-11).

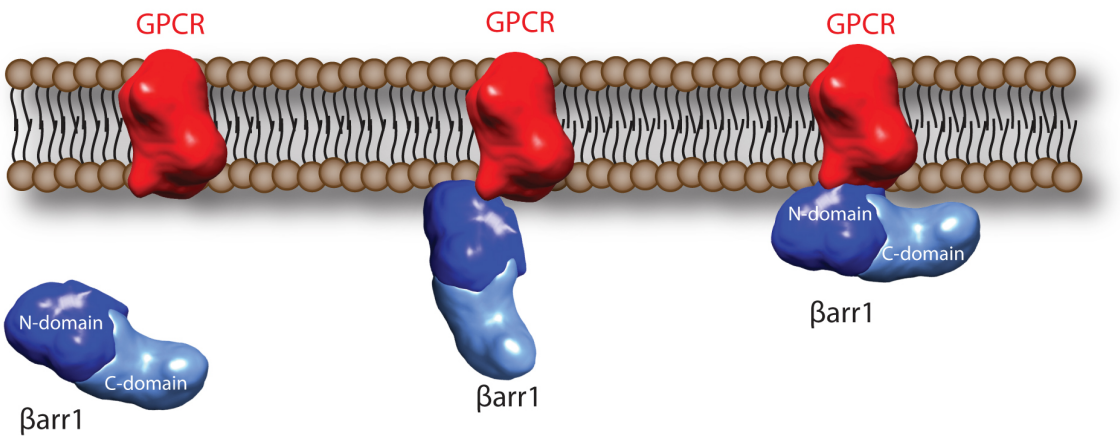


Figure 3-11. Model of β-arrestin binding to GPCR.
Illustration of interaction between βarr and a GPCR. There is an initial interaction between the GPCR and the N-domain of βarr (Figure 3-7, top), and then there is a subsequent second interaction between the GPCR and the central polar core domain of βarr that allows Arrestin to have a more snuggly conformation on the receptor (Figure 3-7, bottom).

In conclusion, we present the first architecture of a pharmacologically and functionally relevant GPCR-βarr signaling complex determined by electron microscopy. As the major signaling paradigms and networks of GPCR signaling are astonishingly well conserved among the members of this superfamily, the architecture presented here should serve as a general template to better understand the structural basis of GPCR desensitization and regulation.

3.4 Discussion

The interaction between G protein coupled receptors (GPCRs) and β-arrestin has been an important model to understand because of its significance in the silencing and G protein independent signaling processes. For the first time, this receptor/arrestin interaction can be characterized structurally due to the complex stabilization contributed from Fab or single chain antibody binding. This *in celltro* purification approach tremendously helped in obtaining high-quality, homogeneous β_2V_2R-βarr-Fab30(ScFv) complex for characterization through a number of different approaches.

In these studies, the architecture of β-arrestin in complex with a GPCR is revealed. The negative stain EM of both the Fab and ScFv stabilized GPCR-βarr1 complex helped illustrate the dynamic nature of the interface between arrestin and the receptor. The 2D class averages from native β_2V_2R-βarr-Fab30 complex showed a variety of different conformations between β-arrestin and the receptor. Primarily there were two types of conformers: the 'hanging chad' conformation where βarr appearing to be hinged at a single a single connection with the receptor (Figure 3-7, top), and a more 'snuggly' conformation where βarr is shown to be more tightly associated with the receptor (Figure 3-7, bottom). After cross-linking, more than 90% of the population consisted of this snuggly conformation (Figure 3-8). This could suggest that the interaction forming the 'hanging chad' conformation is the primary stronger interaction, with the remainder of the interface being a weaker more transient interaction that had become more stabilized from the cross-linking.

The hydrogen deuterium exchange mass spectroscopy (HDXMS) data suggests that the three loops near the core of β-arrestin are primarily involved in receptor binding due to the reduced amounts of deuterium exchange shown from the difference maps (Figure 3-10). These data significantly helped in properly orienting βarr and the receptor within the EM map. The fit, and more importantly the orientation, of βarr within the EM map shows a slight angular rotation of βarr relative to the receptor. The negative stain 3D reconstructions show a completely flattened complex, a very common result from the preparation of negative stain. However, the 2D class averages from cryo-EM reveal that there appears to be more of an angle between the Fab antibody and the receptor that would correlate with the HDXMS data (Figure 3-12).

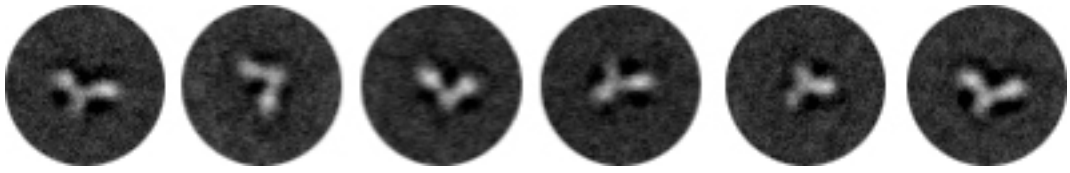


Figure 3-12. Representative 2D cryo-EM class averages of the β_2V_2R-βarr-Fab30 complex.
Representative 2D cryo EM class averages of the β_2V_2R-βarr-Fab30 complex revealing less flattening between the Fab and receptor densities.

While cryo-EM was performed on the cross-linked β_2V_2R-βarr-Fab30 complex, 3D reconstruction from those particle projections was unable to produce an EM map that would maintain a stable cross correlation value as it was iteratively. Some of the pitfalls of attempting cryo-EM with a specimen of such a small size (~125 kDa) is obtaining particle projections with enough contrast, and being able to computationally align the particle projections. The cryo-EM 2D class averages clearly show well-defined domains of the β_2V_2R-βarr-Fab30 complex, however due to the lack of large, more pronounced domains, when trying to align the particle projections it becomes easy for those particle projections to become misaligned. Additionally, the cryo-EM 2D class averages show some variation in the angle between the Fab and the receptor densities suggesting that the complex is inherently flexible, adding to the heterogeneity of the sample. Ideally cryo-EM would be the preferred technique to characterize this complex as supposed to negative stain EM, because it would preserve the specimen in a more native state.

Electron microscopy is a useful technique to obtain low to intermediate resolution of a protein complex's architecture. X-ray crystallography can achieve a much higher resolution, however, a very homogeneous, stable specimen is needed for crystal nucleation and formation. The native β_2V_2R-βarr-Fab30 complex showed a significant amount of heterogeneity, and even after cross-linking the cryo-EM 2D class averages reveal angular differences between Fab and the receptor. The goal for any structural characterization of a specimen is to obtain the highest resolution possible (from an x-ray crystal structure, for example), but there will need to be additional modifications in the formation of the β_2V_2R-βarr-Fab30 complex in order to achieve a more stable specimen. It has been shown that advances in protein engineering and binding of different antibodies can help stabilize GPCR-protein complexes as seen in the β_2AR-Gs complex (Rasmussen et al., 2011).

In these studies a number of different approaches were used to characterize the complex formed between a GPCR and β-arrestin. Negative stain and cryo-EM have helped illustrate the dynamic and transient nature of these proteins, and this has given better understanding the role of β-arrestin in GPCR binding and silencing. Furthermore, the

findings here can aid in future endeavors to promote even better characterization of this relationship.

3.5 Experimental Procedures

Protein expression and purification.

T4L-β_2V_2R (with N-terminal fusion of T4 lysozyme). T4L-β2V2R was co-expressed with untagged β-arr1 and untagged $GRK2^{CAAX}$ in *Sf*9 cells. 60-66h post-infection, cells were stimulated for 30 min (or indicated time) at 37°C with indicated ligand to induce receptor phosphorylation and ternary complex formation. Subsequently, cells were harvested, lysed in 20mM Hepes, pH 7.4, 150mM NaCl, 100nM-1uM ligand and protease inhibitors by douncing and purified Fab was added to this cell lysate. Following 1h incubation at room temperature, proteins were solubilized by adding indicated detergent (0.5% MNG or 1% DDM) and gentle stirring at room temperature for 1h. Solubilized material was purified on a FLAG M1 resin essentially using the protocol described for β_2AR purification previously. In order to avoid potential protein aggregation at high concentration, free cysteines in the eluted proteins were blocked using iodoacetamide as described previously. Proteins were concentrated using a 100 kDa Vivaspin or Millipore concentrator and loaded on to a preparative size exclusion column (Superdex 200, 16/600). The column was run at 0.3 ml/min and the column running buffer consists of 20mM Hepes, pH 7.4, 150mM NaCl, 100nM BI-167107, 0.01% MNG, 100μM TCEP). Elution fractions corresponding to the complex peak were either pooled or used individually for further studies. Selection, characterization and purification of this Fab has been described previously.

Biochemical and functional characterization of the complex.

In order to pharmacologically validate this novel procedure, we stimulated the cells with multiple ligands of varying efficacies (ranging from inverse agonists, partial agonists to full agonists). Subsequently, the cells were lysed, incubated with the Fab and the complex was purified using FLAG M1 affinity resin. Purification of complex was assessed by the extent of β-arr1 co-purified with the receptor. In order to further investigate agonist

dependence of the complex formation, we assessed whether formation of this ternary complex through Fab stabilization follows ligand occupancy of the receptor. Cells were stimulated with varying doses of the agonist BI167107 or Inverse agonist ICI-118551 at 37°C for 30 min. Subsequently, purification of the complex was performed as described above. Similarly, a time course of agonist and inverse agonist stimulation was carried out to identify the optimal time point to achieve maximal complex formation. In order to validate the active conformation of βarr 1 in this ternary complex, we evaluated the ability of this complex to interact with purified clathrin heavy chain (a kind gift from Prof. Tom Kirchhausen at Harvard Medical School). Purified complex (1μM in 20mM HEPES, 150mM NaCl, 0.001% MNG, 100nM BI) was incubated with purified clathrin heavy chain (3μM) at room temperature for 1h. Subsequently, the Fab in the complex was immunoprecipitated using Protein A agarose and co-elution of clathrin was visualized by Western blotting. To assess the biochemical stability of the complex over time, purified complex was stored at 4°C for indicated period of time and subsequently analyzed by analytical size exclusion column (superdex 200, 10/300).

Glutaraldehyde cross-linking of the pre-formed complex.

In order to mildly cross-link the complex, an "on-column" cross linking method was used. First, a bolus of glutaraldehyde was injected to a pre-equilibrated Superdex 200 (10/300 global) column and run at 0.25 ml/min for 20 min (i.e. a total of 5ml buffer). Subsequently, the column flow was paused, injection loop was flushed using buffer followed by injection of purified complex (200μl volume, at 5-10 μM concentration). Subsequently, the column was run at 0.25 ml/min and 0.3 ml fractions were collected. Cross-linking efficiency was visualized by running the individual fractions on a 4-20% SDS gel, followed by staining with simply blue stain. Various concentration of the glutaraldehyde were tested in order to identify an optimal percentage (0.25 % in this case) which yields maximal cross linking efficiency and minimal protein aggregation.

Hydrogen-Deuterium Exchange Mass Spectrometry

Fragmentation tuning for β-arr1

In order to optimize experimental conditions to achieve the best peptide fragmentation pattern for β-arr1, we first performed fragmentation tuning experiments. In brief, β-arr1 protein solution was diluted with 3 volumes of a buffer containing 150 mM NaCl, 8.3mM Tris-HCl, pH7.2 in H2O and then divided into equal sized aliquots. Each aliquot was mixed individually with 6 volumes of a series of quench solutions containing 0.8% formic acid, 16.6% glycerol, and various concentrations of GuHCl (ranging from 0.05M to 4M) and TCEP (ranging from 0.015M to 0.5M). The protein/buffer/quench solution mixtures were incubated at 0°C for varying times and subjected to on-line pepsin digestion and LC/MS analysis. Protein fragmentation maps under different conditions were generated and compared. The condition that produced the best peptide fragmentation pattern (the best coverage along the amino acid sequence, the most number of fragments, and highest number of high quality peptides) was used for all the HDXMS experiments described in this study.

Hydrogen-deuterium exchange experiments

The V2Rpp/β-arr1/Fab30 and β2AR/β-arr1/Fab30 complexes were prepared as described above. Free β-arr1 protein was used as a control. For each protein complex, three sets of samples were prepared: 1) non-deuterated (ND), 2) fully deuterated (FD), and 3) time-dependent on-exchange samples. A protein: buffer: quench solution ratio of 1:3:6 (volume) was used for all sample preparations. The FD sample sets were prepared by mixing protein samples with D2O buffer (150mM NaCl, 8.3 mM Tris-HCl, pH7.2 in D2O) and incubated at room temperature for 12 hours before quenching. The ND sample sets were prepared using a similar procedure with H2O buffer (150 mM NaCl, 8.3 mM Tris-HCl, pH7.2 in H2O) without the incubation step. The on-exchange sample sets were prepared by adding 3 volumes of D2O buffer at 0°C and incubating for varying time-points (10s, 10^2S, 10^3S, 10^4S and 10^5S). Six volumes of ice-cold quench solution was then added to each sample, followed by snap-frozen on dry ice and stored at -80 °C.

On-line pepsin digestion, LC/MS analysis and data processing

Samples prepared above were thawed right before the on-line pepsin digestion at 0°C using an immobilized porcine pepsin column. Peptide fragments were collected

contemporaneously on a C18 trap column desalting, separated by a Magic C18AQ column using a linear gradient of acetonitrile from 6.4% to 38.4% over 30 min, and followed by LC/MS analysis using an Orbitrap Elite mass spectrometer (Thermo Scientific). Both MS1 and MS2 spectra were collected using data-dependent acquisition strategy. Peptide identification was performed using LC/MS data sets collected from ND samples and SEQUEST database search engine using Proteome Discoverer (Thermo Fisher Sci). The SEQUEST database search results were submitted to DXMS Explorer (Version 2.0, Sierra Analytics Inc.), filtered using several threshold parameters to create an initial peptide pool. The quality of the MS1 data for each filtered peptide was then checked by assigning an initial quality score by DXMS Explorer software, followed by a quality control process which includes manual investigation of peak isotopic envelope and adjusting/improving the quality score. Only peptides with high quality in the MS1 spectra were kept in the final peptide pool. The retention times and m/z ranges of each peptide from the final peptide pool was manually verified and adjusted across all LC/MS data sets from on-exchange samples and FD samples to ensure that DXMS Explorer has select the correct peptide for all experiments. Results from FD samples were used to monitor the back-exchange rates during on-line pepsin digestion and LC/MS analysis. The centroids of isotopic envelopes of non-deuterated, partially deuterated and fully deuterated peptides were measured using DXMS Explorer, and then converted to deuteration level with corrections for back-exchange. A deuterium accumulation plot was created for each peptide as a further quality check and data refinement process. Rainbow maps were generated and DXMS data comparison was performed using different macros in Excel.

Chemical Cross-linking and Mass Spectrometry

Chemical cross-linking reaction

We first performed protease digestion and LC/MS analysis for the gluteraldehyde cross-linked β_2V_2R-βarr-Fab30 complex. However, we were not able to detect any cross-linked peptides between the β_2V_2R and βarr. To facilitate the identification of cross-linked peptides, we used an equimolar mixture of protonated (d0) and deuterated (d6) Disuccinimidyl glutarate (DSG, a homobifunctional amine-reactive crosslinker) to cross-

link the β_2V_2R-βarr-Fab30 complex. The cross-linked peptides are characterized with "doublet" peak signatures in mass spectra. The β_2V_2R-βarr-Fab30 complex was prepared in buffer containing 20 mM HEPES, 150 mM NaCl, 0.01% MNG, 100nM BI, 100uM TCEP. Cross-linker solution was freshly prepared in DMSO at a concentration of 10 mM. For the cross-linking reaction, 100-fold excess of DSG-d0:DSG-d6 equimolar mixture was added to the β_2V_2R-βarr-Fab30 complex solution and incubated for 30 min at 25 °C. Unreacted cross-linker was quenched by incubation with 100 mM ammonium bicarbonate for 20 min. The cross-linked β_2V_2R-βarr-Fab30 complex was separated by SDS-PAGE, and the corresponding protein band was in-gel digested overnight at 37 °C with trypsin at a final concentration of 10 ng/μL and subjected to LC/MS analysis as described below.

LC/MS analysis and cross-linked peptide identification

LC/MS analyses were performed on a LTQ Orbitrap XL mass spectrometer (Thermo Scientific) with a Finnigan Nanospray II electrospray ionization source. Tryptic peptides were injected onto a 75 μm × 150 mm BEH C18 column (particle size 1.7 μm, Waters) and separated using a Waters nano ACQUITY Ultra Performance LC™ (UPLC™) System (Waters, Milford, MA). The LTQ Orbitrap XL was operated in the data dependent mode using the TOP10 strategy. In brief, each scan cycle was initiated with a full MS scan of high mass accuracy [375–1800 m/z; acquired in the Orbitrap XL at 6×10^4 resolution setting and automatic gain control (AGC) target of 10^6], which was followed by MS/MS scans (AGC target 5000; threshold 3000) in the linear ion trap on the 10 most abundant precursor ions. Selected ions were dynamically excluded for 30 s. Singly charged ions were excluded from MS/MS analysis. The LC/MS/MS data was processed and analyzed by pLink software. In brief, acquired raw data from LTQ Orbitrap XL mass spectrometer was first be converted to mgf files using Mascot Distiller or Proteowizard 3.0 and saved under a pLink master directory, where a composite database containing protein sequences of β_2V_2R, βarr and Fab30 in FASTA format is also stored. In the pLink configuration file, both the names and paths of the mgf and database file were indicated correspondingly. The total numbers of fixed modifications will be set to 1 for cysteine carbamidomethylation. The name and number of cross linkers used in the experiment

were indicated as well in the configuration file. The filter for precursor mass accuracy was set to ± 10 ppm. After pLink analysis, identified inter-cross linked peptides were manually inspected in the raw file for the appearance of a pair of doublet MS peaks.

Specimen preparation and EM imaging of negative-stained samples.

β_2V_2R-βarr-Fab30 complex was prepared for electron microscopy using the conventional negative staining protocol (Ohi et al., 2004), and imaged at room temperature with a Tecnai T12 electron microscope operated at 120 kV using low-dose procedures. Images were recorded at a magnification of 71,138x and a defocus value of ~1.5μm on a Gatan US4000 CCD camera. All images were binned (2 x 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. All particles were manually excised using Boxer [part of the EMAN 1.9 software suite] (Ludtke et al., 1999) apart from tilt pairs where particles were selected using WEB (Frank et al., 1996).

2D classification and 3D reconstructions of negative stained β_2V_2R-βarr-Fab30(ScFv30) complex.

2D reference-free alignment and classification of particle projections was performed using SPIDER (Frank et al., 1996). 16,286 particles of native β_2V_2R-βarr-Fab30 complex were classified to 150 classes. For the native β_2V_2R-βarr-ScFv30 complex, 13,703 particles were separated into 100 classes.

Crosslinked β_2V_2R-βarr-ScFv30 complex had 14,109 particles that were initially separated into 150 classes, and then an additional secondary classification for particle projection of the front view was performed. The secondary classification of front view β_2V_2R-βarr-ScFv30 particles had 3,865 particles separated into 8 classes. For the crosslinked β_2V_2R-βarr-Fab30 complex, a total of 13,106 particle pairs were interactively selected from 0º and 60º tilted images (100 pairs). The untilted particle images were subjected to 10 cycles of reference-free alignment and classification into 50 classes.

For both Fab30 and ScFv30 complexes the random conical tilt technique (Radermacher et al., 1987) was used to calculate a first back projection map from individual classes using the images of the tilted specimen. After angular refinement, the corresponding particles

from the images of the untilted specimens were added, and the images were subjected to another cycle of refinement. We thus generated reliable initial models for complexes in variable orientations. After contrast transfer function (CTF) correction according to local defocus values obtained by CTFTILT (Mindell and Grigorieff, 2003), particle projections of the individual class averages were subjected to further refinement and alignment in SPIDER. For the β_2V_2R-βarr-ScFv30 3D reconstruction 1,364 particle projections from 0º and 60º images contributed to the map. For the β_2V_2R-βarr-Fab30 complex, 1,050 particle projections from 0º and 60º images contributed to the map.

Cryo-EM specimen preparation and imaging.

2 μl of β_2V_2R-βarr-Fab30 cross-linked complex were adsorbed on glow-discharged Quantifoil R2/2 200 mesh grids, and vitrified using a Vitrobot (FEI Mark IV). The specimen was visualized on a Tecnai F20 electron microscope (FEI) equipped with a field emission electron source operated at 120kV. Images were recorded on a Gatan US4000 couple-charge device (CCD) camera under low-dose conditions, at a nominal magnification of 66,964x, and defocus values ranging from -2.0 to -4.0 μm.

Cryo-EM 2D classification of β_2V_2R-βarr-Fab30 complex.

