

# **Structure and Function of WD40 domain Proteins**

Chao Xu<sup>1</sup> and Jinrong Min<sup>1,2</sup>

1. Structural Genomics Consortium, University of Toronto, 101 College St., Toronto, Ontario, M5G 1L7, Canada.
2. Department of Physiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Corresponding Author:

Jinrong Min

Structural Genomics Consortium

University of Toronto

101 College St., Toronto, On

M5G 1L7, Canada

Tel: 416-9463868

Fax: 416-946-0588

Email: jr.min@utoronto.ca

## **Abstract**

The WD40 domain exhibits a  $\beta$ -propeller architecture, often comprising seven blades. The WD40 domain is one of the most abundant domains and also among the top interacting domains in eukaryotic genomes. In this review, we will discuss the identification, definition and architecture of the WD40 domains. WD40 domain proteins are involved in a large variety of cellular processes, in which WD40 domains function as a protein-protein or protein-DNA interaction platform. WD40 domain mediates molecular recognition events mainly through the smaller top surface, but also through the bottom surface and sides. So far, no WD40 domain has been found to display enzymatic activity. We will also discuss the different binding modes exhibited by the large versatile family of WD40 domain proteins. In the last part of this review, we will discuss how post-translational modifications are recognized by WD40 domain proteins.

## Identification of WD40 repeat and WD40 $\beta$ -propeller domain

WD40 repeat was first identified in the  $G_{\beta}$  subunit of the heterotrimeric G proteins and the CDC4 protein as a repetitive sequence motif of about 43 amino acids, which includes highly conserved residues like the GH dipeptide and a conserved aspartic acid residue 6 residues before the signature WD dipeptide (Fong et al., 1986). The name of WD40 was derived from the conserved WD dipeptide and the length of approximately 40 amino acids residues in a single repeat. The definition of this motif was further refined by Neer et al (Neer et al., 1994; Smith et al., 1999) as a 44-60 residue sequence unit that typically contains the GH dipeptide 11-24 residues from its N-terminus and the WD40 dipeptide at the C-terminus, following a sequence pattern as shown in Fig. 1A. Each WD40 repeat comprises a four-stranded antiparallel  $\beta$ -sheet (Fig. 1B).

The WD40 repeats in a protein fold into a  $\beta$ -propeller architecture (Fig. 1B and 1C). Since the first WD40 domain structure was determined 15 years ago (Lambright et al., 1996; Wall et al., 1995), tens of WD40 structures have been determined so far. In the vast majority of cases, the three strands of the last blade form a  $\beta$ -sheet with the outer strand from the first blade, creating a “velcro” closure to further stabilize the  $\beta$ -propeller folding in addition to the extensive hydrophobic interactions between the  $\beta$ -sheets (Fig. 1B and 1C). WD40 repeats are thought to arise through intragenic duplication and recombination events (Andrade et al., 2001), and diversify during evolution. It was therefore thought that the sequences and structures of the repeats in the same protein are more similar to each other than those from different proteins. Therefore, we chose four WD40 proteins determined in our laboratory and compared it to the  $G_{\beta}$  protein, and found that this rule does not hold for WDR61 and WDR92 (Fig. 1D). It seems that the WD40 proteins arise at different evolution timescales. Those WD40 proteins with low self-similarity arose at an early evolution stage, such as WDR61 and WDR92, and proteins like WDR5 arise relatively recently.

WD40 domain containing proteins are very abundant in eukaryotic organisms, and are rarely present in prokaryotes, although a few cases were reported in bacteria (Stirnemann et al., 2010). This domain is among the top ten most abundant domains in eukaryotic genomes (Stirnemann et al., 2010). In the human genome, there exist 349 WD40 domain containing proteins based on SMART database (Letunic et al., 2009), and this number should be very conservative considering the difficulties in identifying WD40 domains due to the low sequence conservation. The WD40 domain proteins are involved in a wide range of cellular functions, such as signal transduction, vesicular trafficking, cytoskeletal assembly, cell cycle control, apoptosis, chromatin dynamics and transcription regulation. The malfunctions of some WD40 proteins have been reported to be responsible for different diseases (Li and Roberts, 2001). The common function of these WD40 domain proteins is to serve as a rigid scaffold for protein-protein interaction, and to coordinate downstream events, such as ubiquitination and histone methylation. WD40 domain was ranked as the top interacting domain in *S. cerevisiae* (Stirnemann et al., 2010). WD40 domains were also among the top interacting domains based on the limited human interaction data (Stirnemann et al., 2010).

### **Architecture of WD40 domains**

A single WD40  $\beta$ -propeller can contain 4 to 8 WD40 repeats in principle (Paoli, 2001). So far, only 7- or 8-blade WD40  $\beta$ -propeller has been confirmed structurally (Fig. 2A and 2B). Lower number of repeats of  $\beta$ -propeller have been confirmed in other non-WD40 repeat proteins (Paoli, 2001). Interestingly, on the basis of the geometry modeling, it was predicted that the 7-fold  $\beta$ -propeller is the most ideal  $\beta$ -sheet geometry (Murzin, 1992). Consistently, 7-fold  $\beta$ -propeller proteins dominate in the solved WD40 structures and identified WD40 proteins.

Sec13 is a WD40  $\beta$ -propeller protein and functions as a component of the COPII coat and the nuclear pore complex (NPC). The WD40 domain in the Sec13 protein contains only 6 WD40 repeats and it forms an open  $\beta$ -propeller, needing to borrow one more WD40

repeat from its complex partner ACE1 (ancestral coatomer element 1) protein Sec31 or Sec16 to complete a closed  $\beta$ -propeller fold in the COPII coat (Whittle and Schwartz, 2010) (Fig. 2C). This phenomenon also exists in the nuclear pore complex. In the subcomplexes of the Seh1 and Nup85 nucleoporin pair and the Sec13-Nup145C nucleoporin pair, both Seh1 and Sec13 form a 6-repeat  $\beta$ -propeller, and the  $\beta$  sheet at the N terminus of Nup85 or Nup145C donate an insertion blade respectively to Seh1 or Sec13, completing the  $\beta$ -propeller of Seh1 and Sec13 (Brohawn et al., 2009).