Particles from cryo-EM images were excised using Boxer [part of the EMAN 1.9 software suite] (Ludtke et al., 1999). The CTF parameters were determined for each micrograph using *ctfit*, and CTF correction was applied accordingly using *Applyctf* [part of EMAN 1.9 package]. A total of 22,257 particle projections were subjected to 2D reference-free alignment and classification by refine2d.py [EMAN 1.9 software suite] separated into 300 class averages.

3.6 Acknowledgements

I would like to thank Arun Shukla and the many other members of the Lefkowitz lab for their expertise in purifying the β_2V_2R-βarr-Fab30(ScFv) complexes, performing other biochemical assays, and mass spectroscopy. Annie Dosey was a tremendous help in all

of the particle picking, and other data processing. At time of thesis submission, the works in this chapter were in preparation to be submitted for publishing.

3.7 References

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.

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.

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.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol *142*, 334-347.

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.

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.

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.

Shukla, A.K., Manglik, A., Kruse, A.C., Xiao, K., Reis, R.I., Tseng, W.C., Staus, D.P., Hilger, D., Uysal, S., Huang, L.Y*., et al.* (2013). Structure of active beta-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. Nature *497*, 137-141.

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

4.1 Abstract

Histone H3 lysine 4 (H3K4) methylation is catalyzed by the highly evolutionarily conserved multiprotein complex known as Set1/COMPASS or MLL/COMPASS-like complexes from yeast to human, respectively. Here we have reconstituted fully functional yeast Set1/COMPASS and human MLL/COMPASS-like complex *in vitro* and have identified the minimum subunit composition required for histone H3K4 methylation. These subunits include the methyltransferase C-terminal SET domain of Set1/MLL, Cps60/Ash2L, Cps50/RbBP5, Cps30/WDR5, and Cps25/Dpy30, which are all common components of the COMPASS family from yeast to human. 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 Cps50 and Cps30 localizing on the top two adjacent lobes and Cps60-Cps25 forming the base at the bottom. EM analysis of the human complex reveals a striking similarity to its yeast counterpart, suggesting a common subunit organization. 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. Additionally, Cps40 has been shown to regulate trimethlyation, and has been localized near the WD40 domains of Cps30 & Cps50. This structural arrangement suggested that COMPASS family members function as exo-methylases, which we have confirmed by *in vitro* and in *vivo* studies.

4.2 Introduction

Methylation of histone lysines, including H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20, play a crucial role in the regulation of key biological process, such as cell cycle progression, transcription, and DNA repair (Bhaumik et al., 2007; Shilatifard, 2006;

Zhang and Reinberg, 2001). Except for H3K79 methylation, which is catalyzed by the Dot1 family proteins, all other histone lysine methylations are carried out by the SET (Su (var), Enhancer of Zeste, and Trithorax) domain-containing enzymes. While the majority of the SET domains can act as histone methyltransferases in an isolated form, the MLL/Set1 family methyltransferases and Polycomb group proteins, which respectively catalyze histone H3K4 and H3K27 methylation, must assemble within their respective complexes for maximal catalytic and biological activities (Shilatifard, 2006).

The founding member of the MLL/Set1 family protein, Set1, forms a multiprotein complex named COMPASS (COMplex of Proteins ASsociated with Set1) in *Saccharomyces cerevisiae* (Miller et al., 2001). Set1/COMPASS was the first identified histone H3K4 methylase capable of mono-, di-, and trimethylating H3K4, and several of the COMPASS subunits are required for proper methylation (Krogan et al., 2002; Miller et al., 2001; Roguev et al., 2001; Schlichter and Cairns, 2005; Schneider et al., 2005). In addition to the evolutionarily conserved SET domain located at the C-terminus of Set1, most associating subunits are also conserved from yeast to human, forming Set1/COMPASS and MLL/COMPASS-like complexes (Eissenberg and Shilatifard, 2010; Shilatifard, 2008).

Histone H3K4 methylation on chromatin, in particular di- and tri-methylation, correlates with actively transcribed genes and is introduced by the COMPASS family of H3K4 methylases (Eissenberg and Shilatifard, 2010; Shilatifard, 2008). Previous work by us and others has shown that H3K4 methylation is subject to various layers of regulation mechanisms that are highly conserved from yeast to human (Shilatifard, 2008). Different subunits of yeast and human COMPASS have also been shown to regulate H3K4 di- and/or trimethylation as well as Set1 stability. This includes Cps60 (Ash2L in human), Cps50 (RbBP5), Cps40 (Cfp1), Cps30 (WDR5), and Cps25 (Dpy30).

The recently determined crystal structure of the MLL1/SET domain in complex with histone H3 peptide substrate (Southall et al., 2009) revealed that the active site was relatively solvent exposed due to an outward shifted inserted SET motif (iSET). The isolated MLL1/SET domain, free of any associated components of the natural MLL1 complex, exhibits a weak *in vitro* histone methyltransferase activity (Southall et al.,

2009). However, the addition of residual components of the MLL1 complex, namely WDR5, RbBP5, Ash2L, and Dpy30, greatly stimulated catalytic activity. This observation led to a model, in which the association of the MLL1 complex subunits induces reorientation of iSET to the optimal conformation of its catalytic active site (Southall et al., 2009). Given the highly conserved C-terminal catalytic SET domain between yeast Set1 and human MLL1, and the common subunit composition of the COMPASS family, the above model might be applicable to yeast COMPASS as well. However, until now there has been no structural information regarding the architectural arrangement of any COMPASS or COMPASS-like complex.

The roles of the WD40 repeat proteins, Cps50 and Cps30, in regulating H3K4 methylation has remained elusive. Furthermore, additional evidence for the role of Cps40 further regulating trimethylation of H3K4 has been recently found (Takahashi et al., 2009). It has been shown that Tyr1052 on Set1 plays an essential role in the regulation of H3K4 trimethylation regulation, and a mutation of this Tyr to Phe can suppress the loss of Cps40 in H3K4 trimethylation levels. This gives further evidence that Cps40 serves an important role in trimethylation regulation as well.

To better understand the organization and enzymatic regulation of the COMPASS family, we sought to reconstitute and characterize yeast and human COMPASS and to identify the minimal core complex required for the full histone H3K4 methylase activity. Here we present the identification, reconstitution, as well as the structural and biochemical characterization of such active core COMPASS for both yeast and human complexes capable of high levels of H3K4 methylase activity towards histone H3.

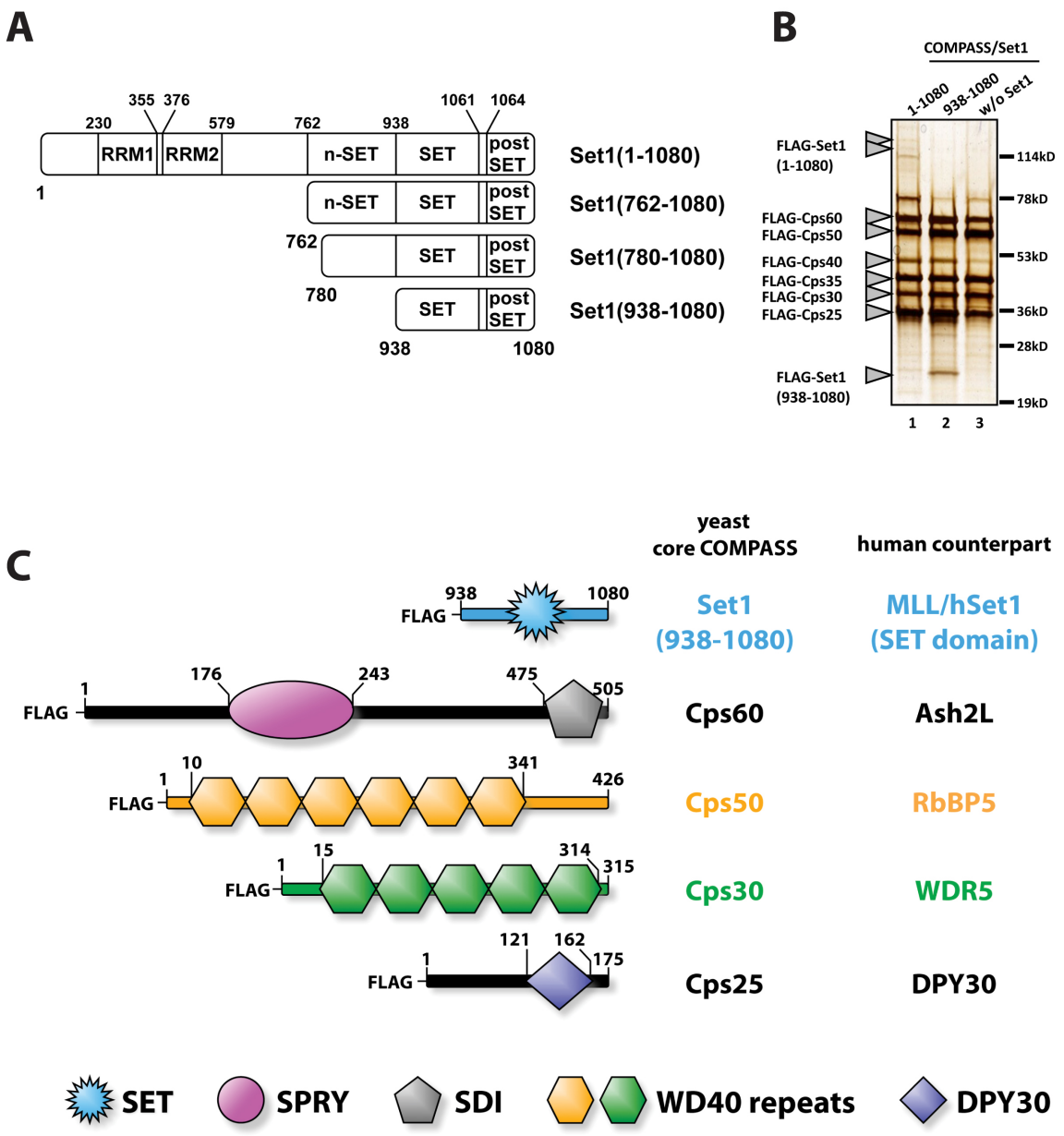


Figure 4-1. Schematic of core COMPASS complex

A) schematic diagram of N-terminally truncated Set1 proteins. Domain structures of each Set1 construct are denoted. B) SDS-PAGE analysis of FLAG affinity purified COMPASS complexes. In addition to Cps60, Cps50, Cps40, Cps35, Cps30, and Cps25, Set1 was expressed as full-length (lane 1), or 938-1080 including SET and post-SET domains (lane 2), or was omitted (lane 3). C) Schematic diagram of the components of yeast core COMPASS visualized by EM. The positions of predicted or confirmed domain structures are shown along their amino acid numbers. Figure courtesy of the Shilatifard Lab.

4.3 Results

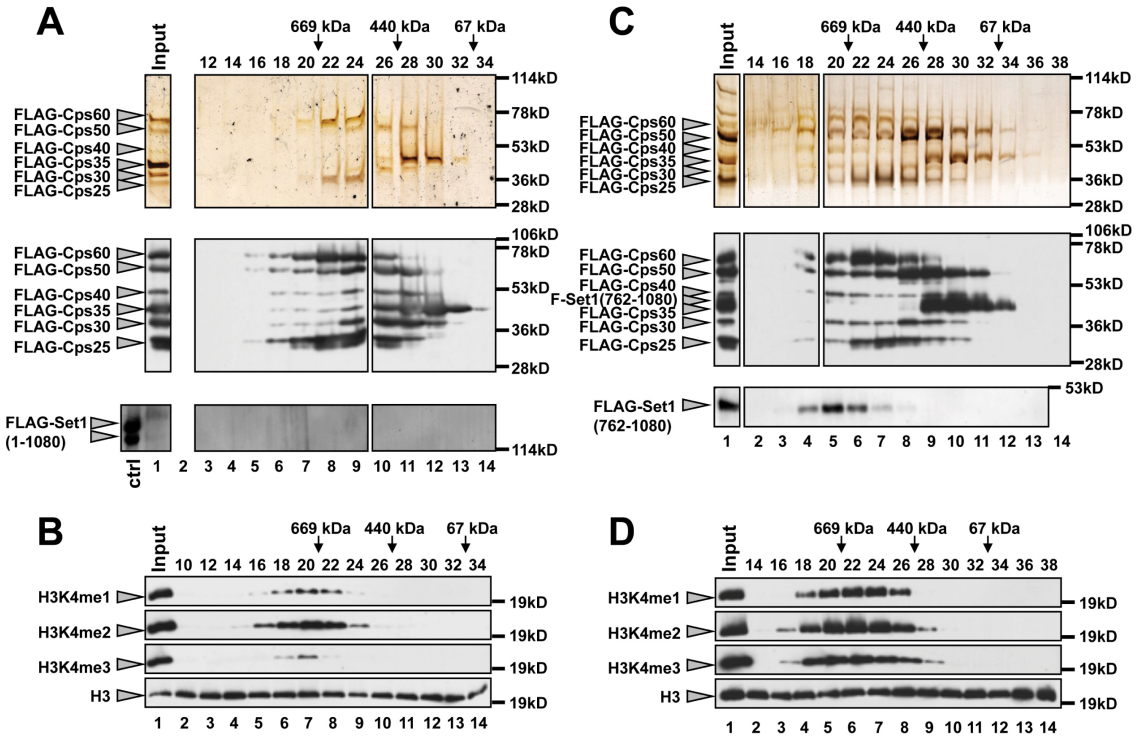
COMPASS complex reconstitution

To obtain stable COMPASS for biochemical characterization, we reconstituted the yeast and human complex through baculovirus-mediated co-transfection and overexpression of different subunit combinations in insect cells. Based on affinity purification of the recombinant Set1 and Cps proteins, we obtained nearly pure preparations of reconstituted COMPASS complexes (Figures 4-1 to 4-4) that are enzymatically active in *in vitro* H3K4 methylation assays with full-length histone H3 (Figure 4-5A). This approach enabled us to prepare any combination of wild-type (WT) or modified COMPASS components, including truncated forms of Set1, for studies that have been previously hampered by the absence of subunits such as Cps50 or Cps30(Schneider et al., 2005). Accordingly, COMPASS containing full length Set1, truncated Set1 (762-1080), Set1 (780-1080), Set1 (938-1080) or without Set1 were successfully prepared (Figure 4-1 and 4-2). Curiously, full length Set1 purified from yeast COMPASS consistently runs on SDS/PAGE as a doublet, although the reasons behind this property of Set1 or its biological significance remain unclear.

COMPASS histone methylase activity

As a first step, we investigated the effects of specific Cps subunits on *in vitro* H3K4 methylation by preparing full-length Set1 (1-1080) with various Cps combinations (Figure 4-1A). Each introduced component was purified through FLAG affinity purification as shown by anti-FLAG or anti-Set1 Western blots (Figure 4-5A, top and middle panels, respectively). Consistent with our previous reports of *in vivo* and *in vitro* H3K4 methylation using cps60Δ and cps25Δ strains (Schneider et al., 2005), the absence of Cps60 and Cps25 significantly impaired the catalytic activity of Set1 with the total loss of H3K4 trimethylation, and more than an 80% reduction of H3K4 mono- and dimethylation (Figure 4-5A, lane 7). In contrast to earlier *in vivo* studies (Dehe et al., 2006; Lee et al., 2007; Nedea et al., 2008; Schneider et al., 2005), almost no mono-methylation or very marginal di- and trimethylation changes of H3K4 were observed in the absence of Cps40 and Cps35 under conditions where Set1 levels are not a limiting

reagent in our *in vitro* experiments (Figure 4-5A). This suggests that full-length Set1 stability *in vivo* is reduced in either cps40Δ or cps35Δ strains, compared to the strain expressing all five subunits. To this end, we found that Cps40 co-eluted with other COMPASS components in the presence of the N-terminally extended region of Set1, including full-length Set1, Set1 (762-1080), Set1 (780-1080), but not Set1 (938-1080) (Figure 4-2). Cps35 was also shown to require a region beyond 762-1080 of Set1, because it did not co-elute in the macromolecular fractions with any other truncated Set1 variants (Figure 4-2).



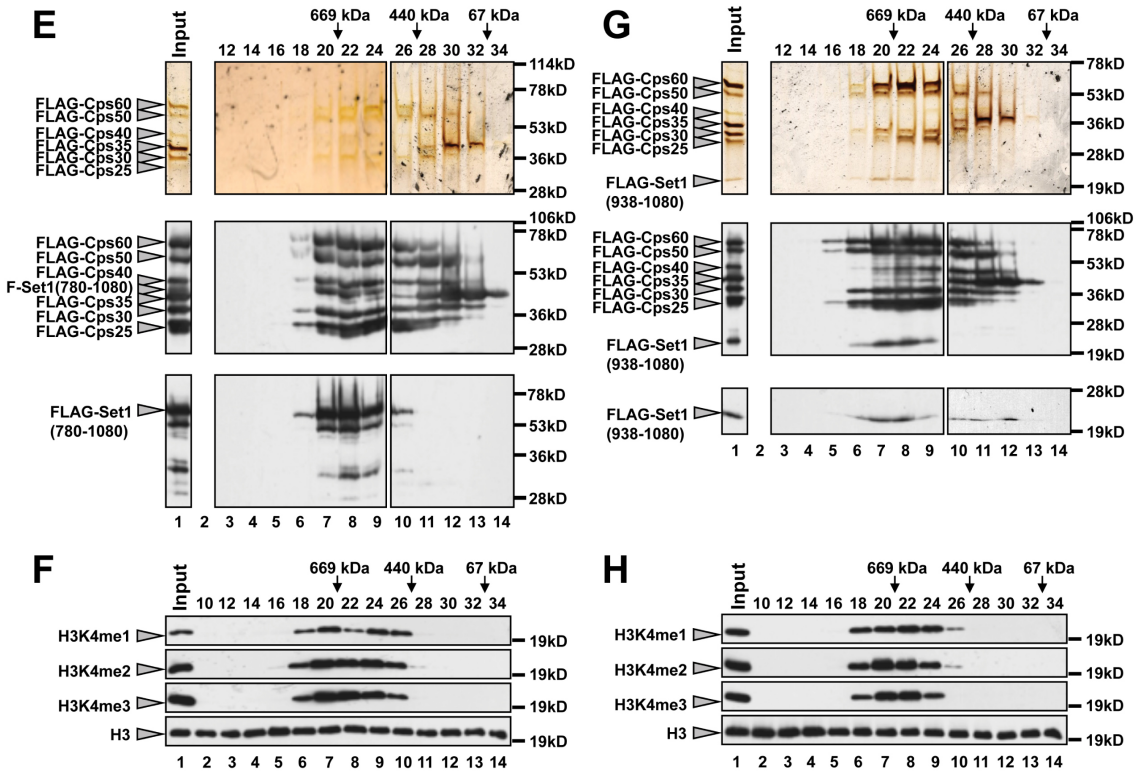


Figure 4-2. Recombinant COMPASS containing full-length or N-terminally truncated Set1 and their in vitro H3K4 methylation activities.

To investigate a region minimally required within Set1 polypeptide for COMPASS complex formation and in vitro H3K4 methyltransferase activity, full-length Set1 (A and B), Set1(762-1080) (C and D), Set1 (780-1080) (E and F), and Set1(938-1080) (G and H) were co-expressed with all Cps subunits except Cps15 in Sf9 insect cells and prepared by FLAG affinity purification. Purified samples were applied onto Superose 6 gel filtration column to separate the formed COMPASS complexes and unformed subunits along fractions denoted on the top of each figure. Each fractionated COMPASS complex was characterized by silver-staining (top boxes of (A), (C), (E) and (F)), western blotting by using anti-FLAG (middle panels of (A), (C), (E) and (F)) or anti-Set1 antibody (bottom boxes of (A), (C), (E) and (F)). In vitro H3K4 methylation activities toward free histone H3 of each fraction were examined by western blotting using anti-H3K4me1, me2, and m3 specific antibodies (B, D, G, and H). It is noted that COMPASS complexes eluted around fraction 18 through 20 exhibiting H3K4 methyltransferase activities irrespective of used Set1 species. Importantly, in those fractions, full-length Set1 containing COMPASS associated with all of Cps subunits whereas Cps40 was dissociated from Set1(938-1080) containing COMPASS and Cps35 was so from either Set1(762-1080), Set1(780-1080), or Set1(938-1080) containing COMPASS, suggesting Set1(938-1080), Cps60, Cps50, Cps30, and Cps25 proteins form the complex, which are shown to be

sufficient for in vitro H3K4 methyltransferase activity. Figure courtesy of the Shilatifard Lab.