In addition to single WD40  $\beta$ -propeller, some proteins contain multiple WD40  $\beta$ -propellers in an array. For instance, Aip1 contains two 7-blade WD40  $\beta$ -propellers (Fig. 2D). Aip1 (Actin-interacting protein 1) is a WD40 domain containing protein involved in enhancing actin filament disassembly in the presence of actin-depolymerizing factor (ADF)/cofilin, and also capping the barbed end of ADF/cofilin-bound actin filament (Mohri et al., 2004; Voegtli et al., 2003). DDB1 contains three 7-blade WD40  $\beta$ -propellers (Fig. 2E) (Li et al., 2006), which is required for DNA repair (Wittschieben et al., 2005). It also functions as a component of the DDB1-CUL4-Roc1 E3 ubiquitin ligase complexes, which mediate the ubiquitination, and subsequent proteasomal degradation of its target proteins (Chen et al., 2001; Higa et al., 2006; Li et al., 2006).

The majority of the  $\beta$ -propeller domains has a “velcro” closure, which contributes to fold stabilization. But, this interlock mechanism is absent in some cases, such as Aip1 (Voegtli et al., 2003) and DDB1 (Voegtli et al., 2003). That means that  $\beta$ -propeller fold is mainly stabilized through the intersheet hydrophobic interactions. So far, DDB1 structure had been determined in three different crystal forms (Fig. 2F). Superposition of these three structures reveals that the individual WD40  $\beta$ -propeller is very rigid, and even the first WD40 domain without the Velcro closure is also very rigid (Fig. 2F). That means that the “velcro” closure is not essential for the WD40 domain formation. Interestingly, the first two  $\beta$ -propellers of DDB1 can be well superimposed, but the third  $\beta$ -propeller of DDB1 exhibited a significant shift relative to the first two  $\beta$ -propellers (Fig. 2F). It will be interesting to know if there is any biological significance for this interdomain shift.

## **WD40 proteins function as a protein-protein interaction adaptor in many different cellular processes**

Based on the solved WD40 domain complex structures and mutagenesis-coupled binding assays, we know that WD40 domain can interact with different partners using all sides of its surface, such as the top side, the bottom side and the circumference, which has been discussed in an excellent review recently (Stirnemann et al., 2010). In the following, we will discuss how WD40 domain recruits its substrates, ligand/partner in different modes.

### **1. WD40 proteins function as an interchangeable substrate receptor to target different substrates selectively**

SCF ubiquitin ligase is a multi-component complex consisting of SKP1, CUL1, Rbx1 and a F-box containing protein (Michel and Xiong, 1998). SCF ubiquitin ligase mediates the ubiquitination of proteins involved in cell cycle progression, signal transduction and transcription (Cardozo and Pagano, 2004). In the SCF complex, CUL1 serves as a rigid scaffold that orientates the SKP1-F-box protein and Rbx1 subunits. The functional specificity of the SCF complex depends on the F-box protein as a substrate recognition component. In the human genome, there are about 70 F-box containing proteins, which function as an interchangeable substrate receptor in the SCF ubiquitin ligase system to recruit different substrates specifically for ubiquitination and degradation by proteasome (Fig. 3A) (Jin et al., 2004). The F-box protein usually contains two domains, i.e., the N-terminal F-box domain and the C-terminal substrate binding domain, such as WD40, LRR (leucine-rich repeats, and others) (Jin et al., 2004). They use the F-box domain to interact with SKP1 and the other domain as the substrate receptor. Of the human 70 F-box proteins, about 20 of them contain the WD40 domain, which usually targets phosphorylated targets, such as Cyclin E (Hao et al., 2007), Catenin (Wu et al., 2003) and Sic1 (Nash et al., 2001; Orlicky et al., 2003).

CUL4-DDB1-ROC1 complex is another E3 ubiquitin ligase (Groisman et al., 2003; He et al., 2006; Higa et al., 2006). It shares a conserved molecular architecture with other SCF ubiquitin ligases (Fig. 3B) (Angers et al., 2006). About 50 single  $\beta$ -propeller WD40

proteins have been identified to interact with the three WD40 domain protein DDB1. These single  $\beta$ -propeller WD40 proteins recruit different targets for ubiquitination. For instance, the CUL4-DDB1 ubiquitin ligase targets the replication-licensing factor CDT1 for ubiquitination in response to UV irradiation (Hu et al., 2004). It also ubiquitinates histone H2A, histone H3 and histone H4 at sites of UV-induced DNA damage (Guerrero-Santoro et al., 2008; Wang et al., 2006). The ubiquitination of histones facilitates their removal from the nucleosome and promote subsequent DNA repair. The CUL4-DDB1 complex can also interact with some WD40 proteins involved in transcription repression and chromatin modification, such as TLE1-3, WDR5, and EED. WDR5 and EED are the core components of the MLL and PRC2 histone methylation complexes that are essential for histone H3 methylation and epigenetic control at K4 and K27, respectively. Therefore, the CUL4A-DDB1 ubiquitin ligase regulates histone methylation through ubiquitin-dependent proteolysis (Higa et al., 2006), although the detailed mechanism is still unclear. Structural analysis indicates that DDB1 recruits some of the substrate receptors through a conserved H-box (Bergametti et al., 2002; Li et al., 2010).