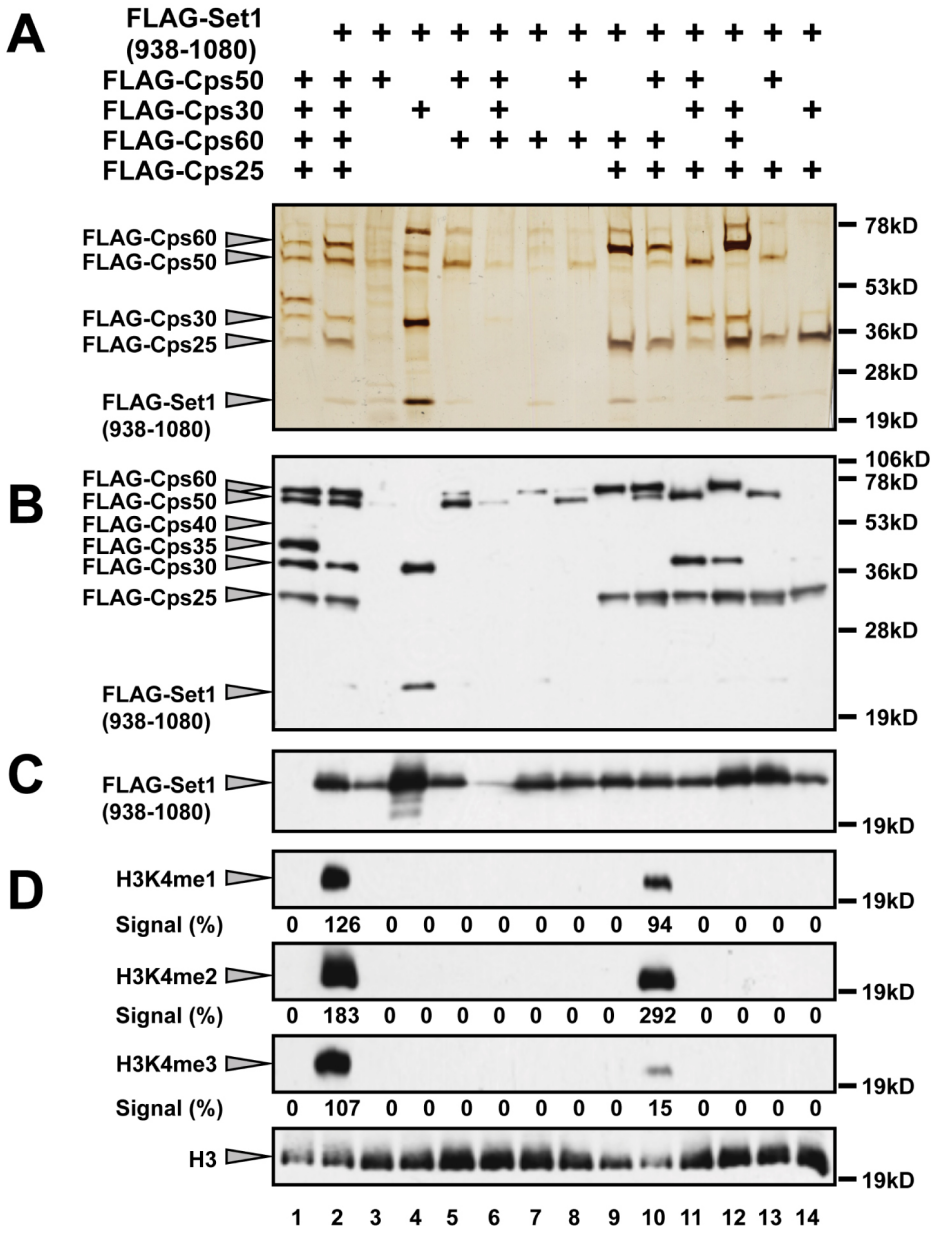


Figure 4-3. Core COMPASS retains in vitro H3K4 methyltransferase activity in the absence of Cps30.
Set1(938-1080) and a single or various combinations of the Cps50, Cps30, Cps60, and Cps25 subunits as denoted on the top of box A except lane 1 were prepared. As a negative control, COMPASS/-Set (Cps60, Cps50, Cps40, Cps35, Cps30, and Cps25) was used in lane 1. Purification of Set1(938-1080) and specified Cps subunits were confirmed by silver-staining (A) and Western blotting using anti-FLAG (B) and anti-Set1 (C) antibodies. The silver-stained gel exhibitied different staining characteristics for each

purified protein, which are not always proportional to its loaded amount, leading to some discrepancy of band intensities between silver-stained gel and western blotting using anti-FLAG and anti-Set1. Several samples were purified with some unknown additional proteins like a band shown at 78kD. In vitro H3K4 methyltransferase activity toward free histone H3 was tested by Western blotting using anti-H3K4me1, me2, and me3 antibodies (D). Note that the low levels of Set1 in lane 6 may also contribute to the absence of activity observed in that condition. Figure courtesy of the Shilatifard Lab.

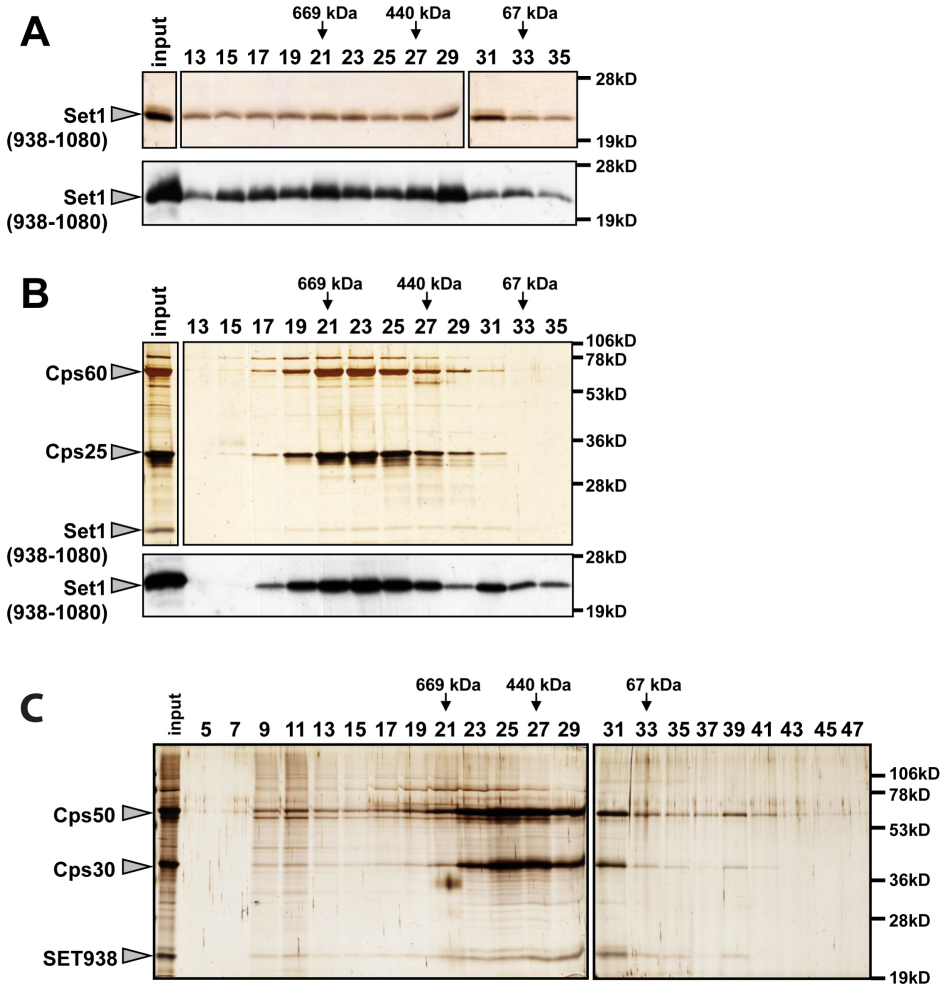


Figure 4-4. Set1(938-1080) correctly folds and interacts with Cps60-Cps25 subcomplex in the presence of Cps60 and Cps25.
(A) Set1(938-1080) was over-expressed in Sf9 insect cells, prepared by FLAG affinity purification, then applied onto Superose 6 gel filtration column chromatography. Input and each fraction denoted on the top of the upper box were subjected to SDS-PAGE and silver-stained (upper box) or western blotting using anti-Set1 antibody (lower box). (B) Set1(938-1080), Cps60, and Cps25 were co-expressed in Sf9 insect cells and prepared as in (A). Input and each fraction were analyzed by silver-staining (upper panel) and

western blotting using anti-Set1antibody (lower box). (C) Set1(938-1080), Cps50, and Cps30 were co-expressed and prepared from over-expressing Sf9 insect cells through FLAG affinity purification, and then applied onto Superose 6 gel filtration. Fractionated samples denoted on the top of the figure were subjected to SDS-PAGE and silver-staining. Set1(938-1080) forms heterogeneous oligomer even in the presence of Cps50 and Cps30. Figure courtesy of the Shilatifard Lab.

Next, we evaluated the intrinsic enzymatic activity of soluble Set1 protein in isolation. We did not detect any H3K4 methylation by Set1 alone or in the presence of putative Set1-stabilizing components Cps50 and Cps30 (Figure 4-5A, lanes 2 and 3), thereby further confirming that Cps60 and Cps25 are required for H3K4 methylation activity by Set1. In the reconstituted system, Set1 (938-1080) was sufficient in replacing the full-length protein in terms of stability and enzymatic activity (Figure 4-5A & 4-1B), except in the absence of Cps60 and Cps25 (Figure 4-5A, lane 7 & 4-5B, lane 7). Whereas full-length Set1-containing COMPASS supported low levels of residual *in vitro* H3K4 mono- and dimethylation in the absence of Cps60 and Cps25, this activity was completely abolished in Set1 (938-1080)-containing COMPASS.

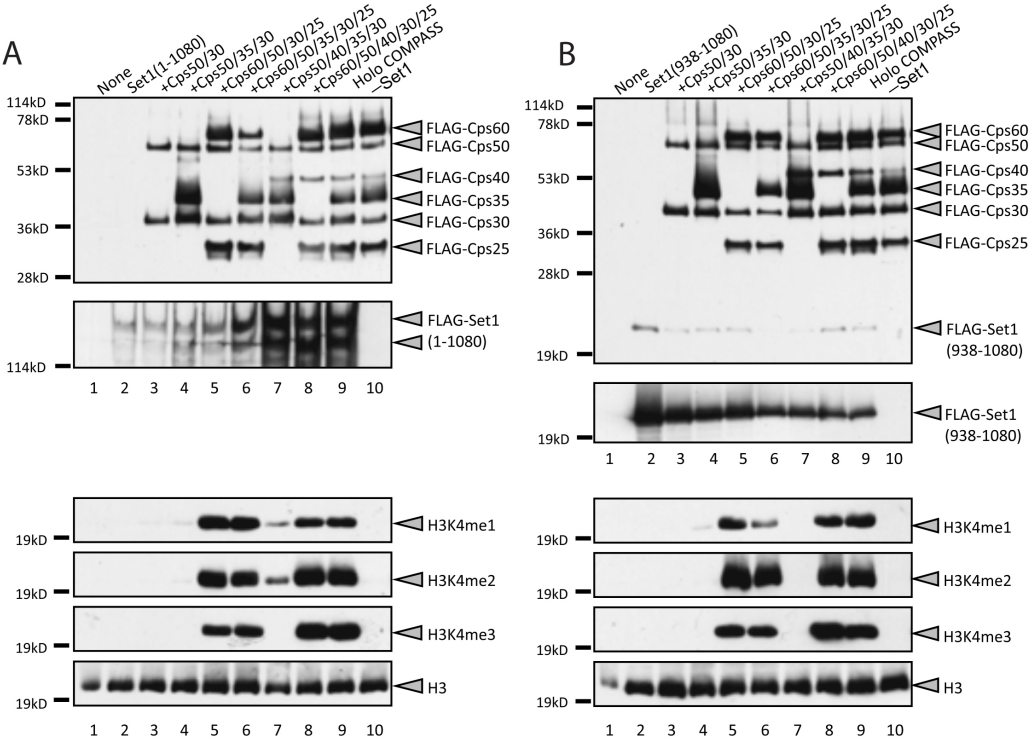


Figure 4-5. Recombinant Set1/COMPASS complexes and *in vitro* histone methyltransferase activities.

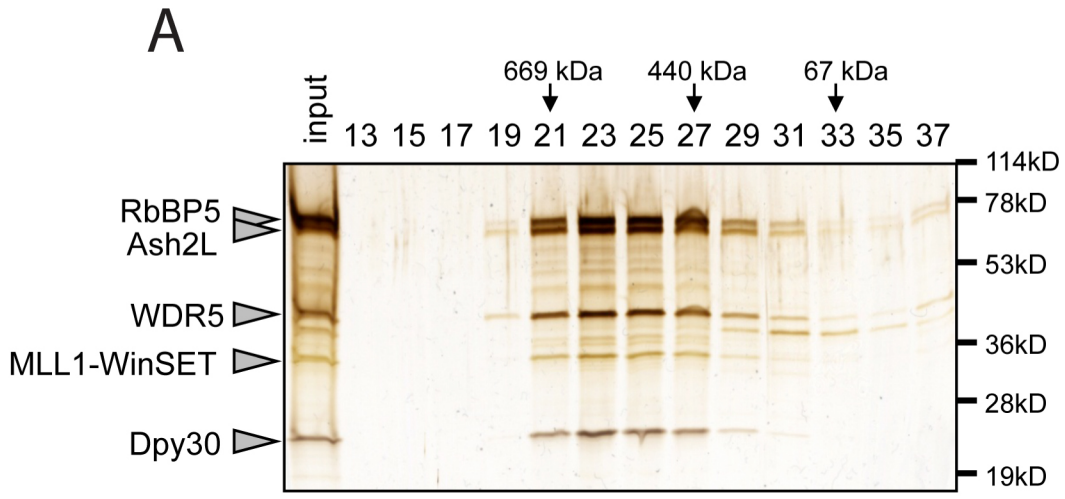
(A**)** Purifications of recombinant COMPASS consisting of full-length Set1 with various combinations of Cps60, Cps50, Cps40, Cps35, Cps30, and Cps25. Subunit composition of purified COMPASS was confirmed by Western blotting using anti-FLAG (top panel) and anti-Set1 (middle panel) antibodies. *In vitro* H3K4 methyltransferase activities toward free histone H3 of various recombinant COMPASS complexes were examined by Western blotting using anti-H3K4me1, me2, and me3 antibodies (bottom panels). (B**)** Recombinant COMPASS complexes were prepared using Set1(938-1080) instead of full-length Set1. COMPASS composition and *in vitro* H3K4 methyltransferase activities were analyzed in the same way as (A). Figure courtesy of the Shilatifard Lab.

Recent studies with H3 peptide substrates have suggested that WDR5, RbBP5, Ash2L, and Dpy30 (also known as the WRAD complex), the human homologs of Cps30, Cps50, Cps60, and Cps25, respectively, display histone H3K4 methyltransferase activity independent of the Set1 domain of MLL1 (Cao et al., 2010; Patel et al., 2009; Patel et al., 2011). Despite the highly conserved components between the yeast COMPASS and the human MLL/COMPASS-like complexes, we could not detect any H3K4 methylation signals in a reconstituted Cps30, Cps60, Cps50, and Cps25 complex in the absence of Set1 (Figure 4-5A and 4-5B; see lane 10 in both figures). This is consistent with our previous *in vivo* studies where we were unable to detect any histone H3K4 methylation in strains with *SET1* deletions (Schneider et al., 2005), suggesting that Set1 is the only histone H3K4 methylase in yeast, and that the yeast WRAD complex does not demonstrate substantial HMTase activity associated with it *in vivo* or *in vitro*. Indeed, our enzymatic analyses (Figure 4-6) demonstrated the presence of less than 1% of H3K4me1 and virtually undetectable levels of H3K4me2-3 when using WRAD as compared to wild-type Set1/COMPASS or MLL/COMPASS-like complexes. Given the extremely weak catalytic rates reported for the reconstituted WRAD complex as a SET-domain independent histone H3K4 methylase (Cao et al., 2010; Patel et al., 2009; Patel et al., 2011), the biological relevance of such activity needs to be further demonstrated.

The 'core' components of COMPASS complex required for trimethylation

Insect cell-expressed core COMPASS also enabled us to investigate the regulation of histone H3K4 methylation by Cps50 and Cps30. Systematic analysis for *in vitro* H3K4 methyltransferase activity (Figure 4-3) shows that core COMPASS is rather insensitive to the lack of Cps30 with comparable H3K4 mono- and dimethylation, but partially

defective trimethylation activity (Figure 4-3D, lane 2 & 10). This suggests that COMPASS is enzymatically active in the absence of Cps30, but the processivity of the SET domain from H3K4 di- to trimethylation is regulated by Cps30. In contrast, reconstituted core COMPASS lacking either Set1 (938-1080), Cps60, Cps50, or Cps25 completely lacked enzymatic activities (Figure 4-3D, lane 1, 6, 11, and 12; note that the low levels of Set1 in lane 6 may also contribute to the absence of activity observed in that condition). Thus, we have identified Set1 (938-1080), Cps60, Cps50, Cps30, and Cps25 as the necessary core COMPASS components for high levels of *in vitro* H3K4 mono-, di- and trimethylation using histone H3 as a substrate. Both Cps60 and Cps25 are essential for *in vitro* enzymatic activity regardless of the combination of other core COMPASS subunits (Figure 4-5A lanes 3, 4, and 7; and Figure 4-3D, lanes 3-8, 11-14). It is noted that the Cps60-Cps25 subcomplex (Roguev et al., 2001) co-eluted with Set1(938-1080) upon size exclusion chromatography when reconstituted together (fractions 21 to 23; Figure 4-4B). This result demonstrates a direct interaction between the SET domain and the Cps60-Cps25 module (Figure 4-4B). In contrast, in the absence of Cps60 and Cps25, soluble Set1 (938-1080) elutes broadly on the same size exclusion chromatography system (fractions 13 to 29; Figure 4-4A). Given that co-expression of neither Cps50 nor Cps30 had any effects on Set1 (938-1080) elution profile (Fig 4-4C), Cps60 and Cps25 may play an additional role in the correct assembly of COMPASS through interactions with Set1.



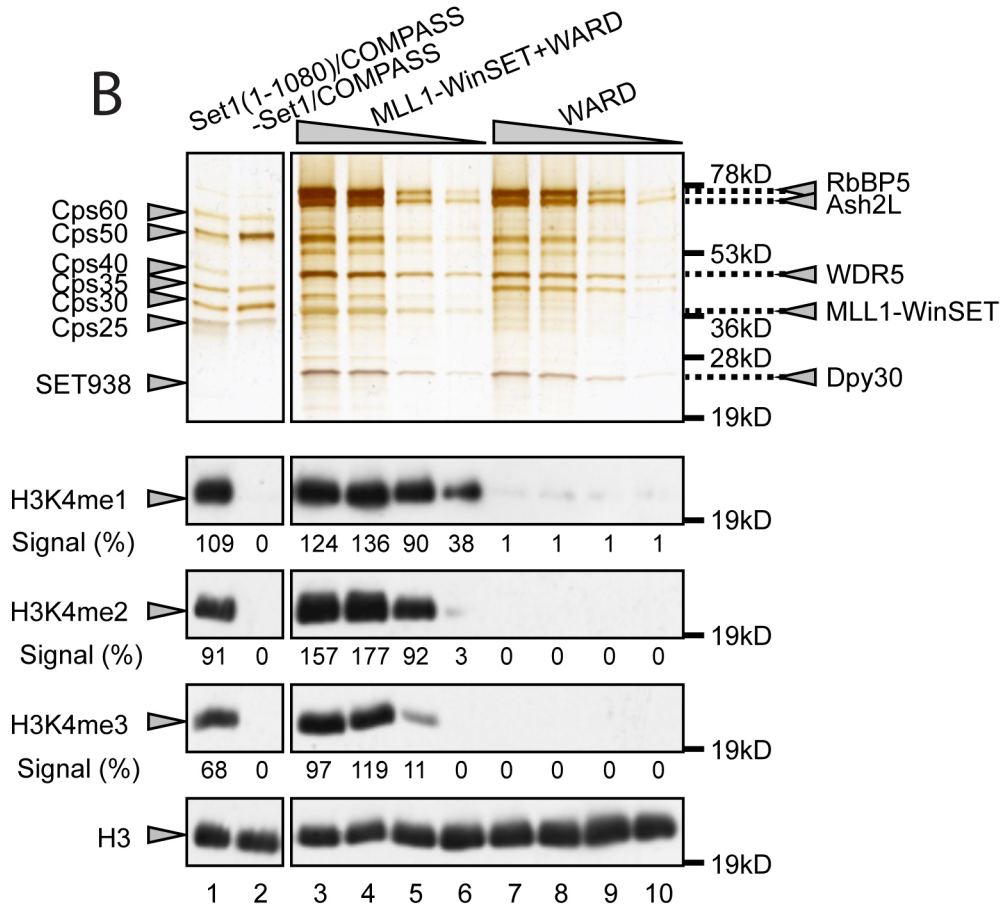


Figure 4-6. Purification and histone methyltransferase activity of human core MLL complex.

(A) Full-length WDR5 (human homolog of Cps30), Ash2L (Cps60), RbBP5 (Cps50), Dpy30 (Cps25), and the C-terminal fragment of MLL1 including SET domain and the adjacent Win-motif were fused with FLAG tag in their N-termini, co-expressed in Sf9 insect cells, and prepared by FLAG affinity purification. Purified samples were separated by Superose 6 gel filtration column along fractions denoted on the top of each lane, and then characterized by SDS-PAGE and silver-staining. (B) FLAG-affinity purified recombinant COMPASS and MLL1 complex were analyzed with in vitro histone mehtyltransferase assay using free histone H3 from H. sapiens. The methylated histone H3 were detected by using anti-H3K4me1, me2, and me3 specific antibodies. The same amounts of enzymes used in the histone mehtyltransferase assay were served in the silver-stained gel for comparison (top). Several minor bands showing up in MLL1 complex (MLL1-WinSET+WARD), and WARD complex are impurities from insect cells. Figure courtesy of the Shilatifard Lab.

Yeast COMPASS subunit:	Human COMPASS subunit:
Cps60 (Bre2)	Ash2l
Cps50 (Swd1)	RbBP5
Cps40 (Spp1)	CFP1
Cps35 (Swd2)	Wdr82
Cps30 (Swd3)	WDR5
Cps25 (Sdc1)	Dpy30

Table 4-1. Yeast COMPASS and human COMPASS subunits.

Electron microscopic mapping of core COMPASS

To gain insight into the architecture of yeast core COMPASS, we employed electron microscopy to visualize preparations consisting of purified Set1 (938-1080), Cps60, Cps50, Cps30, and Cps25 (Figure 4-7A, fraction 18; Figure 4-1B). Raw images of negative stained specimen revealed a mono-disperse population of complexes with similar sizes (Figure 4-8A). To analyze these particles, 35,527 projections were interactively selected and grouped into 100 classes by reference-free alignment and classification (Figure 4-9A). The 2D class averages revealed flexible conformers of two major populations of particle projections: one displays a Y-shaped particle with a triangular base connecting through an extension to two adjacent oval lobes, one of which is in closer proximity to the base (Figure 4-7B, right; Figure 4-9A); the second population reveals a similar triangular base and an arm extension connecting to a single circular doughnut-like domain with a distinct stain accumulation region in the center (Figure 4-7B, left). The doughnut-shaped domain appears identical to the projection profile of a WD40 domain (inset in Figure 4-7B), which defines both Cps50 and Cps30, further suggesting that a subpopulation of the reconstituted COMPASS complexes might be missing one of the two WD40 domain-containing subunits.

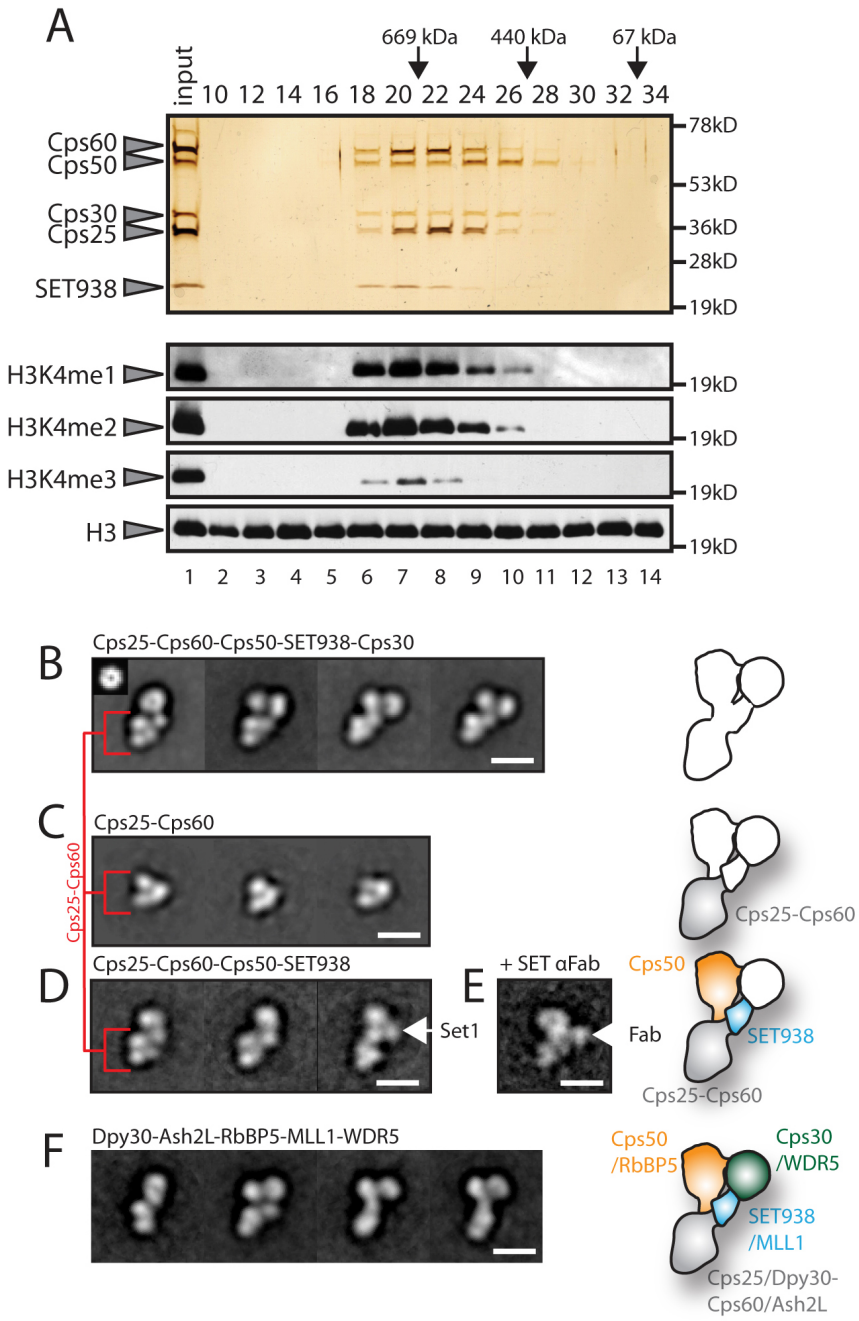


Figure 4-7. EM mapping of core COMPASS.
(A) Purification of Core Set1/COMPASS by size exclusion chromatography and analysis of the active fractions by HMTase assay. Core Set1/COMPASS consisting of Cps25, Cps60, Cps50, Set1(938-1080) was reconstituted and fractionated by size exclusion chromatography over a *Superose-6 PC* 3.2/30 (GE Healthcare). The H3K4 methylase activities of the resulting fractions were tested as shown. (B) Representative 2D averages of negative stained core COMPASS (fraction 18 in A) including Cps25, Cps60, Cps50, Set1(938-1080) which is denoted as SET938, and Cps30. The inset shows the 2D projection profile of a WD40 domain for comparison (scale bar 10nm). (C)

Representative 2D averages of the Cps25-Cps60 assembly. (D) Representative 2D averages of negative stained Cps25-Cps60-Cps50-SET938 complex. (E) 2D average of Cps25-Cps60-Cps50-SET938 complex labeled with a Fab fragment against SET938. (F) Representative 2D class averages of negative stained human core MLL1/COMPASS-like complex. Schematic representations: Cps25/DPY30-Cps60/Ash2L (grey), SET938/MLL1 (blue), Cps50/RbBP5 (orange), and Cps30/WDR5 (green).

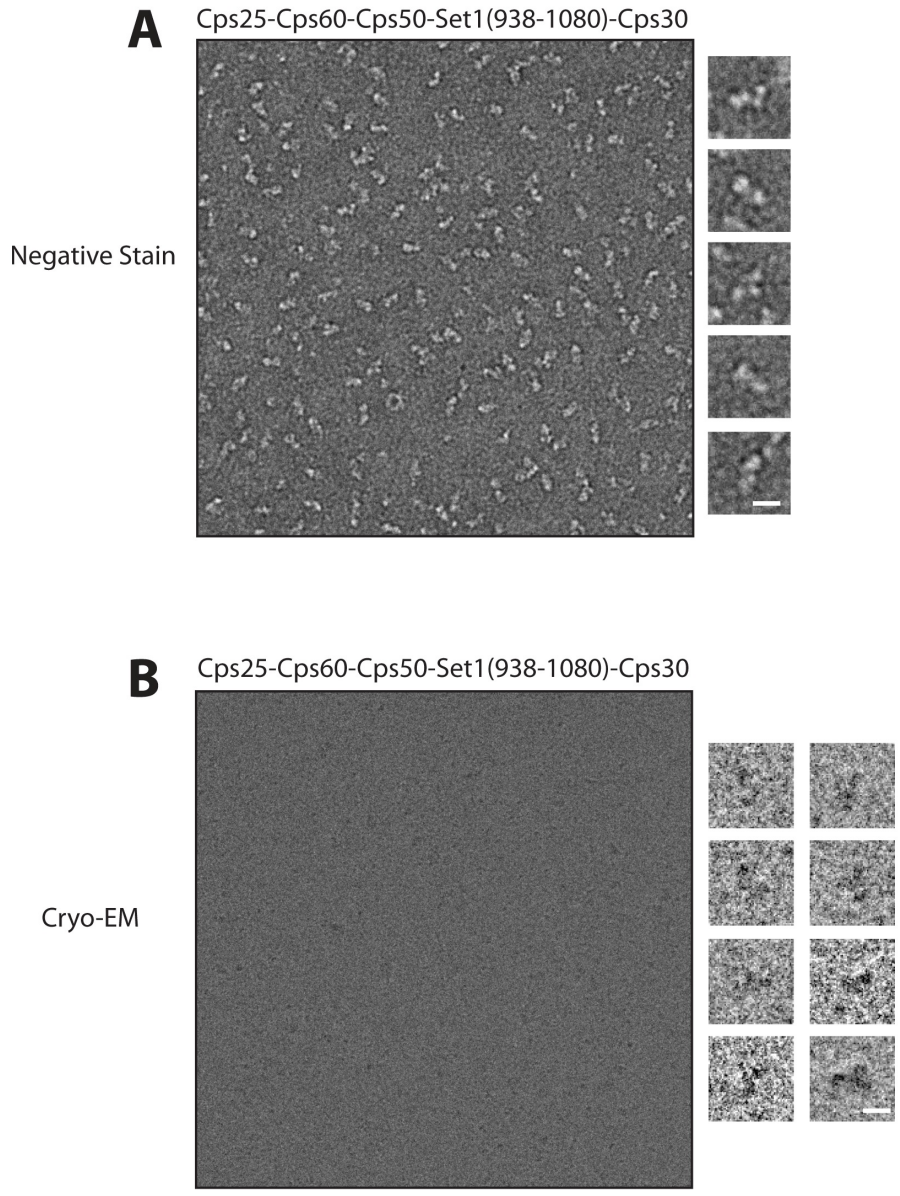


Figure 4-8. Negative stain and cryo-EM images of core COMPASS
(A) Raw image of negative stained Cps25-Cps60-Cps50-Set1(938-1080)-Cps30. (B) Raw cryo-EM image of Cps25-Cps60-Cps50-Set1(938-1080)-Cps30. Scale bar = 10 nm.

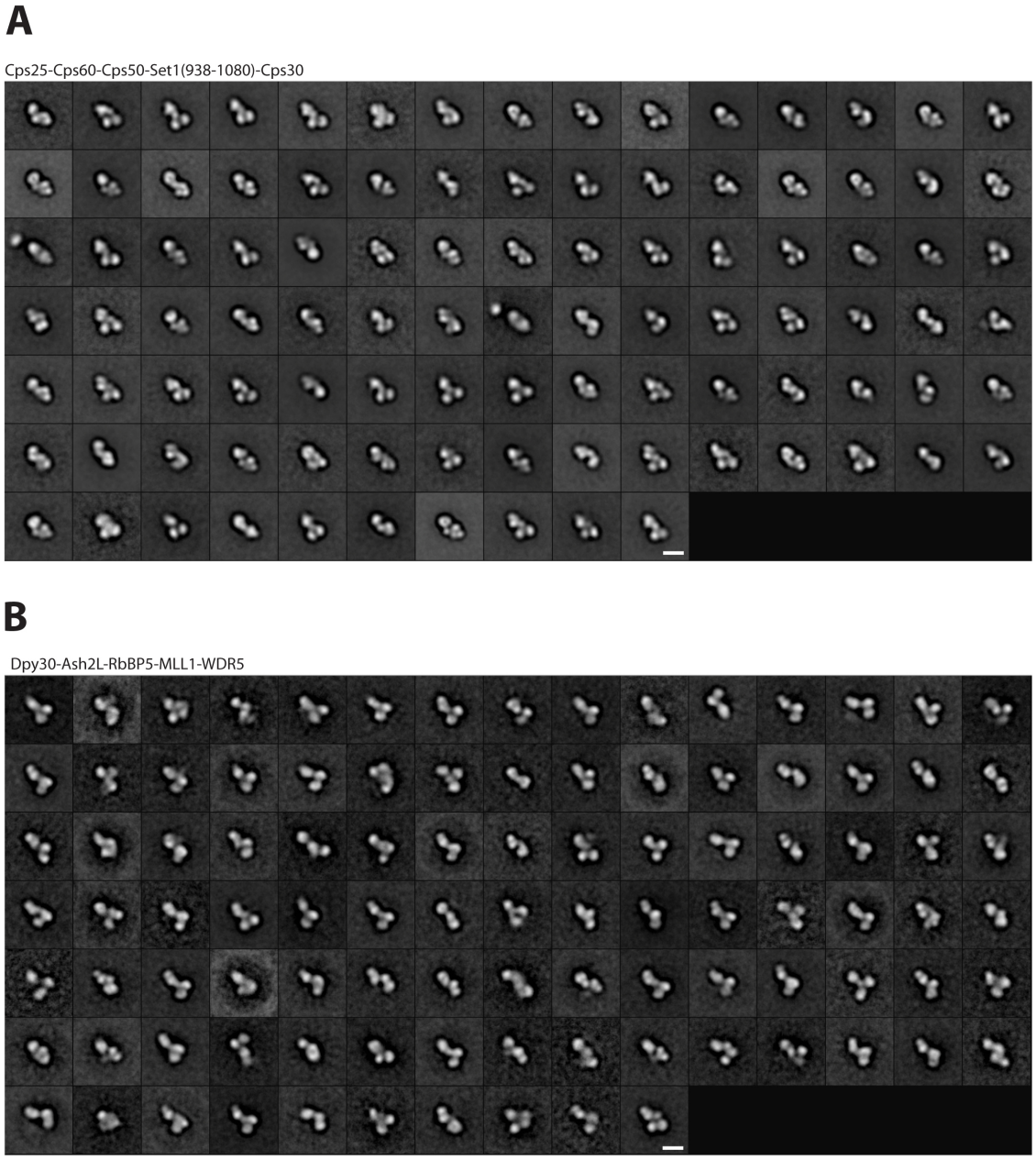


Figure 4-9. 2D class averages of yeast and human core COMPASS complex.
(A) 35,527 projections of negative stained yeast Cps25-Cps60-Cps50-Set1(938-1080)-Cps30 were classified into 100 classes. (B) 2D class averages of human MLL1 core complex. 12,454 projections of negative stained Dpy30-Ash2L-RbBP5-MLL1-WDR5 were classified into 100 classes. Scale bar = 10 nm.

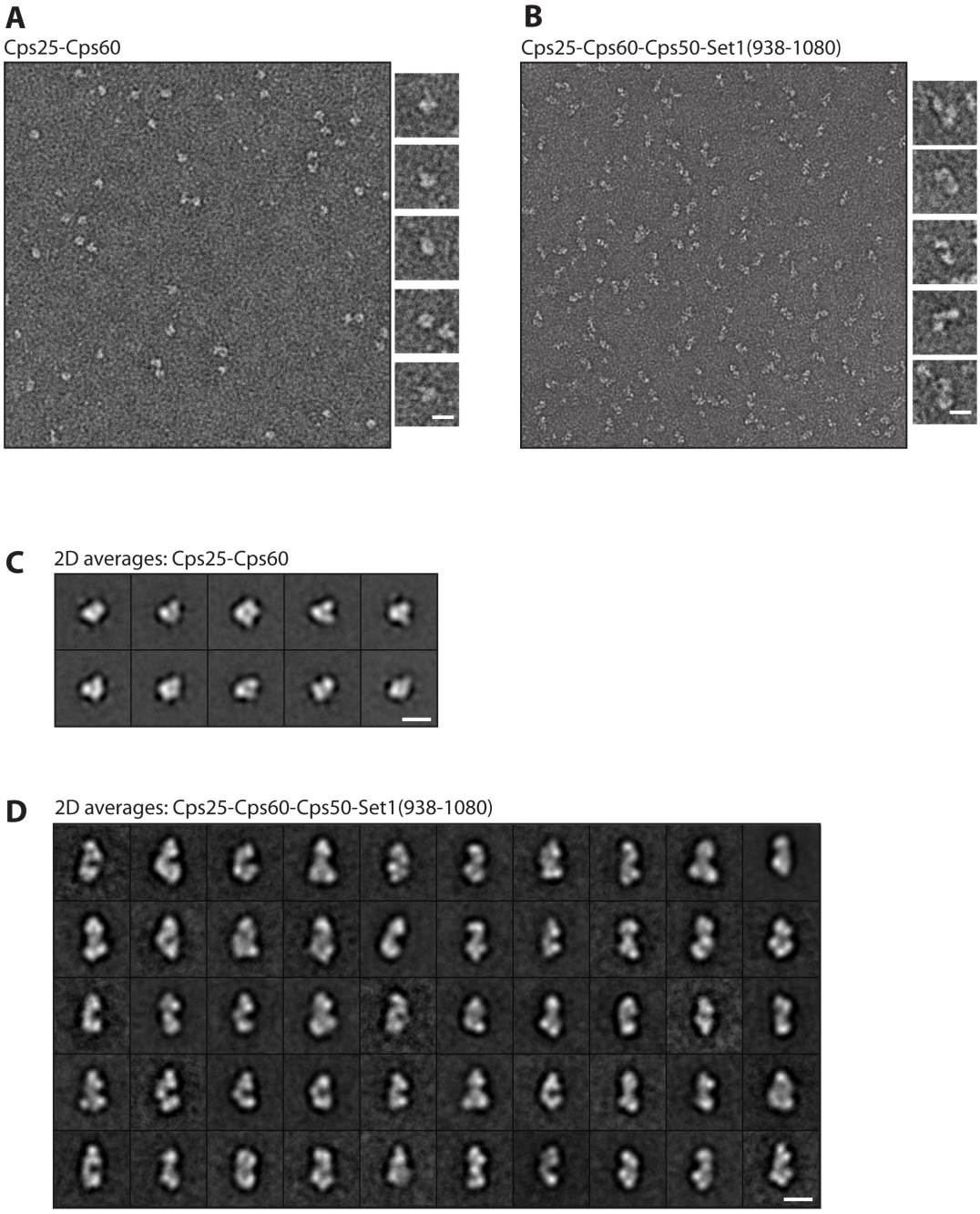


Figure 4-10. Images and 2D classification of the Cps25-Cps60 and Cps25-Cps60-Cps50-Set1(938-1080) complexes.
(A) Raw image of the negative stained Cps25-Cps60 complex. (B) Raw image of the negative stained Cps25-Cps60-Cps50-Set1(938-1080) complex. (C) 2D class averages of the negative stained Cps25-Cps60 complex. 1,809 particles were classified into 10

classes. (D) 2D class averages of the negative stained Cps25-Cps60-Cps50-Set1(938-1080) complex. 3,798 particles were classified into 50 classes. Scale bar = 10 nm.

EM of COMPASS sub-complexes for domain assignment

In order to identify each domain within the core COMPASS, we visualized individual subcomplexes by negative stain EM. We first examined the structure of the Cps60-Cps25 subcomplex, where classification and averaging (Figure 4-7C, Figure 4-10A & C) revealed a triangular structure with similar dimensions and shape to the triangular base found in the images of core COMPASS (Figure 4-7B). This result suggests that Cps60 and Cps25 form the base of core COMPASS. We next analyzed EM images of reconstituted core COMPASS lacking the Cps30 subunit (core COMPASS/-Cps30) (Figure 4-10B & D). 2D averages from this analysis show the well-defined triangular base, with an arm extension and an additional oval lobe near the top of the complex (Figure 4-7D). Missing, however, is the second, more distant oval lobe that is present in core COMPASS. This indicates that the Cps30 subunit is the more terminal of the two lobes away from the triangular base, while the lobe closer to the middle of the complex belongs to the WD40 domain of Cps50. In addition, a few averages reveal a small density extending from the connection between the Cps60-Cps25 base and the Cps50 domain, which might be attributed to the SET domain (Figure 4-7D, right panel). To test this hypothesis, we incubated core COMPASS/-Cps30 with a Fab fragment conjugated to SET and analyzed the complexes by single-particle EM. 2D projection averages of Fab-labeled complexes clearly reveal the extra Fab density at the level of the region connecting the Cps60-Cps25 base and Cps50. (Figure 4-7E). Next, we examined the architecture of the reconstituted human core COMPASS consisting of MLL SET domain, RbBp5, Ash2L, WDR5 and Dpy30. Negative stain 2D averages from this preparation revealed an architecture with striking resemblance to the yeast complex (Figure 4-7F). Thus, both the structure and function and COMPASS-like complexes appear to be highly conserved from yeast to human.