## **2. Same WD40 proteins recruit different substrates in a similar mode**

Clathrin is a major component of the polyhedral vesicle coat, which is a “triskelion”-shaped molecule composed of 3 heavy chains and 3 light chains, forms cage-like enclosures and drives vesiculation of cellular membranes (Kirchhausen and Harrison, 1981; Ungewickell and Branton, 1981). Two different adapter protein complexes link the clathrin lattice either to the plasma membrane or to the trans-Golgi network.

The N-terminal domain of clathrin is a WD40  $\beta$ -propeller, which directs incorporation of cargo into coated pits. The comparison of the complex structures of the WD40 domain of clathrin respectively with peptides derived from two different cargo adaptors,  $\beta$ -arrestin 2 and the  $\beta$ -subunit of the AP-3 complex, shows a very similar binding mode, i.e., a peptide-in-groove binding mode (Fig. 3C) (ter Haar et al., 2000). Some other endocytosis related proteins also have the similar "clathrin-box" motif, and the authors predicted that they all likely bind the clathrin WD40 domain in the same binding mode (ter Haar et al., 2000).

WDR5 is a component of the MLL subfamily of histone H3K4 methyltransferases. It was shown that WDR5 is required for the di- to trimethylation conversion of histone H3K4 by presenting the histone H3K4 for further methylation by the MLL complex (Ruthenburg et al., 2006; Schuetz et al., 2006; Wysocka et al., 2005). Later on, it was found that a conserved arginine containing motif, namely the "Win" motif (WDR5 interaction motif) from the MLL is essential for the assembly and enzymatic activity of the MLL1 core complex (Patel et al., 2008). Superposition of WDR5 with these arginine containing peptides, including H3K4 (Couture et al., 2006; Han et al., 2006; Ruthenburg et al., 2006; Schuetz et al., 2006), Win motif (Patel et al., 2008; Song and Kingston, 2008) and a N-terminal fragment from WDR5 (Schuetz et al., 2006) shows that WDR5 uses the same binding surface (top surface) to bind these different peptides (Fig. 3D). The conserved arginine residue in the peptide inserts into the central channel of the  $\beta$ -propeller from the top surface and forms a large number of interactions with WDR5 through a network of hydrogen bonds and cation- $\pi$  interactions and contributes largely to the complex formation (Schuetz et al., 2006).

### **3. Same WD40 proteins recruit different substrates in distinct modes**

On the other hand, some WD40 proteins bind their ligands in distinct modes, such as the  $G_{\beta}$  subunit of the heterotrimeric G proteins (Fig. 3E) (Gaudet et al., 1996; Johnston et al., 2008; Lambright et al., 1996) and the transcriptional repressor TLE (Fig. 3F) (Jennings et al., 2006). Heterotrimeric G proteins transmit information in various transmembrane signaling systems and thereby regulate a variety of cellular functions. Heterotrimeric G proteins are activated by G-protein-coupled receptors (GPCRs), which catalyze the exchange of GTP for GDP on  $G_{\alpha}$  to initiate signal transmission. The comparison of the complex structures of heterotrimeric G proteins (Lambright et al., 1996) and  $G_{\beta\gamma}$  subunit with its regulator phosducin (Gaudet et al., 1996) shows that  $G_{\beta}$  protein binds  $G_{\alpha}$ ,  $G_{\gamma}$ , or phosducin in different modes.  $G_{\beta}$  binds  $G_{\alpha}$  through two separate interfaces, i.e., the "switch interface" at the top of the WD40 domain of  $G_{\beta}$  with a  $\alpha$ -helix from  $G_{\alpha}$  and the "N-terminal interface" between the N-terminal  $\alpha$ -helix of  $G_{\alpha}$  and the side of the WD40 domain of  $G_{\beta}$  (Fig. 3E) (Lambright et al., 1996).  $G_{\beta}$  binds phosducin also through two

separate interfaces. The “switch interface” is almost conserved between these two complexes, but the phosducin binds a side “N-terminal interface” adjacent to the sixth and seventh blades of  $G_{\beta}$ , instead of the first and seventh blades for  $G_{\alpha}$  (Fig. 3E) (Gaudet et al., 1996). A critical role of the  $G_{\beta\gamma}$  dimer in heterotrimeric G proteins is to facilitate the engagement and activation of the  $G_{\alpha}$  subunit by G-protein-coupled receptors. Crystal structures of the  $G_{\beta\gamma}$  dimer in complex with a GPCR peptide from the parathyroid hormone receptor-1 reveals that PTH1R contact only  $G_{\beta}$  residues on the side in a distinct mode from phosducin or  $G_{\alpha}$  (Johnston et al., 2008).

TLE proteins are a family of transcriptional corepressors that bind to a number of transcription factors and inhibit NF-kappa-B-regulated gene expression. In addition, they inhibit the transcriptional activation mediated by CTNNB1 and TCF family members in Wnt signaling. TLE proteins bind these transcriptional repressors by recognizing divergent short peptides via different binding modes (Fig. 3F). The C-terminal WRPW/Y motif (Hairy/Hes/Runx) binds the top surface of the WD40 domain of TLE1 in an extended conformation, whereas the internal EH1 motif binds the top surface of the WD40 domain of TLE1 as a  $\alpha$ -helix (Fig 3F) (Jennings et al., 2006).