We next decided to investigate the architecture of the COMPASS complex with the inclusion of the Cps40 subunit. This subunit has previously been indicated in regulating trimethylation activity. In order to achieve complex formation with the Cps40 subunit, a

longer Set1 truncation (762 – 1080) that includes the n-SET domain is needed for Cps40 to associated with the complex designated '+Cps40-SET762.' Using negative stain EM we compared 2D class averages of the recombinant core COMASS complex (Cps60-Cps50-Cps30-Cps25-SET938) to recombinant and natively purified +Cps40-SET762 complex (Cps60-Cps50-Cps40-Cps30-Cps25-SET762). When comparing the class averages of both populations, with 1 or 2 WD40-like domains, the +Cps40-SET762 complex has an additional rod like density interacting with the WD40-like domains of Cps30 and Cps50 (Figure 4-11).

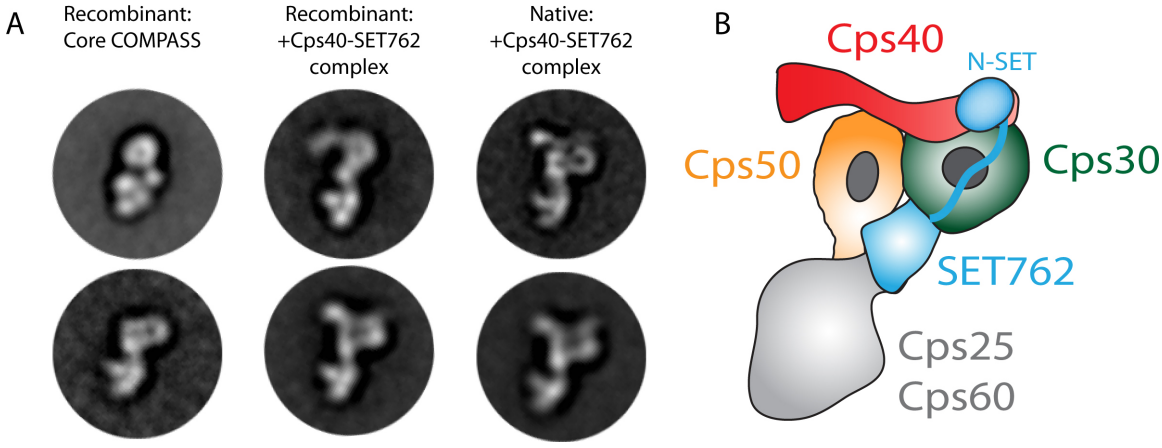


Figure 4-11. Class averages and model of +Cps40-SET762 complex.
A) Left, representative 2D class averages of the recombinant 'core' COMPASS complex (Cps60-Cps50-Cps30-Cps25-SET938). Middle, representative class averages of recombinant '+Cps40-SET762' COMPASS complex (Cps60-Cps50-Cps40-Cps30-Cps25-SET762). In the top class averages left and middle respectively, shows sub-population of complex with only 1 WD40 domain (Cps50). In the top-middle class average, the additional density of Cps40 is seen compared to COMPASS complex without Cps40 (top-left). Right, representative class averages of native '+Cps40-SET762' COMPASS complex. B) Model of Cps60-Cps50-Cps40-Cps30-Cps25-SET762 complex with domain designation.

3D architecture of core COMPASS

To further characterize the architecture of the core COMPASS, we sought to examine the 3D structure of the yeast assembly. In the first step, we calculated several 3D models from negative stained particles belonging to the individual groups produced by classification (Figure 4-12). For this approach, we used the corresponding 60º tilted projections and the random conical tilt (RCT) method to calculate initial reconstructions,

which were further refined after the inclusion of 0º particle views. The 3D reconstructions of particles belonging to classes with only one globular domain connected to the Cps60-Cps25 base reveal that these particles have indeed only one WD40-like domain. In contrast, 3D reconstructions from Y-shaped particles reveal the Cps60-Cps25 module having two globular domains connected to it at different distances (Figure 4-12).

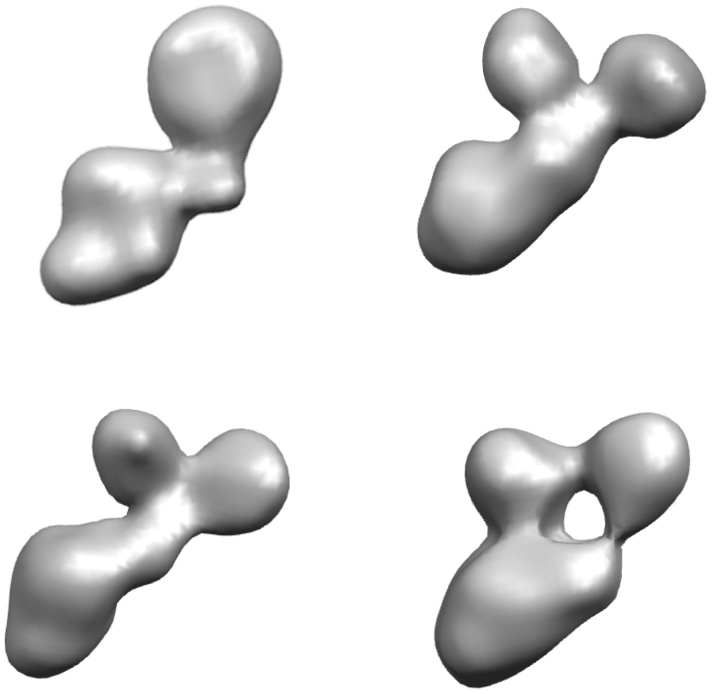


Figure 4-12. 3D reconstructions from random conical-tilt (RCT) for yeast core COMPASS embedded in negative stain.
3D reconstructions from representative conformers of the COMPASS complex.

In these 3D reconstructions, we notice variability in the disposition of the Cps60-Cps25 base as well as in the proximity between the two WD40-like lobes of Cps50 and Cps30. We reasoned that the various conformations might be the result of inherent complex flexibility, resulting in deformations due to the negative stain preparation on the carbon support of the grids. Therefore, we focused our efforts on cryo-EM of the complex in holey carbon grids, and successfully obtained particle images from specimen suspended in thin vitreous ice (Figure 4-8B).

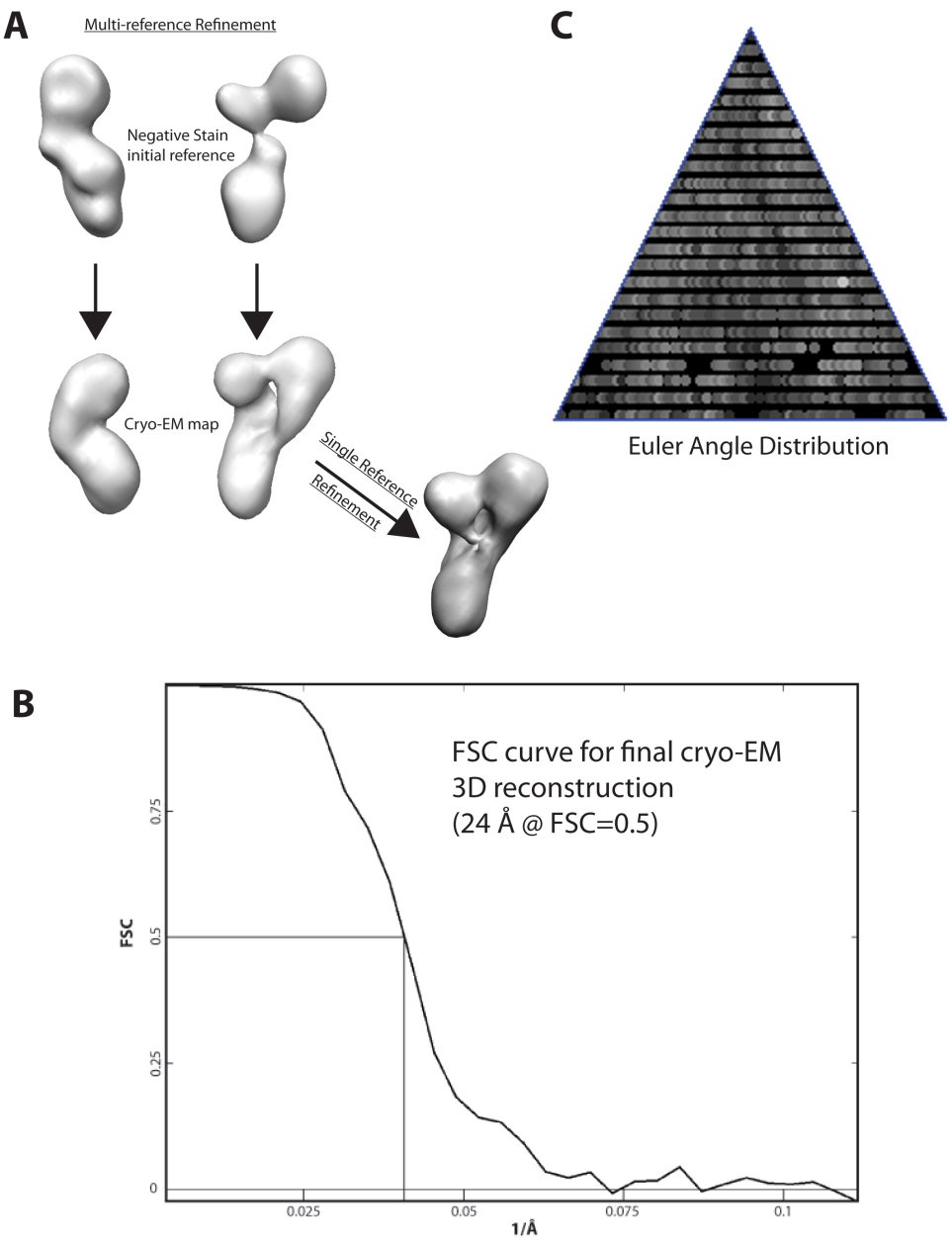


Figure 4-13. Overview of multi-reference alignment, with FSC and angular distribution.
(A) Two initial references from negative stain 3D maps were used for initial multireference alignment of cryo-EM data, enabling us to separate and obtain a 3D map from projections of particles encompassing all subunits (right volume). Following this first classification, the separated 11,671 particle projections belonging to particles with all subunits were subjected to single reference refinement and 3D reconstruction, thereby producing the final cryo-EM 3D map. (B) FSC curve of final cryo-EM map. The final map resolution is 24 Å based of the FSC = 0.5 criterion. (C) The Euler angular distribution of projections used in the final cryo-EM map.

Since our negative stain analysis of core COMPASS indicated the presence of two major populations (+/- one WD40 domain), we subjected 21,583 cryo-EM images to multiple reference-supervised alignment (Brink et al., 2004; Menetret et al., 2005) using two initial 3D references (Fig. 13A): a 3D volume representing only the Cps60-Cps25 base and one WD40-like domain (Figure 4-13A, left), and a 3D volume that included the base and both WD40-like domains (Figure 4-13A, right). This approach allowed us to effectively separate the two major populations within the cryo-EM dataset, resulting in stable particle assignments after several cycles of multi-reference alignment. We then employed single reference alignment and reconstruction to produce a final cryo-EM map of the core COMPASS at a resolution of 24 Å (FSC = 0.5) (Figure 4-13B). The cryo-EM 3D map displayed features that are clearly distinct from those of the starting reference 3D map from negative stain EM. To validate our reconstruction, we also employed different initial references for cryo-EM projection alignment, thereby, producing very similar final maps (Figure 4-14). Furthermore, 2D classification of the cryo-EM projections (Figure 4-15A) produces averages that are in very good agreement with re-projections of the 3D map (Figure 4-15B).

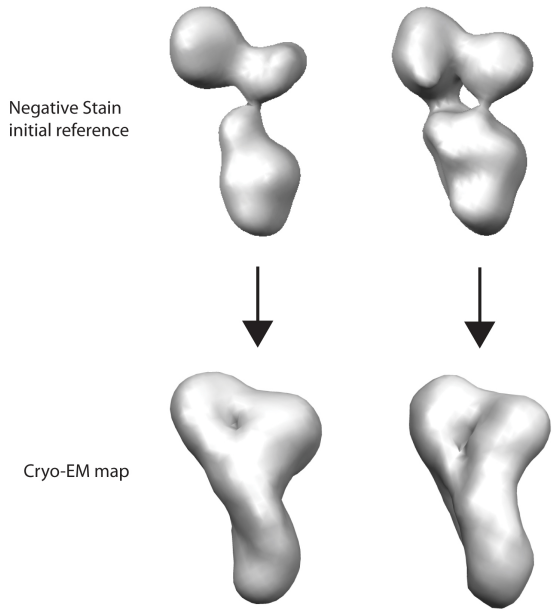


Figure 4-14. Alignment and reconstruction of cryo-EM projections with different initial models from negative stained core COMPASS (top) resulted in very similar cryo-EM 3D maps (bottom).

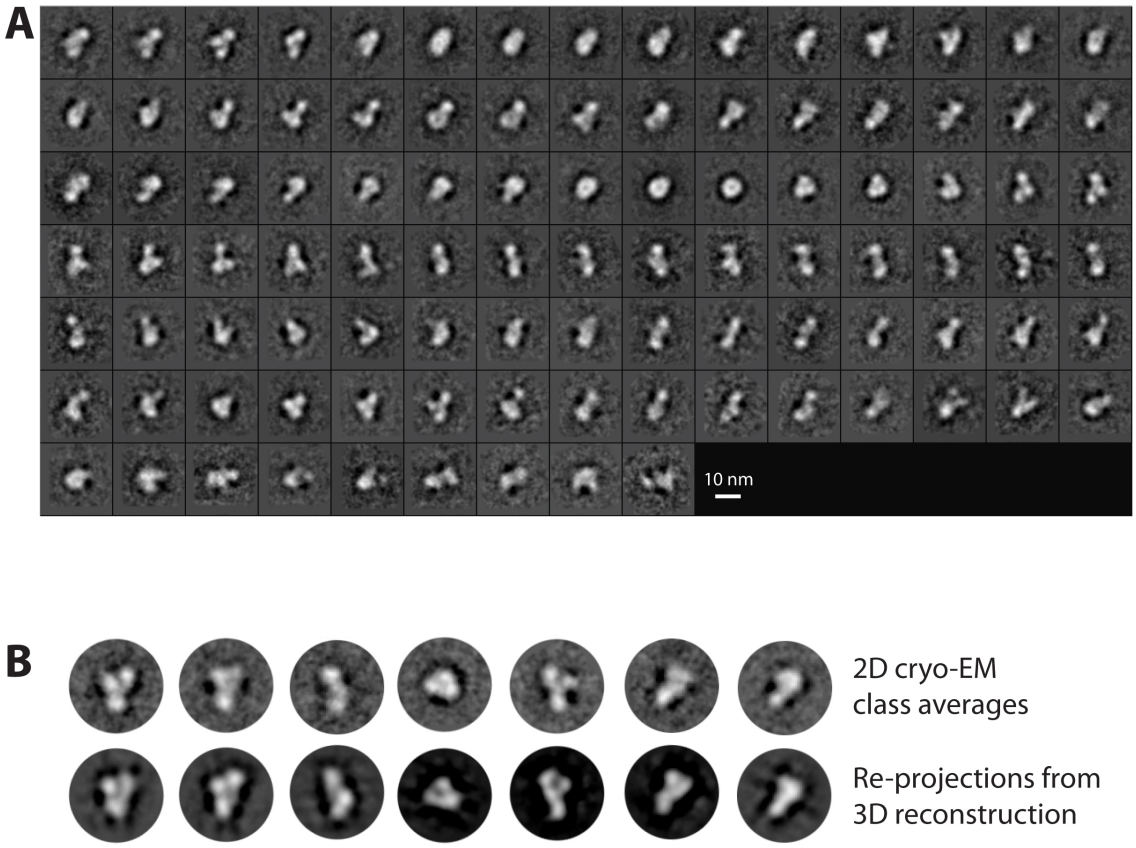


Figure 4-15. 2D classification and projection comparison.
(A) 2D classification of cryo-EM projections of yeast core COMPASS used for 3D reconstruction. 11,671 particle projections were classified into 100 classes. (B) Comparison of class averages from 2D classification of projections (top) with re-projections from final cryo-EM 3D reconstruction (bottom).

Molecular modeling of core COMPASS

To obtain deeper insights to the architecture of the core COMPASS, we generated a model by docking the available or homologous crystal structures into their corresponding positions within the cryo-EM map and according to our assignment from the 2D projection analysis (Figure 4-16). Due to the limits in resolution, we performed rigid body manual docking based on visual inspection of the fit. The two neighboring globular domains of the 3D map represent the two WD40 domains, with the lobe more distal to

the Cps60-Cps25 base belonging to Cps30 (Figure 4-16 & 4-17B, green) and the other to Cps50 (orange). To model Cps50, which contains a β-propeller fold and a C-terminal of ~80 amino acid tails, we docked a homologous WD40 domain (orange; PDBID:2XL2) in its corresponding position. A recent crystal structure of WDR5, a Cps30 homolog, has been obtained with peptides of both the C-terminal tail of RbBP5 (Cps50 homolog) and the Win motif that is N-terminally adjacent to the SET domain of MLL1 (Set1 homolog) bound onto the opposite faces of WDR5(Avdic et al., 2011; Odho et al., 2010). We thus docked the co-crystal structure (PDBID:3P4F) of the MLL1 Win peptide-WDR5-RbBP5 tail peptide in the position of Cps30 with an orientation that places the RbBP5 tail peptide close to the Cps50 position on the cryo-EM map. Our Fab labeling experiments position the SET domain in the region connecting the two WD40 domains and the Cps60-Cps25 base. Accordingly, the Set1/SET domain from the homologous crystal structure (Southall et al., 2009; Takahashi et al., 2009) was docked in this position. Although yeast Set1 does not have the canonical Win motif upstream of the SET domain, we identified a similar Ala-Arg-Ser motif at positions 943-945 of the Set1 primary sequence, and found that core COMPASS harboring either Set1 (R944A) or Set1(S945A) substitutions almost completely loses its H3K4 methyltransferase activity (Figure 4-18). This suggests that the Ala-Arg-Ser motif is important in complex formation like the Win motif in the MLL1 core complex. Based on this result, we oriented the N-terminus of Set1 close to Cps30 and opposite to Cps50 (Figure 4-19). In addition to these interactions, and given the spatial constraints of this region, we could only dock the Set1 domain in a single orientation. In this model, one side of Set1 (938-1080) appears to extensively interact with Cps30, while another side is forming a bridge with density stemming from the Cps50 WD40 domain (Figure 4-17A). Thus, our modeling places the active site of Set1 in the middle of a central channel that runs through the complex, starting from the connection between Cps60-25 and Cps50, and running adjacent to Set1, exiting behind the interface of Cps30 with Cps50 (Figure 4-17A,B). In this configuration, the peptide bound to the active site of the SET domain, as shown in the crystal structure (PDBID:2W5Z), would reside directly in the middle of the COMPASS channel (shown in red in Figure 4-17A). Thus, this configuration may limit the substrates that COMPASS can recognize and methylate.

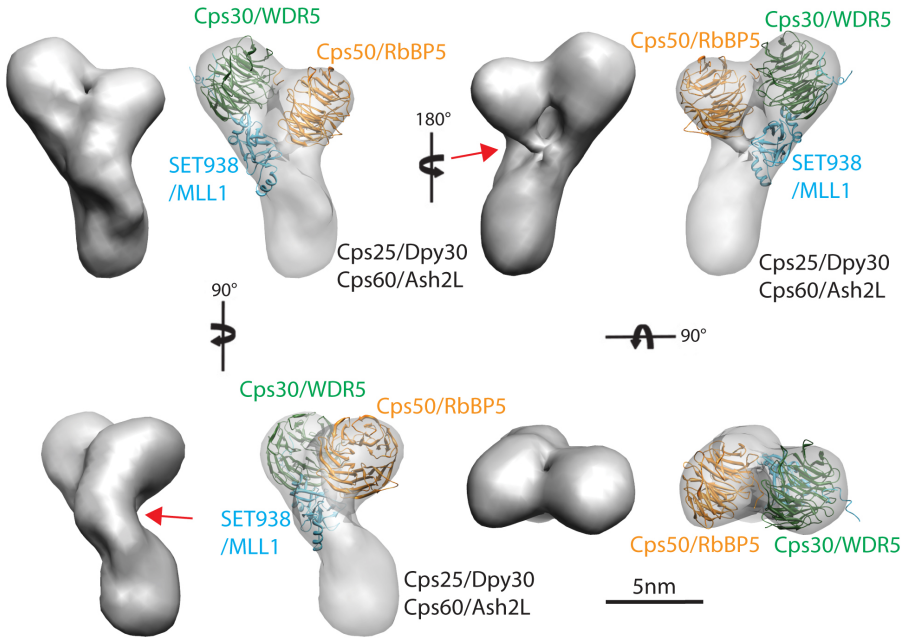


Figure 4-16. 3D cryo-EM reconstruction and modeling of core COMPASS.
Different views of the cryo-EM 3D map for the Cps25-Cps60-Cps50-SET938-Cps30 COMPASS complex. Each view shows the solid rendered map accompanied by a transparent map with modeled crystal structures of the WD40 domain of Cps50 (orange), Cps30 (green), and MLL1/SET938 (cyan). The red arrow indicates the expected position of the Cps50 arm that bridges with the Cps25-Cps60 module.

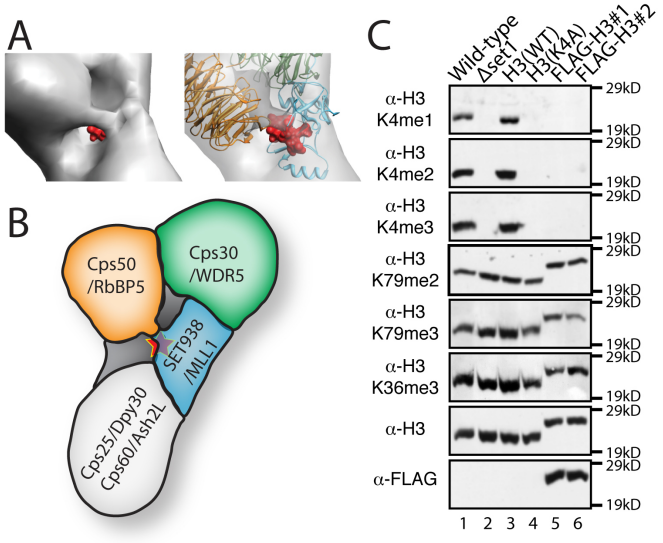


Figure 4-17. COMPASS family members are exomethylases.
(A) Zoom-in view of the central channel formed in the complex, with the histone peptide of the MLL1 (as co-crystallized) shown in the red space-filling model. **(B)** Schematic

model of core COMPASS family architecture. The red star indicates the histone peptide-binding region of SET938. **(C)**The N-terminally extended version histone H3 was constructed by inserting FLAG sequence (DYKDDDDK) between the Met start codon and the second Ala codon. COMPASS shows no H3K4 methylase activity with the FLAG-extended H3 tails *in vivo* (this figure) and *in vitro* (Figure 4-20).

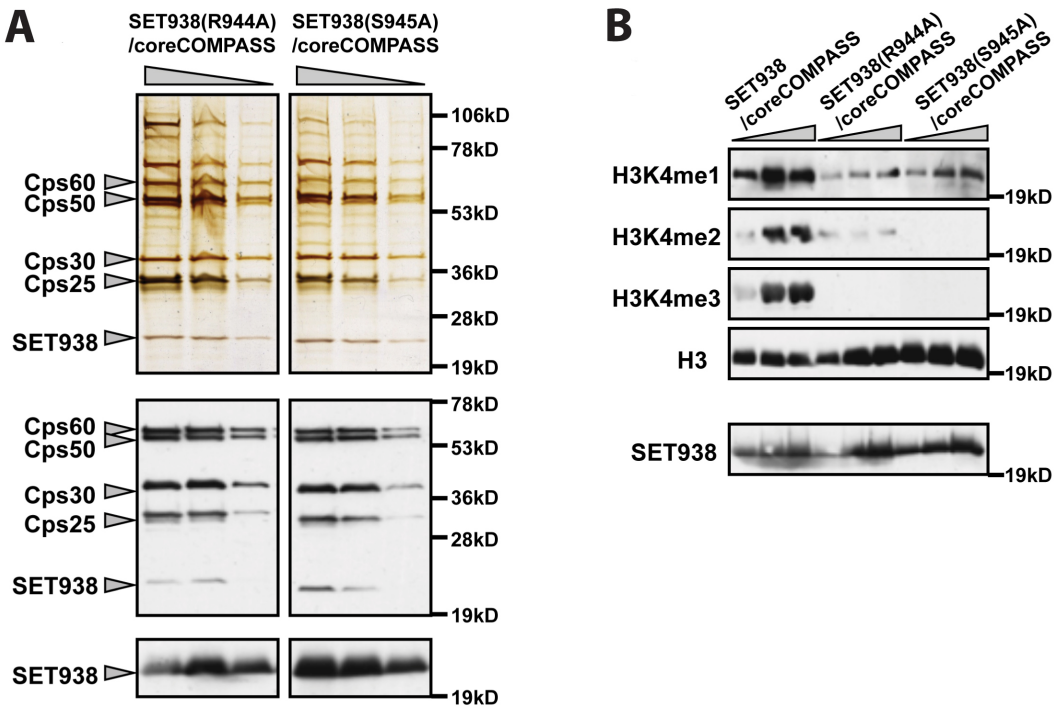


Figure 4-18. Set1 point mutations.
(A) Set1(938-1080) harboring R944A or S945A point mutation was co-expressed with other core COMPASS subunits (Cps60, Cps50, Cps30, and Cps25) in Sf9 insect cells and prepared by FLAG affinity purification. Purified core COMPASS was characterized with silver-staining (top box), anti-FLAG western blotting (middle box), and anti-Set1 western blotting (bottom box). (B) In vitro H3K4 methylation activities toward free histone H3 of each purified core COMPASS were examined by western blotting using anti-H3K4me1, me2, and me3 specific antibodies. The amounts of core COMPASS were evaluated by anti-Set1 western blotting (bottom box). Figure courtesy of the Shilatifard Lab.

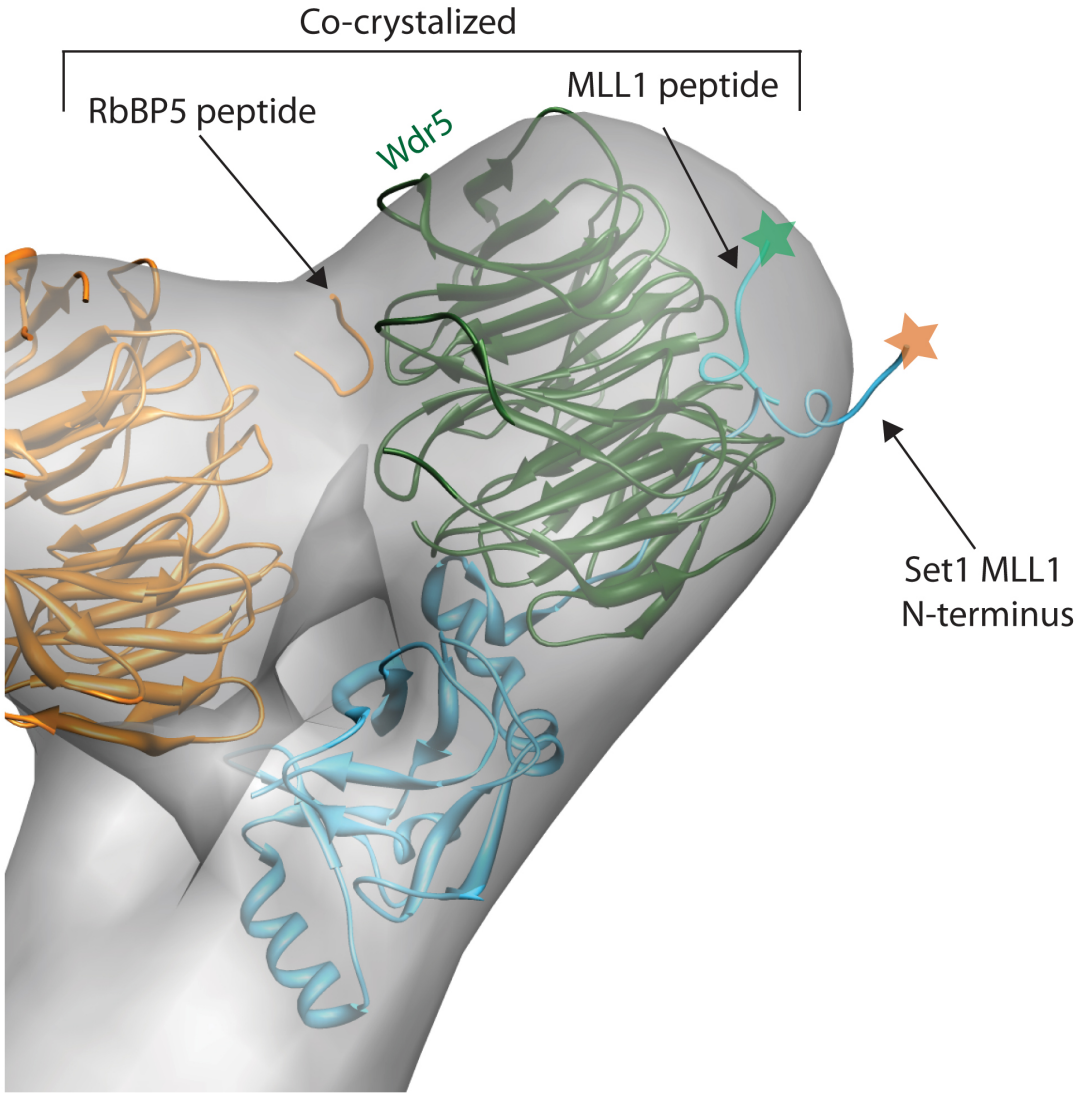


Figure 4-19. Enlarged view of molecular docking.
Close up view of the crystal structure docking of Cps30 homolog (Wdr5, green) and Set1 homolog (MLL1, cyan). The co-crystallized peptide (also in cyan, with Ala-3770 indicated by a green star) in the Wdr5 crystal structure is a segment of the MLL1 protein. The orange star on the N-terminal end of the Set1 MLL1 structure indicates the position of Glu-3790. The orange peptide in the Wdr5 structure is an RbBP5 (Cps50) peptide, shown in proximity of the docked WD40 domain for Cps50 (orange).

COMPASS family members are exo-methylases

Our structural analysis and modeling suggest that the centrally located active site within the COMPASS channel may only be reached by flexible peptide terminals, indicating

that COMPASS family may function primarily as exo- and not endo-methylases. To test this hypothesis, we engineered a Flag sequence to the N-terminus of the only copy of histone H3 in our yeast strain and tested H3K4 methylation by Set1/COMPASS. As shown in figure 4-17C (lanes 1, 3), WT H3 can be methylated on K4 by Set1/COMPASS *in vivo*. However, H3 bearing a single Flag sequence on its N-terminus (making the H3K4 site an internal site) is no longer methylated, although this H3 can be fully methylated on K36 and K79 (Figure 4-17C lanes 5, 6). This finding illustrates that Set1/COMPASS specifically methylates the N-terminal tail of histone H3 and that the addition of a short heterologous sequence to the substrate's N-terminus, making the site of methylation an internal site, blocks methylation by Set1/COMPASS in yeast cells. To further confirm this observation in a reconstituted system, we also tested whether N-terminally 10-His-tagged H3 can be methylated either by Set1/COMPASS or MLL/COMPASS-like complex. Unlike WT H3, an internally engineered H3K4 site can no longer serve as a substrate for COMPASS (Figure 4-20 lanes 1-14) further suggesting that the COMPASS family preferentially methylates N-terminal and not internally lysine sites.

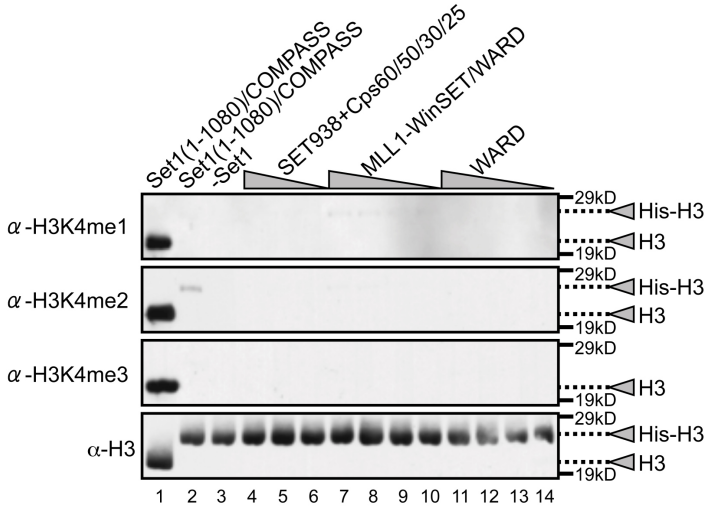


Figure 4-20. COMPASS family are exo-methylases *in vitro*.
Histone H3K4 methylation of N-terminally extended histone H3 by reconstituted Set1/COMPASS and MLL1/COMPASS-likes complexes in vitro. The recombinant N-terminally extended histone H3 was prepared in the histidine-tagged form (His-H3; MGHHHHHHHHHHSSGHIEGRHM-H3) through Ni-NTA column chromatography.

His-H3 was served for the in vitro histone methyltransferase assay using the same enzymes including COMPASS and MLL1 complex as the figure 1B and S5B, respectively. As a positive control, recombinant Set1(1-1080)/COMPASS was used for H3K4 methylation of free histone H3 (lane 1). Consistent with in vivo results, the N-terminal appendage of histone H3 demonstrated diminished H3K4 methylation levels by COMPASS as well as MLL1 complex except very faint bands of H3K4me2 (lane 7) by COMPASS and H3K4me1 by MLL1 complex (lane 7-10). Figure courtesy of the Shilatifard Lab.

4.4 Discussion

In this study, we established the first 3D architecture of *S. cerevisiae* and human COMPASS complexes through characterization of the fully functional core assembly, consisting of the SET domain of Set1, Cps60/Ash2L, Cps50/RbBP5, Cps30/WDR5 and Cps25/DPY30 (Figure 4-16). *In vitro* reconstitution of yeast COMPASS allowed us to prepare multiple truncated forms of Set1 in combination with other subunits (Figure 4-5), and analyzed the generation of H3K4 methylation by these complexes. The characterization of core COMPASS clearly shows that five components are required and sufficient for all forms (mono-, di-, and tri-) of H3K4 methylation *in vitro,* and therefore the complex likely mediates this modification on active genes *in vivo*. Furthermore, the obtained 3D architecture of core COMPASS, revealing the centrally located SET domain of Set1 that directly contacts Cps50, Cps30 and the Cps60-Cps25 subcomplex, provides the fundamental architectural blueprint mediating its enzymatic function and stability.

The core COMPASS subunits identified here are present in all six Set1/COMPASS and MLL1-4 COMPASS-like complexes (Shilatifard, 2008). While a partially reconstituted core MLL complex lacking Dpy30 (Cps25) was able to trimethylate H3K4 *in vitro*(Dou et al., 2006), another study showed that the presence of Dpy30 enhanced the catalytic activity of the MLL complexes (Patel et al., 2009). Recent studies (Patel et al., 2009; South et al., 2010) reveal a conserved physical interaction between Cps60/Ash2 and Cps25/Dpy30 both in yeast and humans. Using both biochemical and structural data, we provide evidence that the Cps60-Cps25 subcomplex interacts directly with the SET domain of Set1 (Figure 4-2E) and activates the *in vitro* H3K4 methyltransferase activity

of core COMPASS (Figure 4-3D). We propose that our reconstituted core yeast and human COMPASS can serve as a model for COMPASS family members in metazoans.

The single-particle EM 2D and 3D analysis of core COMPASS and its subcomplexes have provided insights to the underlying molecular mechanisms of H3K4 methylation. The position of the SET domain, immediately adjacent to each component of core COMPASS, suggests a centralized organization that reflects the importance of the SET motif in catalysis, together with complex stabilization by its closely associated subunits. The histone methyltransferase active site within the SET domain is likely located in the middle of a central channel formed by flanking Cps50, Cps30, and the Cps60-25 subcomplex, suggesting that H3K4 methylation takes place inside the channel. This interpretation provides a structural explanation for the regulation of H3K4 di- and trimethylation by Cps60-Cps25 observed *in vivo* and *in vitro*. The Cps60-Cps25 subcomplex could directly alter the structure of the SET domain to allow an inward shift of iSET, the precise positioning of which was previously proposed to allow MLL1's SET domain to trimethylate H3K4 (Southall et al., 2009). Additionally, the 3D map and modeling of the complex presented here have suggested that COMPASS family members may function as exo-methylases. We have shown that both in a reconstituted system and *in vivo* the COMPASS family does indeed prefer N-terminal lysines as substrate (Fig 17C).

Our reconstitution studies further suggest a novel regulatory role for Cps30 in the progression from H3K4 di- to trimethylation by COMPASS (Figure 4-3). Consistent with our *in vitro* findings with yeast COMPASS, metazoan Cps30 (WDR5) has been shown to be specifically required for H3K4 trimethylation, but not for dimethylation, in *X. laevis* embryos as well as in human *HOX* genes (Wysocka et al., 2005). Although the mechanistic aspect of Cps30-mediated regulation is unclear, it is possible that Cps30, through its proximity and interactions, directly affects the structure of the SET domain in a distinct way from that of Cps60-25, leading to an optimal conformation for H3K4 trimethylation.

4.5 Experimental Procedures

Plasmids and yeast strains.

Full-length Set1, Set1(762-1080), Set1(780-1080), Set1(938-1080), and the Cps subunits of COMPASS (Cps60, Cps50, Cps40, Cps35, Cps30, Cps25) except for Cps15 were fused with the FLAG epitope tag on their N-termini and cloned into the transfer vector, pBacPAK8 (Clontech). The MLL1 core complex components including MLL1-Win-SET(3745-3969), full-length WDR5, Ash2L, RbBP5, and Dpy30 were tagged with FLAG epitope tag on their N-termini and cloned pBacPAK8 vector in the same manner. Yeast shuffle strain, YBL574, was transformed with modified pWZ414-F12 plasmid, which encodes an N-terminally FLAG-tagged hht2 gene, and used for the analysis of exo-methylation by COMPASS.

Protein preparation.

Recombinant COMPASS and MLL1 complexes were prepared through the BacPAK Baculovirus Expression System (Clontech). To prepare COMPASS complexes of various combinations of the Set1 and Cps subunits, exponentially growing Sf9 insect cell cultures were co-transfected with a mixture of viruses expressing specific combinations of COMPASS components and subsequently FLAG-purified. Collected cells were disrupted with dounce homogenization in modified FLAG lysis buffer (10 mM Tris-HCl [pH 8.0], 500 mM NaCl, 10 μM ZnOAc, 1 mM MgOAc, 1 mM imidazole, 10% glycerol, 0.1% NP-40, 10 mM β-mercaptoethanol), and suspensions were clarified with ultracentrifugation. FLAG-tagged COMPASS components were bound to M2 agarose beads (Sigma) for at least 2 hours at 4°C. Following several washes with modified FLAG lysis buffer, COMPASS complexes were eluted with 0.1 mg/ml 3 x FLAG peptide (Sigma) containing modified FLAG lysis buffer. MLL1 complexes were prepared in the same manner. An aliquot of the eluate was applied onto a Superose 6 gel filtration column equilibrated with 50 mM Tris-HCl [pH 8.0], 50 mM NaCl buffer running on a SMART system (GE Healthcare). The purity of prepared samples was analyzed by SDS-PAGE and silver-staining. MLL1 core complex was also prepared in the same manner.

In vitro and in vivo H3K4 methylation analysis.

Recombinant COMPASS complexes were incubated with 0.5 μg of free histone H3 and 200 μM S-adenosylmethionine in methyltransferase reaction buffer (50 mM Tris-HCl

[pH 8.8], 20 mM KCl, 5 mM $MgCl_2$, 0.5 mM dithiothreitol) for at least 2 hours at 30°C. The methylation of histone H3 was examined by Western analysis using anti-H3K4me1, me2, and me3 specific antibodies. Histone methylation in vivo was determined by Western analysis of cleared cell lysates using anti-H3K4me1, me2, me3, H3K79me2, me3, and H3K36me3 specific antibodes.

Specimen preparation, EM imaging, 2D classification and 3D reconstruction of negative stained COMPASS complexes.