#### **4. Utilization of insertion motifs of WD40 domains in ligand binding**

Bub3 is a protein involved in the regulation of the spindle assembly checkpoint. Bub3 and its related proteins (such as Mad3, Bub1) from the spindle assembly checkpoint system inhibit the action of the anaphase-promoting complex (APC), preventing early anaphase entry and mitotic exit. This serves as a mechanism for the fidelity of chromosomal segregation (Lens et al., 2003). Both Bub1 and Mad3 bind constitutively to Bub3 throughout the cell cycle, implying that this interaction is important for the function of both proteins (Hardwick et al., 2000). Crystal structures of Bub3 with a conserved motif (GLEBS motif) from either Bub1 or Mad3 shows that Bub3 uses the top surface of its WD40 domain to recognize the GLEBS motif (Larsen et al., 2007) (Fig. 4A). Interestingly, 2 insertion elements from the Bub3 WD40 domain are involved in binding the GLEBS motif (Larsen et al., 2007). The extended long loop (DA loop) between the D strand of blade 5 to the A strand of blade 6 becomes ordered upon binding to the

GLEBS motif. This DA loop forms a three-stranded  $\beta$ -sheet with two strands from the Bub3 DA loop and one strand from the GLEBS motif. Another intra-blade loop in the blade 7 (BC loop) is also involved in GLEBS motif binding through hydrophobic interactions (Fig 4A).

RBBP4 and RBBP7 are a pair of WD40 domain containing proteins with 92% of sequence identity, which often exist together in different chromatin modifying complexes, such as the Polycomb Repressor Complex 2, a histone H3K37 methyltransferase involved in Hox genes silencing, X-inactivation, germline development, stem cell pluripotency and differentiation, and cancer metastasis (Cao and Zhang, 2004); Sin3/HDAC complex, a corepressor complex involved in a variety of cellular processes, including cell cycle progression, genomic stability, embryonic development, and homeostasis; the NuRD complex (the nucleosome remodeling and histone deacetylase complex), an chromatin remodeling and modification complex with a broad cellular and tissue distribution, and harbors histone deacetylation and chromatin remodeling ATPase activities in the same complex (Feng and Zhang, 2003). Very interestingly, only RBBP7 is present in the HAT complex with HAT1, the only known type B histone acetyltransferase in charge of acetylating newly synthesized histones during the process of chromatin assembly (Verreault et al., 1996). RBBP7 in the HAT complex stimulates the acetyltransferase activity of Hat1 (Verreault et al., 1998). On the contrary, only RBBP4 is identified in the CAF1 complex, a complex mediating chromatin assembly in DNA replication and DNA repair. Hence, it seems that RBBP4 and RBBP7 may have different functions, although they have high sequence and structure similarity. On the other hand, it is also possible that RBBP4 and RBBP7 perform the same function in these chromatin complexes, RBBP4 and RBBP7 is mutually exclusive in these complexes. That means that in each individual complex, either RBBP4 or RBBP7 is present and sufficient for the complex function.

It was first reported by Stillman's lab that both RBBP4 and RBBP7 bind directly to the first  $\alpha$ -helix of histone H4, a region that is not accessible when H4 is in chromatin (Verreault et al., 1998). The crystal structures of RBBP4 and RBBP7 have been determined recently, either alone or in complex with the first helix of histone H4 or a N-

terminal fragment from the GATA-1 cofactor FOG-1 (Lejon et al., 2011; Murzina et al., 2008; Song et al., 2008). The complex structures of RBBP7 with the histone H4 peptide show that the first  $\alpha$ -helix of histone H4 is bound in a groove formed between an N-terminal  $\alpha$ -helix and the extended CD loop in blade six (Fig. 4B). Surprisingly, the first  $\alpha$ -helix of histone H4 adopts a different conformation in this complex than it does in either the nucleosome or in the complex with ASF1 (English et al., 2006), another histone chaperone. That implicates that the first  $\alpha$ -helix of histone H4 unfolds to interact with RBBP4 or RBBP7 (Murzina et al., 2008).

The NuRD complex has been shown to associate with some transcription factors, such as the GATA-1 cofactor FOG-1. FOG-1-mediated recruitment of NuRD is essential for cell lineage re-enforcement during haematopoiesis (Gao et al., 2010). The crystal structure of the NuRD subunit RBBP4 in complex with a N-terminal fragment from the GATA-1 cofactor FOG- reveals that this FOG-1 peptides binds to the top surface of the WD40 domain of RBBP4. This is another example that the WD40 domain could utilize all sides of its surface for partner interaction, especially in those multiple component complexes.

## **5. Utilization of the inter-blade binding grooves of WD40 domain in ligand binding**

The yeast polarity protein Sro7 is a WD40 domain containing protein, whose primary role is in membrane fusion. Sro7 consists of two seven-bladed WD40  $\beta$ -propellers followed by a C-terminal 60-residue-long tail (Hattendorf et al., 2007). The C-terminal tail is bound to the bottom surface of the first WD40 propeller (Figure 5A and 5B) and is thought to have an auto-inhibitory effect on Sro7 binding to Sec9 (Hattendorf et al., 2007). This C-terminal tail contains two short  $\alpha$ -helices, which are wedged into two inter-blade binding grooves separately, i.e., the first  $\alpha$ -helix bound in the groove formed between the third and fourth blades, and the second  $\alpha$ -helix in the groove formed

between the second and third blades (Figure 5A and 5B). In addition, the first blade of the second WD40  $\beta$ -propeller harbors an insertion motif, which contains a  $\alpha$ -helix bound in the groove formed between the first and second blades. The three  $\alpha$ -helices all interact with the WD40 domains mainly through hydrophobic interactions. Both the C-terminal tail and the insertion motif of Sro7 are involved in stabilizing the WD40 fold and tying up the two WD40 domains together (Hattendorf et al., 2007). Interestingly, neither of two WD40 propellers in Sro7 have the “velcro” closure.