Core COMPASS and sub-complexes were prepared for electron microscopy using the conventional negative staining protocol (Ohi et al., 2004), and imaged at room temperature with a Tecnai T12 electron microscope operated at 120 kV using low-dose procedures. Images were recorded at a magnification of 71,138x and a defocus value of ~1.5μm on a Gatan US4000 CCD camera. All images were binned (2 x 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. All particles were manually excised using Boxer [part of the EMAN 1.9 software suite](Ludtke et al., 1999)apart from tilt pairs where particles were selected using WEB (Frank et al., 1996). For antibody labeled COMPASS, Fab fragments were generated using Pierce Fab Micro Preparation Kit (ThermoFischer). Antibody was incubated with COMPASS complex at equimolar concentrations at room temperature for 5 minutes before being adsorbed to EM grids. 2D reference-free alignment and classification of particle projections was performed using SPIDER (Frank et al., 1996). The random conical tilt technique (Radermacher et al., 1987) was used to calculate a first back projection map from individual classes using the images of the tilted specimen. FREALIGN (Grigorieff, 2007) was then used for further refinement of the orientation parameters and for correction of the contrast transfer function (CTF) to produce final 3D reconstructions.

2D classification and 3D reconstructions of negative stained COMPASS complexes.

2D reference-free alignment and classification of particle projections was performed using SPIDER (Frank et al., 1996). 1,809 particle images of Cps25-Cps60, and 3,798 particle images of Cps25-Cps60-Cps50-Set1(938-1080) were classified to 10 and 50 classes, respectively. 12,454 particle images of Dpy30-Ash2L-RbBP5-MLL1-WDR5

were classified into 100 classes. For Cps25-Cps60-Cps50-Set1(938-1080)-Cps30, a total of 35,527 tilt-pair particles pairs were interactively selected from 0º and 60º tilted images (503 pairs). The untilted particle images were subjected to 10 cycles of reference-free alignment and classification into 100 classes. The random conical tilt technique (Radermacher et al., 1987) was used to calculate a first back projection map from individual classes using the images of the tilted specimen. After angular refinement, the corresponding particles from the images of the untilted specimens were added, and the images were subjected to another cycle of refinement. Using the resulting maps as reference models, FREALIGN (Grigorieff, 2007) was used for further refinement of the orientation parameters and for correction of the contrast transfer function (CTF) to produce final 3D reconstructions.

Cryo-EM specimen preparation and imaging.

2 µl of Cps25-Cps60-Cps50-Set1(938-1080)-Cps30 were adsorbed on glow-discharged Quantifoil R2/2 200 mesh grids, and vitrified using a Vitrobot (FEI Mark IV). The specimen was visualized on a Tecnai F20 electron microscope (FEI) equipped with a field emission electron source operated at 200kV. Images were recorded on a Gatan US4000 couple-charge device (CCD) camera under low-dose conditions, at a nominal magnification of 66,964x, and defocus values ranging from -2.0 to -4.0 µm.

Cryo-EM 3D reconstructions of Cps25-Cps60-Cps50-Set1(938-1080)-Cps30.

Particles from cryo-EM images were excised using Boxer [part of the EMAN 1.9 software suite] (Ludtke et al., 1999). The CTF parameters were determined for each micrograph using *ctfit*, and CTF correction was applied accordingly using *Applyctf* [part of EMAN 1.9 package]. A total of 21,583 particles were initially subjected to multiple reference-supervised alignment with EMAN (Brink et al., 2004; Menetret et al., 2005) to separate the projections of full core COMPASS complexes from those missing one of the two WD40 domain. In this process, we used two negative stain 3D reconstructions as initial references: one with two WD40 domains, and another with one WD40 domain, as these were the two populations shown by negative stain analysis. Based on this process, 11,671 particles were stably assigned to the two-lobed COMPASS initial model. These

particles were then subjected to single reference refinement and 3D reconstruction using model-based projection matching in EMAN. The 3D cryo-EM reconstruction was distinctly different from the initial reference, diminishing the possibility of model bias. In addition, very similar 3D reconstructions were obtained using different 3D volumes as starting references, while reprojections of the 3D map were in good agreement with class averages from 2D reference-free alignment and classification of the cryo-EM projections. The final resolution of the core COMPASS complex was calculated to 24 Å @ FSC=0.5 (19.5 Å at FSC = 0.153).

4.6 Acknowledgements

Data from this chapter is contained in the manuscript "Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human" which was published in the *Proceedings of the National Academy of Sciences*, December 2011. Yoh-hei Takahashi in the Shilatifard lab reconstituted, purified, and performed all of the biochemical assays for all of the yeast and human COMPASS complexes in this study. Min Su assisted with trying different computational approaches. Austin Oleskie was a tremendous resource in helping to pick many, many, many particles.

4.7 References

Avdic, V., Zhang, P., Lanouette, S., Groulx, A., Tremblay, V., Brunzelle, J., and Couture, J.F. (2011). Structural and biochemical insights into MLL1 core complex assembly. Structure *19*, 101-108.

Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. Nat Struct Mol Biol *14*, 1008-1016.

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.

Cao, F., Chen, Y., Cierpicki, T., Liu, Y., Basrur, V., Lei, M., and Dou, Y. (2010). An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. PLoS One *5*, e14102.

Dehe, P.M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., Rodriguez-Gil, A., Mkandawire, M., Landsberg, K., Shevchenko, A.*, et al.* (2006). Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. J Biol Chem *281*, 35404-35412.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol *13*, 713-719.

Eissenberg, J.C., and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol *339*, 240-249.

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.

Grigorieff, N. (2007). FREALIGN: high-resolution refinement of single particle structures. J Struct Biol *157*, 117-125.

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.

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.

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.

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.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Mol Cell *29*, 577-587.

Odho, Z., Southall, S.M., and Wilson, J.R. (2010). Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. J Biol Chem *285*, 32967-32976.

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.

Patel, A., Dharmarajan, V., Vought, V.E., and Cosgrove, M.S. (2009). On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. J Biol Chem *284*, 24242-24256.

Patel, A., Vought, V.E., Dharmarajan, V., and Cosgrove, M.S. (2011). A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. J Biol Chem *286*, 3359-3369.

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.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J *20*, 7137-7148.

Schlichter, A., and Cairns, B.R. (2005). Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. EMBO J *24*, 1222-1231.

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.

South, P.F., Fingerman, I.M., Mersman, D.P., Du, H.N., and Briggs, S.D. (2010). A conserved interaction between the SDI domain of Bre2 and the Dpy-30 domain of Sdc1 is required for histone methylation and gene expression. J Biol Chem *285*, 595-607.

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.

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.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell *121*, 859-872.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Chapter 5 Discussion and Future Directions

5.1 β_2AR-Gs Complex

Better understanding how G protein coupled receptors (GPCRs) interact and relay signals from the extracellular environment to their intracellular G proteins is a major topic for study. GPCRs are the target for nearly one-third (~33%) of all pharmaceutical drugs. β_2AR has been the model system for understanding how GPCRs function, and there have been a significant amount of biochemical and pharmacological studies exhausting the different mechanisms of activation and silencing of the receptor. While these types of studies have been quite informative, structural information can contribute crucial insights to the function of GPCRs and how they interact with their intracellular companions.

Electron microscopy as a tool for studying membrane proteins

As mentioned earlier, the structural studies of membrane proteins are very challenging for a number of different reasons. Part of these issues are now better dealt with given advances to aid in the purification, solubilization, stabilization, and crystallization of a number of GPCRs and other integral membrane proteins. The current capabilities of electron microscopy allow the use of negative stain EM to assess architectural changes, as shown in characterizing the movement of the alpha helical domain in the β_2AR-Gs complex (Westfield et al., 2011). EM also played a pivotal role in determining ideal conditions that led to the X-ray crystal structure of the first GPCR/G-protein heterotrimeric complex (Rasmussen et al., 2011). The use of negative stain EM to screen numerous conditions of the purified β_2AR-Gs complex to evaluate the best purification and complex conditions and proved to be an essential technique for successful crystallogenesis (Figure 5-1). However, there are potential pitfalls in using EM to study membrane proteins. In the case of receptor-protein complexes, such as β_2AR-Gs and β_2AR-arrestin, alignment of less prominent features of the specimen can lead to complications in both negative and cryo-EM. Specimens in negative stain can

have further collapse and adapt non-native conformations on the carbon support. Also, with cryo-EM, the lower contrast makes it very challenging to align particles accurately. The presence of detergent near the critical micelle concentration (CMC) can also pose imaging and alignment issues for cryo-EM. For example, in the case of β_2AR-Gs the MNG detergent micelles appeared as particles nearly the same size as the specimen, which made both particle selection and alignment extremely difficult.

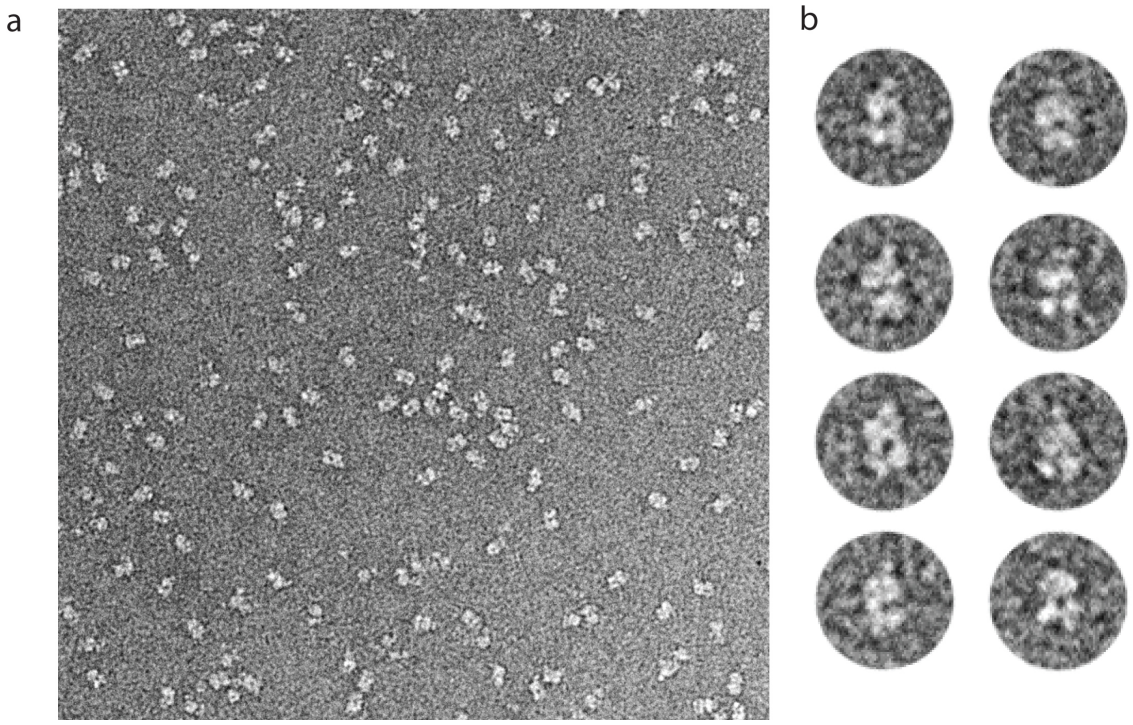


Figure 5-1. a) Image of negative stained β_2AR-Gs complex. b) Boxed out particle projections of β_2AR-Gs complex.

While many biochemical techniques, including size exclusion chromatography (SEC), showed the purified complex to be a single population, when imaged with negative stain EM it was shown that the complex remained very conformationally heterogeneous. Additionally, screening the β_2AR-Gs complex by EM incubated with different camelid antibodies (nanobodies) identified optimal conditions used for crystal formation. This provides further evidence that EM is very useful as a standalone technique, or in conjunction with other structural techniques to achieve structural information of membrane proteins.

Future of EM with membrane proteins/protein complexes

The use of negative stain EM in studying the β_2AR-Gs complex was needed due to the small size of the complex (~130 kDa) and the lack of strongly distinguished features. Negative staining relies on a specimen laying on a carbon support, which may induce preferred orientations while also being dehydrated and embedded in heavy metal solution that can cause particle collapse. Ideally, cyro-EM would be the preferred technique to use because it preserves the specimen in a physiological buffer of vitreous ice that does not deform the sample while allowing different 3D orientations. Currently imaging, aligning, and processing small specimens can be extremely challenging using cryo-EM. However, with new direct electron detection or phase plate technology, imaging of small proteins can be improved (Danev and Nagayama, 2001). While most standard CCD cameras take a single exposure, new direct electron detectors camera systems can collect a series of shorter exposure frames and then align these series of frames to generate a single aligned image that can correct for drift and other environmental aberrations during imaging to generate improved contrast. Phase plate imaging produces a cosine-type phase contrast transfer function (CTF starts at 1) as supposed to standard TEM imaging that produces a sine-type (CTF starts at 0) function producing images with highly improved contrast by better preserving the lower frequency spectrum.

Characterizing different states within a population

Electron microscopy also has the capability to characterize a number of different states co-existing within a given specimen based on applying computational sorting and using single particle analysis. Variations within a population of particles can be illustrated by directly comparing class averages of particle projections. Comparatively, while X-ray crystallography can deliver high-resolution data, this information is only a snapshot of a single conformation of a specimen packed in a crystal lattice with the exception of the unconventional case where two or more molecules are in an asymmetrical unit and may adopt different conformations. When the crystal structure of Gαs with GTPγS was solved (Sunahara et al., 1997), the alpha helical domain was shown stabilized below the nucleotide-binding pocket at the AH/Ras interface (Figure 5-2B). However, in the nucleotide-free state in complex with β_2AR (Rasmussen et al., 2011), the AH-domain is

shown behind the Ras domain (Figure 5-2A). The EM studies reveal that in solution, the orientation of the AH-domain can take on a variety of positions, best observed when Nb37 is used as a domain marker tracking the AH position. These experiments also helped illustrate that the AH-domain positioning is not just a crystallographic artifact. The variation of states is further evidenced by the incubation of the β_2AR-Gs complex with GTPγS. The class averages clearly show a mixture of different transient states of the complex where a portion of the population is intact, another part of the population partially is dissociated, and a percentage of the population is completely dissociated to individual subunits (Figure 5-3) (Westfield et al., 2011).

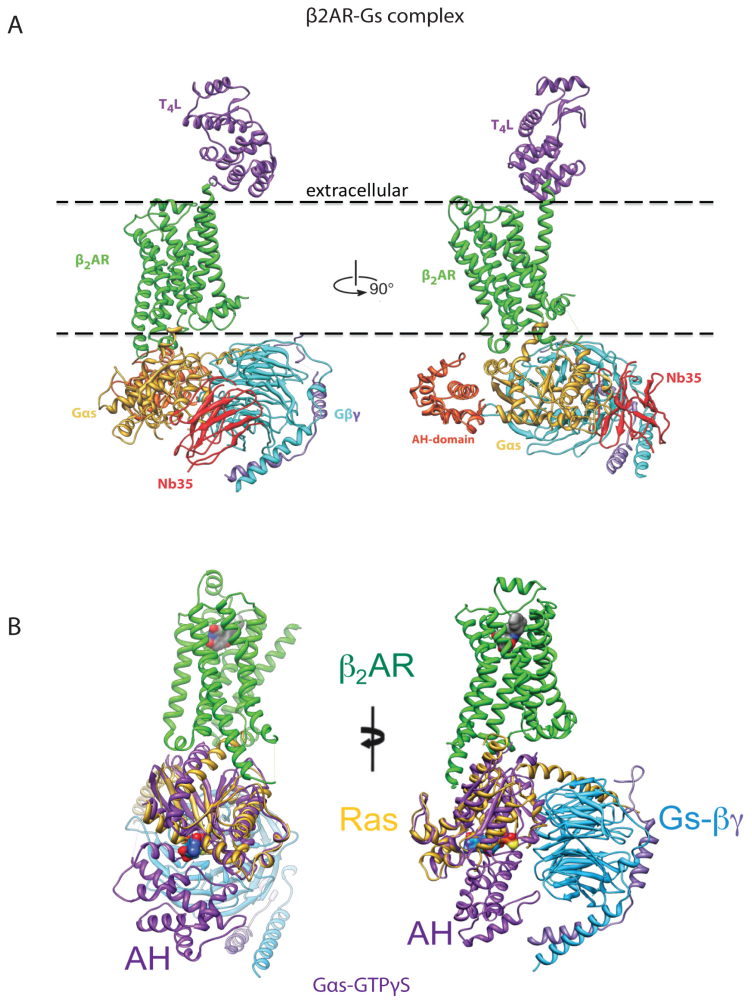


Figure 5-2. Crystal structures of β_2AR-Gs complex and Gαs illustrating variable alpha helical domain positioning.
A) X-ray crystal structure of β_2AR-Gs complex with nanobody 35 (Nb35, red). N-terminal T4 DNA lysozyme (Purple) attached to the β_2AR (green). The structure shows

the alpha helical domain (orange) behind the Ras domain of the Gα subunit (yellow) in complex with the βγ heterodimer (cyan, and light purple respectively). B) The structure shows the crystal structure of Gαs-GTPγS complex (Sunahara, 1997) in purple, superimposed on the Ras-domain of the β_2AR-Gs complex. This confirmation shows the alpha helical domain below the Ras domain of Gαs.

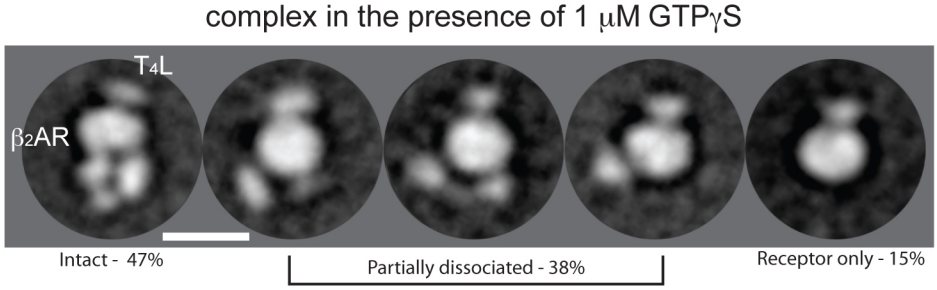


Figure 5-3. Transient states of β_2AR-Gs complex with GTPγS.
Representative class averages of the T4L-β_2AR-Gs complex after rapid mixing with GTPγS (1 μM) and immediate stain embedding reveal both intact as well as partially or fully dissociated complexes. The fraction of corresponding sub-populations is indicated. The scale bars correspond to 10 nm.

Alpha helical (AH) domain flexibility of a G protein

Crystal structures of G proteins have provided a remarkable amount of important information into their activation mechanism. Conformational changes occur in three regions of the α subunit termed, Switch I-III, which reside in the Ras domain and Switch I is located between the Ras and AH-domain (Sprang, 1997). The presence of the AH-domain has been shown to slow the rate of nucleotide exchange as well (Markby et al., 1993). Incubation of the β_2AR-Gs complex with different nucleotides (GDP and GTPγS) and nucleotide mimics (PPi and foscarnet) was shown to stabilize the AH-domain below the Ras domain, further indicating that the AH flexibility is somehow involved in G protein activation. Additionally, a number of biochemical and biophysical studies have also revealed flexibility of the AH-domain. In hydrogen deuterium exchange studies, the solvent accessibility at the Ras-AH interface was increased during receptor binding (Chung et al., 2011), suggesting that receptor activation causes a conformational change moving the AH-domain away from the binding pocket to help facilitate nucleotide exchange. In another study employing electron-electron resonance

spectroscopy, the increased distance between the Ras and AH-domain corresponded with increased nucleotide release (Van Eps et al., 2011).

5.2 β-Arrestin / β_2V_2 / Fab complex

The first architecture of β-Arrestin complexed with a GPCR was characterized through the *in celltro* purification method and by employing negative stain EM. Visualizing the complex has provided important information regarding domain arrangement and subunit interactions. There have been many studies done characterizing the interactions between Arrestins and receptors including fluorescent tags and co-immunoprecipitation, but this is the first study actually visualizing this relationship (Aguila et al., 2011; Sommer et al., 2005).

Arrestin – receptor variable association

From the negative stain analysis, it was shown that there are numerous conformations that Arrestin can occupy with a single connectivity hinged and hanging from the receptor, or with Arrestin sitting "snuggly" perpendicular to the receptor. For a better understanding of these associations the complex was cross-linked to further stabilize its position on the receptor. This led to essentially all particles to have a more tightly fitting, or snuggly, conformation on the receptor. This suggests that there is a transient association between the more loosely associated arrestin conformation (Figure 5-4, second from right), and where Arrestin appears to be more closely associated with the receptor (Figure 5-4, second from left).

The biological significance of the two conformations, loose versus tight, seen in the native complex is not known. While βarr was originally thought to function primarily for receptor desensitization, the new structural insights of these conformations reveals new possibilities of their function. β-arrestin also functions as a scaffolding protein, and it could be hypothesized that the loose conformation is more efficient for scaffolding than the more tightly associated form, in order to make complexes with additional proteins.

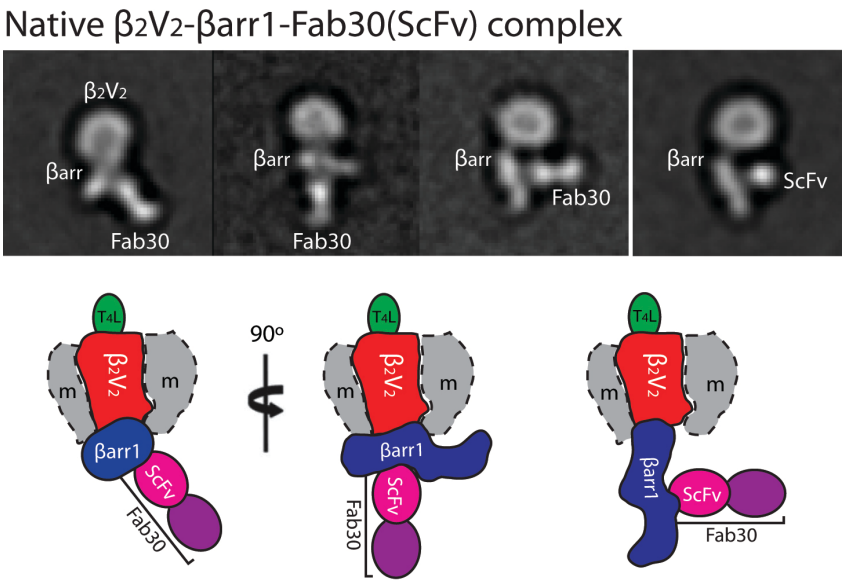


Figure 5-4. Different conformation of β-Arrestin complex.
Top, representative class averages of the different conformations the β_2V_2/β-Arrestin/Fab(ScFv)30 complex. Bottom, representative models for each domain of the complex.

Future studies for β-Arrestin-GPCR

While negative stain EM gives low-resolution information about the overall domain architecture of β-Arrestin and β_2V_2, a more detailed structure from X-ray crystallography would provide atomic resolution insights into the receptor-arrestin interaction. Some of the difficulty in obtaining a high-resolution structure is the ability to purify a conformationally stable complex that can lead to crystal formation. In the case of the β_2V_2/β-Arrestin/Fab(ScFv)30 complex, it became very clear that the inherent flexibility of the native complex would lead to crystallization issues. Cross-linking would improve the stability of the complex, however, it became evident from the negative stain and cryo EM class averages that there was still inherent flexibility between the Fab-βarr-receptor, and additional engineering of the construct may be needed in order to better prepare the specimen for crystallization. Additionally, binding the β_2V_2/β-Arrestin/Fab30 complex with different nanobodies, as was done with the β_2AR-Gs complex, could also lead to improved stability of the complex and improved crystallogenesis.

5.3 COMPASS Histone Methyl transferase

Studying the yeast COMPASS histone methyltransferase has provided critical insight into the necessary subunits to facilitate H3K4 trimethylation, a highly impactful epigenetic signal. The architecture shown by the cryo-EM work suggests a channel formed between the Set1 catalytic domain, the Cps60/Cps25 dimer, and the two WD40 domains of Cps50 and Cps30. The single particle EM of the core COMPASS components has provided many insights into the underlying mechanism of H3K4 methylation. The molecular modeling, specifically the positioning of the catalytic Set1 (truncated) subunit, suggests a centralized organization with the active site likely located in the middle of the central channel formed by the surrounding subunits previously mentioned. The architecture of the central channel along with the other biochemical data further suggests the exo-methylase activity of COMPASS, and it is shown in Figure 4-17C that in both reconstituted complexes and *in vivo,* COMPASS prefers N-terminal lysines as substrate.

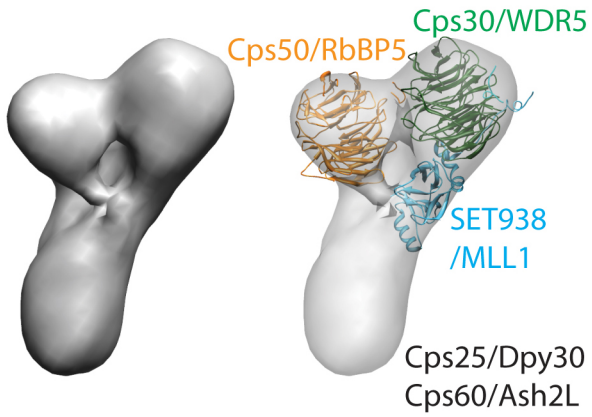


Figure 5-5. Cryo-EM reconstruction of COMPASS complex.
Cryo-EM 3D reconstruction of the COMPASS complex with docked known crystal structures.

Understanding COMPASS beyond the 'core' complex

2D negative stain EM was performed on the COMPASS complex consisting of Cps60-Cps50-Cps40-Cps30-Cps25-SET762 subunits, termed '+Cps40-SET762' to compare it directly to the 'core' complex (Cps60-Cps50-Cps30-Cps25-SET938) with the intention of characterizing the the Cps40 subunit localization and its effect on the overall

architecture (Figure 4-11). Biochemical studies have shown that Cps40 further regulates H3K4 trimethylation (Takahashi et al., 2009), and based on the 2D class averages it is shown to interact with the WD40 domains of Cps50 and Cps30 subunits. Histone methyl transferase (HMT) activity assays have shown that Cps30 is required specifically for enhanced trimethylation of H3K4 (Figure 4-3d), so the close proximity between Cps40 and Cps30 would suggest that this interaction further regulates trimethylation.

Future COMPASS characterization

The structural studies described in Chapter 4 primarily involved the necessary or 'core' components of the COMPASS complex. COMPASS and conserved COMPASS-like complexes in higher organisms have many similar components. Comparison of 2D class averages revealed that the 'core' yeast COMPASS complex and human complex have very similar architecture (Figure 4-7). In future structural studies, the characterization of the full yeast COMPASS complex, as well as its full human COMPASS-like complexes (MLL1-4 complexes) may provide additional perspective into their epigenetic function and regulation. Additionally, COMPASS forms larger complexes with transcriptionally active RNA polymerase II (Pol II) mediated by its interactions with Paf1 (Krogan et al., 2003). Large protein assemblies of this kind are ideal for cryo-EM studies.

While a number of homologous COMPASS complexes are very good candidates for EM characterization, the ideal target for EM 3D reconstruction would be COMPASS bound to a mononucleosome. This would help characterize the interactions and overall topology of the COMPASS/nucleosome relationship. The current 3D reconstruction presented in Chapter 4 helps to generate new models and hypotheses about this interaction, but an actual 3D reconstruction of this physiological association would make an extremely significant contribution to further understanding this system. The difficulty for generating this lies in the biochemistry to purify homogeneous samples of mononucleosome and COMPASS, and preparing this in a stably bound 1:1 complex for single particle analysis.

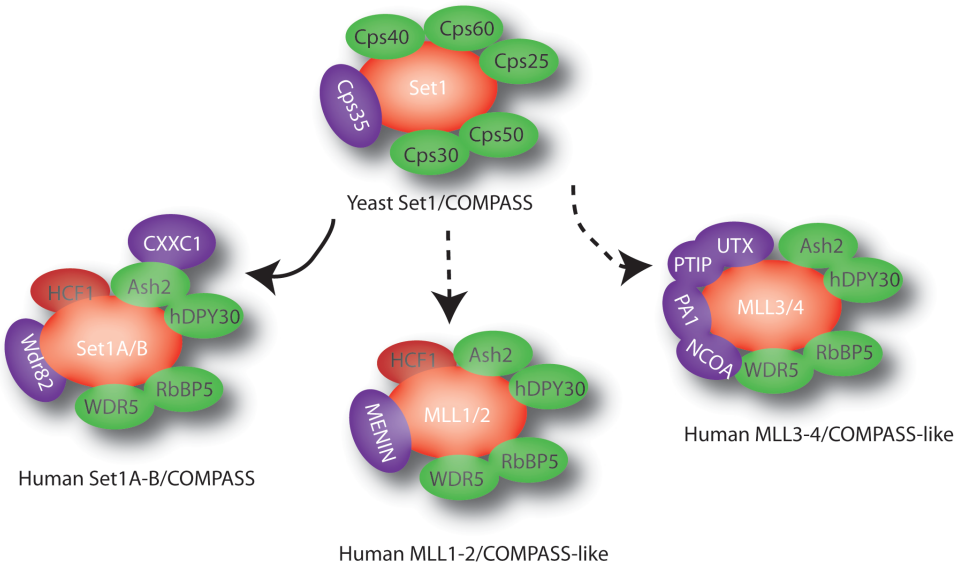


Figure 5-6. Subunit composition of COMPASS and mammalian homologs.
The founding member of H3K4 methyltransferases, Set1/COMPASS was discovered in yeast. All of the conserved core components are shared amongst all complexes; Cps30 (WDR5), Cps50 (RbBP5), Cps25 (Dpy30), and Cps60 (Ash2). Human Set1 / COMASS has Wdr82 and CXXC1 which are homologus to Cps40 and Cps35 in yeast that help regulate trimethylation in vivo. HCF1 is shared between Set1/COMPASS and MLL1-2 complexes. MLL1/2 contains tumor suppressor Menin as a subunit. The MLL3/4 complexes contain PTIP, PA-1, and NCOA6 which helps targeting the complex to specific genes and a H3K27 demethylase UTX.

Implications of histone crosstalk

Previous studies have suggested that COMPASS H3K4 methylation is partially regulated by histone crosstalk with histone H2B monoubiquitination. COMPASS purified from yeast strains lacking H2B monoubiquitination have several fold lower Cps35 presence in the complex (Lee et al., 2007). This, combined with other biochemical evidence led to the proposed hypothesis of Cps35/monoubiquitinated-H2B cross talk. However, more recent studies have suggested that the n-SET domain of Set1, and not Cps35, is what regulates this COMPASS/H2B crosstalk (Kim et al., 2013). Further investigation biochemically as well as structurally could be used to further answer these relationships.

Implications for other SET domain containing complexes

Repressive H3K27 methylation, which has an apparently opposite function to H3K4 methylation, is also introduced by the SET domain-containing multi-protein complex

Polycomb Repressive Complex 2 (PRC2) (Cao et al., 2002). COMPASS and PRC2 share common features, such as closely related SET-containing catalytic subunits that require complex formation for activity *in vivo* and *in vitro* (Kuzmichev et al., 2002; Nekrasov et al., 2005), and also the presence of multiple WD40 proteins (RbBP5 and WDR5 in MLL1/COMPASS, and EED and RbAP46/48 in PRC2)(Kuzmichev et al., 2004). It will thus be interesting to investigate whether and how substrate specificities, enzymatic mechanisms, and complex architecture are common or diverged between COMPASS and PRC2.

COMPASS – PRC2 architectural comparison

Polycomb Repressive Complex 2 (PRC2), part of the Polycomb group (PcG) protein complexes, is essential for gene silencing by repressing transcription of specific genes via histone H3 lysine 27 trimethylation (H3K27-me3) (Schmitges et al., 2011). As mentioned previously, many of the subunits of the PRC2 complex have great similarity to that of the COMPASS complex. An EM structure characterizing reconstituted PRC2 complex was published in October 2012 by Nogales and colleagues (Ciferri et al., 2012).

In Ciferri et al., a pentameric PRC2 complex (Ezh2/EED/Suz12/RbAp48/AEBP2) is reconstituted, and a negative stain 3D reconstruction is calculated initially from tilt-pairs of a single class average, and then subsequent iterative angular refinement from additional 0º projections to a 21Å (0.5=FSC) resolution. A striking similarity of the sample preparation between the two complexes is their inherent flexibility. In the COMPASS complex, a variation of distances was observed between the two WD40 domains (Figure 4-12). Additionally with COMPASS, 2 populations were observed containing one or two (1 or 2) WD40 domains. In the PRC2 complex, their initial structural characterization of the specimen indicated a large amount of flexibility, that in order to preserve complex stability the sample was cross-linked using glutaraldehyde.

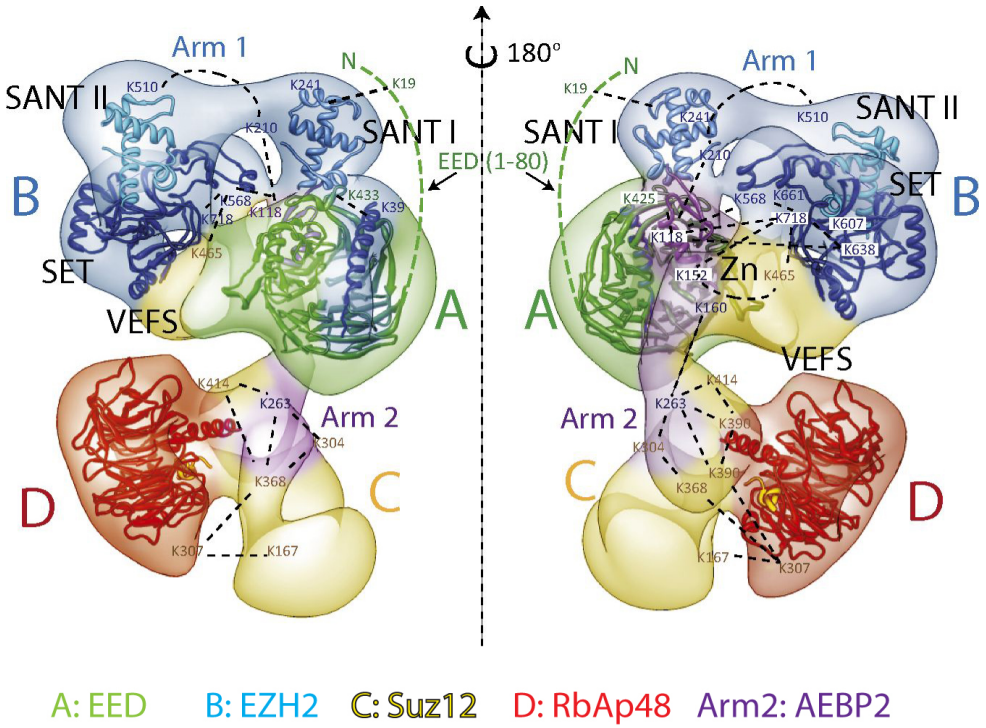


Figure 5-7. EM reconstruction and architecture of reconstituted PRC2-AEBP2 complex.
Modified figure (Ciferri *et al.* 2012) showing the overall architecture and subunit organization of the PRC2-AEBP2 complex. The two WD40 domains of EED and RbAp48 are shown in green and red respectively. The SET-containing Ezh2 subunit is shown in blue. The non-modeled Suz12 and AEBP2 domain are shown in yellow and purple respectively.

When comparing the overall architecture of the two complexes, while their individual subunit composition within the complex is similar, the organization is different. Both complexes contain a SET-domain containing catalytic subunit and two WD40-domain containing subunits. In the COMPASS complex, the two WD40 domains (Cps30 and Cps50) are juxtaposed to each other forming a substrate channel with the catalytic SET-domain and Cps60/25 subcomplex (Figure 4-17a & 5-5). In the PRC2 complex, the two WD40 domains of EED and RbAp48 are somewhat diametrically opposed from one another on the complex (Figure 5-7, domains A and D). A consistent feature between the H3K4 (COMPASS) and H3K27 (PRC2) methylases is the centralization of the SET-domain within the complexes. For PRC2, while the WD40 domains are not juxtaposed to each other as seen in the COMPASS complex, EED and RbAp48 do flank the SET-domain forming a central cavity within the complex. While the subunit organization

between the two complexes are not identical, characteristics of flanking WD40 containing subunits around the catalytic SET-domain appears somewhat consistent.

5.4 Concluding Remarks

Applying electron microscopy to investigate the architecture of the β_2AR-Gs complex has provided an incredible level of insight into the dynamics of receptor/G-protein interactions. Single particle analysis has illuminated the different conformations of the AH-domain, in addition to capturing sequential transient states of the complex's dissociation when incubated with nucleotide (GTPγS). Additionally, the use of EM played a crucial role in screening the conditions that led to the eventual crystallization and high-resolution structure of the first GPCR and G-protein heterotrimer in complex. The AH-domain in the crystal structure was located in an unsuspecting position, presumably due to crystallographic contacts. EM helped illustrate the flexibility of the AH-domain, which is a key finding. Incubating the β_2AR-Gs complex with a nanobody not only stabilized the complex for crystallization, but also additionally showed that nanobodies can be used as domain markers to track the movement of small domains (AH-domain). These techniques have helped pioneer approaches that can be used in the future when working to determine the structure and dynamics of relatively small membrane proteins and protein complexes. With advances in electron microscopic hardware and software, cryo-EM can be used to study smaller membrane protein complexes as well.

The electron microscopy studies of the β_2V_2/β-Arrestin/Fab(ScFv)30 complex give novel insights about the architecture of a GPCR in complex with an arrestin. The *in celltro* approach of antibody-based complex stabilization is also an innovative and useful technique for purification of unstable protein complexes. On-column cross-linking size exclusion chromatography played a critical role in purifying a conformationally homogeneous, sample for single particle analysis. Additionally, mass spectroscopy used to identify cross-linked residues provided information regarding interacting regions between the receptor and arrestin in order to more precisely dock high-resolution structures in the EM maps. Additionally, single-particle EM analysis revealed the first snapshots of the different association states between β-arrestin and a GPCR. While the

biological significance of these different conformations is not fully understood, observing these conditions allows for new models about the β-arrestin/GPCR dynamics.

The EM studies of COMPASS have also offered critical insights to the complex architecture. The yeast COMPASS histone methyltransferase complex was the first H3K4 methylase discovered, and understanding its architecture and component makeup helps delineate the mechanism of its trimethylase activity. These structural insights will lead to understanding the function of COMPASS and other COMPASS-like complexes in higher organisms. Now that the overall architecture of core COMPASS is revealed, further studies can be undertaken to better characterize the epigenetic crosstalk that is involved in the H3K4 regulation and other types of epigenetic signaling.

All of the protein complexes investigated in these studies are centrally involved in important signaling events. The β_2AR-Gs complex and the β_2AR/β-arrestin complex serve as paradigms for GPCR-based intracellular signal instigation and silencing, while the COMPASS complex is involved in various mechanisms of epigenetic regulation. These processes are now better understood based on the structural characterization obtained by electron microscopy combined with other biochemical approaches. Single particle negative stain and cryo-EM are very powerful tools and have immense potential for understanding the architecture and dynamics of macromolecular complexes.

5.5 References

Aguila, B., Simaan, M., and Laporte, S.A. (2011). Study of G protein-coupled receptor/beta-arrestin interactions within endosomes using FRAP. Methods Mol Biol *756*, 371-380.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Chung, K.Y., Rasmussen, S.G., Liu, T., Li, S., DeVree, B.T., Chae, P.S., Calinski, D., Kobilka, B.K., Woods, V.L., Jr., and Sunahara, R.K. (2011). Conformational changes in the G protein Gs induced by the beta2 adrenergic receptor. Nature *477*, 611-615.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. Elife *1*, e00005.

Danev, R., and Nagayama, K. (2001). Transmission electron microscopy with Zernike phase plate. Ultramicroscopy *88*, 243-252.

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.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M.*, et al.* (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell *11*, 721-729.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell *14*, 183-193.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev *16*, 2893-2905.

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.

Markby, D.W., Onrust, R., and Bourne, H.R. (1993). Separate GTP binding and GTPase activating domains of a G alpha subunit. Science *262*, 1895-1901.

Nekrasov, M., Wild, B., and Muller, J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep *6*, 348-353.

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.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

Sommer, M.E., Smith, W.C., and Farrens, D.L. (2005). Dynamics of arrestin-rhodopsin interactions: arrestin and retinal release are directly linked events. J Biol Chem *280*, 6861-6871.

Sprang, S.R. (1997). G protein mechanisms: insights from structural analysis. Annu Rev Biochem *66*, 639-678.

Sunahara, R.K., Tesmer, J.J., Gilman, A.G., and Sprang, S.R. (1997). Crystal structure of the adenylyl cyclase activator Gsalpha. Science *278*, 1943-1947.

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 Eps, N., Preininger, A.M., Alexander, N., Kaya, A.I., Meier, S., Meiler, J., Hamm, H.E., and Hubbell, W.L. (2011). Interaction of a G protein with an activated receptor opens the interdomain interface in the alpha subunit. Proc Natl Acad Sci U S A *108*, 9420-9424.

Westfield, G.H., Rasmussen, S.G., Su, M., Dutta, S., DeVree, B.T., Chung, K.Y., Calinski, D., Velez-Ruiz, G., Oleskie, A.N., Pardon, E.*, et al.* (2011). Structural flexibility of the G alpha s alpha-helical domain in the beta2-adrenoceptor Gs complex. Proc Natl Acad Sci U S A *108*, 16086-16091.