This bind mode is also observed in the PALB2-BRCA2 peptide complex (Oliver et al., 2009). PALB2 plays a critical role in homologous recombination repair through its ability to recruit BRCA2 and RAD51 to DNA breaks and serves as the molecular scaffold in the formation of the PALB2-BRCA1-BRCA2 complex (Sy et al., 2009). In the PALB2-BRCA2 (aa 21-39) peptide complex, the BRCA2 peptide forms a short  $\alpha$ -helix, which is bound in a pocket formed between the fourth and fifth blades of the PALB2 WD40  $\beta$ -propeller (Figure 5C). Similar to the Sro7 structure, the short  $\alpha$ -helix interacts with PALB2 mainly through extensive hydrophobic interactions with PALB2 (Oliver et al., 2009) (Figure 5D).

## **Recognition of post-translational modification marks by WD40 domains**

SCF ubiquitin ligase targets substrates for ubiquitination and degradation, which are often modified by phosphorylation or other post-translational modifications (called degron) (Ang and Wade Harper, 2005). The modification of these target substrates introduces high affinity binding to the SCF ubiquitin ligase and facilitates the ubiquitination. The

crystal structures of Skp1-Cdc4-CPD(Orlicky et al., 2003),  $\beta$ -TrCP1-Skp1- $\beta$ -Catenin(Wu et al., 2003), and Fbw7-Skp1-Cyclin E(Hao et al., 2007) complexes shed light on how the serine/threonine phosphorylation contributes to their binding to the SCF ligase. In the crystal structure of the Skp1-Fbw7 complex bound to a Cyclin E peptide, the pT380 side chain inserts into a shallow, positively charged pocket at the inner rim of the  $\beta$ -propeller channel and interacts with residues from the third and fourth blades (Figure 6A) (Hao et al., 2007). The pT380 phosphate group makes a few hydrogen bonds with residues from the top face of the Fbw7 near the central channel. The pS384 side chain binds outside the central channel on top of the second and third blades, which also interacts with the residues from the top surface of the Fbw7 WD40 domain via a hydrogen bond network. These two phosphorylated residues are the major contributors for the Cycle E peptide to bind to the SCF substrate targeting subunit Fbw7 in this case.

The PRC2 complex (the polycomb repressive complex 2) is a histone methyltransferase targeting H3K27 and H1K26 (Cao et al., 2002; Cao and Zhang, 2004; Czermin et al., 2002; Muller et al., 2002). The H3K27 methylation mark recruits the PRC1 complex through its chromodomain protein Polycomb, which in turn maintains transcriptional silencing through an unknown mechanism (Min et al., 2003; Wang et al., 2004). In addition to Hox gene silencing, the Polycomb Group complexes are also involved in X-inactivation, germline development, stem cell pluripotency and differentiation, and cancer metastasis (Cao and Zhang, 2004). Histone lysine methylation plays an important role in chromatin organization, transcriptional regulation and other nuclear processes. It was recently shown that PRC2 maintains histone methylation patterns during DNA replication in part through its ability to bind to H3K27me3 (Hansen et al., 2008). The structure of the WD40 domain of EED in complex with a EZH2 N-terminal fragment revealed that the top surface of the EED WD40 domain harbors an aromatic cage consisting of Phe97, Tyr148, and Tyr365, a common features for methyllysine binding proteins, such as chromodomain (Min et al., 2003), MBT repeat domain (Eryilmaz et al., 2009; Guo et al., 2009; Min et al., 2007), tudor domain (Adams-Cioaba et al., 2010; Adams-Cioaba and Min, 2009), PHD domain (Adams-Cioaba and Min, 2009) and the methylarginine binding tudor domain (Liu et al., 2010a; Liu et al., 2010b). Recent crystal

structures of EED with the histone H3K27me3 peptide reveal that the histone H3K27me3 peptide is bound to EED through the trimethyllysine residue, which is accommodated by a four-aromatic residue cage (Figure 6B) (Margueron et al., 2009; Xu et al., 2010). Very interestingly, the extended aliphatic chain of the methyllysine lies against the planar ring of W364, which is a unique methyllysine-binding feature for EED (Xu et al., 2010). EED is also able to bind other ARKS motif containing histone peptides, such as H3K9me3 and H1K26me3 with similar binding affinity and other trimethylated peptides with lower affinity. On the basis of our binding and structure studies, we concluded that the trimethylated lysine is the major affinity contributor, and the residues one and two residues preceding the methyllysine in a peptide confers the binding specificity (Xu et al., 2010). More importantly, the binding of different histone marks by EED differentially regulates the activity and specificity of PRC2. The H3K27me3 mark enhances the histone methyltransferase activity of PRC2, whereas the H1K26me3 mark inhibits the PRC2 methyltransferase activity (Xu et al., 2010). Furthermore, the H1K26me3 binding switches the specificity of PRC2 from methylating H3K27 to EED (Xu et al., 2010).

Based on the bioinformatics analysis (Margueron et al., 2009) and our structure-based sequence analysis (Xu et al., 2010), EED seems to be the only methyllysine binder with an aromatic cage. However, a recent report suggested that WDR61 is a histone methyllysine binder (Chan et al., 2009). Nevertheless, the crystal structure shows that this WD40 does not harbor an aromatic cage. Therefore, if WDR61 does bind, it will have different a binding mechanism, which warrant further investigation in the future. WDR5 has also been claimed to be a histone H3K4me2 binder (Wysocka et al., 2005), but the complex structures of WDR5 and the histone H3K4 peptides show that the histone peptide binds the top surface with the arginine H3R2 in histone H3 inserted into the central channel from the top surface of the WD40  $\beta$ -propeller (Schuetz et al., 2006). This H3R2 forms a large number of interactions with WDR5 through a network of hydrogen bonds and cation- $\pi$  interactions (Figure 6C). The H3R2 contributes to most of the binding affinity, which is confirmed by the mutant peptide binding studies (Schuetz et al., 2006). Consistently, the H3R2 methylation severely diminishes histone H3 binding to WDR5

(Couture et al., 2006). Therefore, EED is the only known methyllysine binding WD40 protein (Margueron et al., 2009; Xu et al., 2010).

## **Concluding remarks**

The WD40 domain is one of the most abundant domains in eukaryotic genomes. It is also among the top interacting domains in eukaryotic genomes. The WD40 domain proteins function as an adaptor in many different protein complexes or protein-DNA complexes in very diverse cellular processes. Since the first WD40 domain structure was solved in 1994, significant progress has been made to advance our understanding about the WD40 domain itself and also how it functions. Considering the rich interaction surfaces afforded by the WD40 domain and its insertion and extension motifs, it is clear that there is much more to be investigated about the WD40 domain protein complexes in the future.

## **ACKNOWLEDGMENTS**

This research was supported by the Structural Genomics Consortium, a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Karolinska Institutet, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck & Co., Inc., the Novartis Research Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research and the Wellcome Trust. We apologize to authors whose studies could not be cited here due to space limitations. We also would like to thank the members in the Min lab and Structural Genomics Consortium at the University of Toronto for their contributions and discussions to this work. The coordinates and structure factors for WDR39, WDR61, WDR92, RBBP4 and DDB1 have been deposited at the Protein Data Bank under the accession codes 3FM0, 3OW8, 3I2N, 33GFC and 3E0C, respectively.

## References

- Adams-Cioaba, M.A., Guo, Y., Bian, C., Amaya, M.F., Lam, R., Wasney, G.A., Vedadi, M., Xu, C., and Min, J. (2010). Structural Studies of the Tandem Tudor Domains of Fragile X Mental Retardation Related Proteins FXR1 and FXR2. *PLoS One* 5, e13559.
- Adams-Cioaba, M.A., and Min, J. (2009). Structure and function of histone methylation binding proteins. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 87, 93-105.
- Andrade, M.A., Perez-Iratxeta, C., and Ponting, C.P. (2001). Protein repeats: structures, functions, and evolution. *J Struct Biol* 134, 117-131.
- Ang, X.L., and Wade Harper, J. (2005). SCF-mediated protein degradation and cell cycle control. *Oncogene* 24, 2860-2870.
- Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T., and Zheng, N. (2006). Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* 443, 590-593.
- Bergametti, F., Bianchi, J., and Transy, C. (2002). Interaction of hepatitis B virus X protein with damaged DNA-binding protein p127: structural analysis and identification of antagonists. *J Biomed Sci* 9, 706-715.
- Brohawn, S.G., Partridge, J.R., Whittle, J.R., and Schwartz, T.U. (2009). The nuclear pore complex has entered the atomic age. *Structure* 17, 1156-1168.
- 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 (New York, NY)* 298, 1039-1043.
- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current opinion in genetics & development* 14, 155-164.
- Cardozo, T., and Pagano, M. (2004). The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev Mol Cell Biol* 5, 739-751.
- Chan, D.W., Wang, Y., Wu, M., Wong, J., Qin, J., and Zhao, Y. (2009). Unbiased proteomic screen for binding proteins to modified lysines on histone H3. *Proteomics* 9, 2343-2354.
- Chen, X., Zhang, Y., Douglas, L., and Zhou, P. (2001). UV-damaged DNA-binding proteins are targets of CUL-4A-mediated ubiquitination and degradation. *The Journal of biological chemistry* 276, 48175-48182.
- Couture, J.F., Collazo, E., and Trievel, R.C. (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. *Nature structural & molecular biology* 13, 698-703.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111, 185-196.
- English, C.M., Adkins, M.W., Carson, J.J., Churchill, M.E., and Tyler, J.K. (2006). Structural basis for the histone chaperone activity of Asf1. *Cell* 127, 495-508.

Eryilmaz, J., Pan, P., Amaya, M.F., Allali-Hassani, A., Dong, A., Adams-Cioaba, M.A., Mackenzie, F., Vedadi, M., and Min, J. (2009). Structural studies of a four-MBT repeat protein MBTD1. *PLoS One* 4, e7274.

Feng, Q., and Zhang, Y. (2003). The NuRD complex: linking histone modification to nucleosome remodeling. *Curr Top Microbiol Immunol* 274, 269-290.

Fong, H.K., Hurley, J.B., Hopkins, R.S., Miake-Lye, R., Johnson, M.S., Doolittle, R.F., and Simon, M.I. (1986). Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs. *Proc Natl Acad Sci U S A* 83, 2162-2166.

Gao, Z., Huang, Z., Olivey, H.E., Gurbuxani, S., Crispino, J.D., and Svensson, E.C. (2010). FOG-1-mediated recruitment of NuRD is required for cell lineage re-enforcement during haematopoiesis. *The EMBO journal* 29, 457-468.

Gaudet, R., Bohm, A., and Sigler, P.B. (1996). Crystal structure at 2.4 angstroms resolution of the complex of transducin betagamma and its regulator, phosducin. *Cell* 87, 577-588.

Groisman, R., Polanowska, J., Kuraoka, I., Sawada, J., Saijo, M., Drapkin, R., Kisselev, A.F., Tanaka, K., and Nakatani, Y. (2003). The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 113, 357-367.

Guerrero-Santoro, J., Kapetanaki, M.G., Hsieh, C.L., Gorbachinsky, I., Levine, A.S., and Raptic-Otrin, V. (2008). The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. *Cancer Res* 68, 5014-5022.

Guo, Y., Nady, N., Qi, C., Allali-Hassani, A., Zhu, H., Pan, P., Adams-Cioaba, M.A., Amaya, M.F., Dong, A., Vedadi, M., *et al.* (2009). Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. *Nucleic acids research* 37, 2204-2210.

Han, Z., Guo, L., Wang, H., Shen, Y., Deng, X.W., and Chai, J. (2006). Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell* 22, 137-144.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nature cell biology* 10, 1291-1300.

Hao, B., Oehlmann, S., Sowa, M.E., Harper, J.W., and Pavletich, N.P. (2007). Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell* 26, 131-143.

Hardwick, K.G., Johnston, R.C., Smith, D.L., and Murray, A.W. (2000). MAD3 encodes a novel component of the spindle checkpoint which interacts with Bub3p, Cdc20p, and Mad2p. *J Cell Biol* 148, 871-882.

Hattendorf, D.A., Andreeva, A., Gangar, A., Brennwald, P.J., and Weis, W.I. (2007). Structure of the yeast polarity protein Sro7 reveals a SNARE regulatory mechanism. *Nature* 446, 567-571.

He, Y.J., McCall, C.M., Hu, J., Zeng, Y., and Xiong, Y. (2006). DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. *Genes & development* 20, 2949-2954.

Higa, L.A., Wu, M., Ye, T., Kobayashi, R., Sun, H., and Zhang, H. (2006). CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nature cell biology* 8, 1277-1283.

Hu, J., McCall, C.M., Ohta, T., and Xiong, Y. (2004). Targeted ubiquitination of CDT1 by the DDB1-CUL4A-ROC1 ligase in response to DNA damage. *Nature cell biology* 6, 1003-1009.

Jennings, B.H., Pickles, L.M., Wainwright, S.M., Roe, S.M., Pearl, L.H., and Ish-Horowicz, D. (2006). Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol Cell* 22, 645-655.

Jin, J., Cardozo, T., Lovering, R.C., Elledge, S.J., Pagano, M., and Harper, J.W. (2004). Systematic analysis and nomenclature of mammalian F-box proteins. *Genes & development* 18, 2573-2580.

Johnston, C.A., Kimple, A.J., Giguere, P.M., and Siderovski, D.P. (2008). Structure of the parathyroid hormone receptor C terminus bound to the G-protein dimer Gbeta1gamma2. *Structure* 16, 1086-1094.

Kirchhausen, T., and Harrison, S.C. (1981). Protein organization in clathrin trimers. *Cell* 23, 755-761.

Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., and Sigler, P.B. (1996). The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* 379, 311-319.

Larsen, N.A., Al-Bassam, J., Wei, R.R., and Harrison, S.C. (2007). Structural analysis of Bub3 interactions in the mitotic spindle checkpoint. *Proc Natl Acad Sci U S A* 104, 1201-1206.

Lejon, S., Thong, S.Y., Murthy, A., Alqarni, S., Murzina, N.V., Blobel, G.A., Laue, E.D., and Mackay, J.P. (2011). Insights into Association of the NuRD Complex with FOG-1 from the Crystal Structure of an RbAp48{middle dot}FOG-1 Complex. *The Journal of biological chemistry* 286, 1196-1203.

Lens, S.M., Wolthuis, R.M., Klompmaker, R., Kauw, J., Agami, R., Brummelkamp, T., Kops, G., and Medema, R.H. (2003). Survivin is required for a sustained spindle checkpoint arrest in response to lack of tension. *The EMBO journal* 22, 2934-2947.

Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic acids research* 37, D229-232.

Li, D., and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell Mol Life Sci* 58, 2085-2097.

Li, T., Chen, X., Garbutt, K.C., Zhou, P., and Zheng, N. (2006). Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell* 124, 105-117.

Li, T., Robert, E.I., van Breugel, P.C., Strubin, M., and Zheng, N. (2010). A promiscuous alpha-helical motif anchors viral hijackers and substrate receptors to the CUL4-DDB1 ubiquitin ligase machinery. *Nature structural & molecular biology* 17, 105-111.

Liu, H., Wang, J.Y., Huang, Y., Li, Z., Gong, W., Lehmann, R., and Xu, R.M. (2010a). Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes & development*.

Liu, K., Chen, C., Guo, Y., Lam, R., Bian, C., Xu, C., Zhao, D.Y., Jin, J., Mackenzie, F., Pawson, T., *et al.* (2010b). Structural basis for recognition of arginine methylated

Piwi proteins by the extended Tudor domain. *Proc Natl Acad Sci U S A* 107, 18398-18403.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., *et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762-767.

Michel, J.J., and Xiong, Y. (1998). Human CUL-1, but not other cullin family members, selectively interacts with SKP1 to form a complex with SKP2 and cyclin A. *Cell Growth Differ* 9, 435-449.

Min, J., Allali-Hassani, A., Nady, N., Qi, C., Ouyang, H., Liu, Y., MacKenzie, F., Vedadi, M., and Arrowsmith, C.H. (2007). L3MBTL1 recognition of mono- and dimethylated histones. *Nature structural & molecular biology* 14, 1229-1230.

Min, J., Zhang, Y., and Xu, R.M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & development* 17, 1823-1828.

Mohri, K., Vorobiev, S., Fedorov, A.A., Almo, S.C., and Ono, S. (2004). Identification of functional residues on *Caenorhabditis elegans* actin-interacting protein 1 (UNC-78) for disassembly of actin depolymerizing factor/cofilin-bound actin filaments. *The Journal of biological chemistry* 279, 31697-31707.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* 111, 197-208.

Murzin, A.G. (1992). Structural principles for the propeller assembly of beta-sheets: the preference for seven-fold symmetry. *Proteins* 14, 191-201.

Murzina, N.V., Pei, X.Y., Zhang, W., Sparkes, M., Vicente-Garcia, J., Pratap, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F., *et al.* (2008). Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure* 16, 1077-1085.

Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F.B., Mendenhall, M.D., Sicheri, F., Pawson, T., and Tyers, M. (2001). Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 414, 514-521.

Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371, 297-300.

Oliver, A.W., Swift, S., Lord, C.J., Ashworth, A., and Pearl, L.H. (2009). Structural basis for recruitment of BRCA2 by PALB2. *EMBO Rep* 10, 990-996.

Orlicky, S., Tang, X., Willems, A., Tyers, M., and Sicheri, F. (2003). Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* 112, 243-256.

Paoli, M. (2001). Protein folds propelled by diversity. *Prog Biophys Mol Biol* 76, 103-130.

Patel, A., Dharmarajan, V., and Cosgrove, M.S. (2008). Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *The Journal of biological chemistry* 283, 32158-32161.

Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J., and Verdine, G.L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nature structural & molecular biology* 13, 704-712.

Schuetz, A., Allali-Hassani, A., Martin, F., Loppnau, P., Vedadi, M., Bochkarev, A., Plotnikov, A.N., Arrowsmith, C.H., and Min, J. (2006). Structural basis for molecular recognition and presentation of histone H3 by WDR5. *The EMBO journal* 25, 4245-4252.

Schuster-Bockler, B., Schultz, J., and Rahmann, S. (2004). HMM Logos for visualization of protein families. *BMC Bioinformatics* 5, 7.

Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends in biochemical sciences* 24, 181-185.

Song, J.J., Garlick, J.D., and Kingston, R.E. (2008). Structural basis of histone H4 recognition by p55. *Genes & development* 22, 1313-1318.

Song, J.J., and Kingston, R.E. (2008). WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *The Journal of biological chemistry* 283, 35258-35264.

Stirnemann, C.U., Petsalaki, E., Russell, R.B., and Muller, C.W. (2010). WD40 proteins propel cellular networks. *Trends in biochemical sciences* 35, 565-574.

Sy, S.M., Huen, M.S., and Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc Natl Acad Sci U S A* 106, 7155-7160.

ter Haar, E., Harrison, S.C., and Kirchhausen, T. (2000). Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin. *Proc Natl Acad Sci U S A* 97, 1096-1100.

Ungewickell, E., and Branton, D. (1981). Assembly units of clathrin coats. *Nature* 289, 420-422.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* 87, 95-104.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1998). Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr Biol* 8, 96-108.

Voegtli, W.C., Madrona, A.Y., and Wilson, D.K. (2003). The structure of Aip1p, a WD repeat protein that regulates Cofilin-mediated actin depolymerization. *The Journal of biological chemistry* 278, 34373-34379.

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.

Wang, H., Zhai, L., Xu, J., Joo, H.Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y., and Zhang, Y. (2006). Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol Cell* 22, 383-394.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14, 637-646.

Whittle, J.R., and Schwartz, T.U. (2010). Structure of the Sec13-Sec16 edge element, a template for assembly of the COPII vesicle coat. *J Cell Biol* 190, 347-361.

Wittschieben, B.O., Iwai, S., and Wood, R.D. (2005). DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer,

mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *The Journal of biological chemistry* 280, 39982-39989.

Wu, G., Xu, G., Schulman, B.A., Jeffrey, P.D., Harper, J.W., and Pavletich, N.P. (2003). Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. *Mol Cell* 11, 1445-1456.

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.

Xu, C., Bian, C., Yang, W., Galka, M., Ouyang, H., Chen, C., Qiu, W., Liu, H., Jones, A.E., Mackenzie, F., *et al.* (2010). Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci U S A* 107, 19266-19271.

## Figure legend

**Figure 1. Definition of WD40 repeat and WD40 domains.** (A) HMM logo of WD40 repeat from the website <http://pfam.sanger.ac.uk/family?id=WD40#tabview=tab3> (Schuster-Bockler et al., 2004). The letter size represents the conservation level of each amino acid in the WD40 repeat. (B) Structure of a canonical WD40 repeat protein. Each repeat consists of four  $\beta$ -strands. The residues marked by stars are conserved hydrophobic residues involved in WD40  $\beta$ -propeller stabilization. (C) WD40 domain is stabilized mainly through hydrophobic interactions between the blades. (D). Structure comparison of the repeats in the same WD40 protein or between different WD40 proteins. The PDB accession codes for WDR5, WDR39, WDR61, WDR92 and G $\beta$  are 2H9M, 3FM0, 3OW8, 3I2N and 1GP2, respectively.

**Figure 2. Architecture of WD40 domain proteins shown in cartoon representation.** (A) WDR5 (2H9M); (B) Sif2p (1R5M); (C) Sec13 in complex with Sec31 (2PM6). The SEC13 is colored in green and Sec31 is colored in gray with the insertion repeat colored in salmon; (D) Aip1 (1NR0). The two seven-bladed  $\beta$ -propellers are colored in blue and green, respectively; (E) DDB1 (3EI3). The three seven-bladed  $\beta$ -propellers are colored in red, blue and green, respectively; (F) Superposition of DDB1 from three DDB1 structures. The first two  $\beta$ -propeller of DDB1 can be well superimposed, but the third  $\beta$ -propeller exhibited significant shift. The shifted  $\beta$ -propeller is colored in red (2HYE), green (3EI2) and blue (2B5L), respectively.

**Figure 3. WD40 domain proteins exhibit different binding modes.** (A) The different F-box proteins in the SCF ubiquitin ligase recruit different substrates through the WD40 domain. (B) The different DCAF WD40 proteins recognize different substrates for Cul4-DDB1 ubiquitin ligase. (C) Clathrin (purple) in complex with two different peptides (blue  $\beta$ -arrestin 2/AP-3 1C9I and yellow 1C9L); (D) WDR5 (blue) in complex with MLL (red, 3EMH) and unmodified H3K4 (yellow, 2H9M); (E) TLE1 (green) in

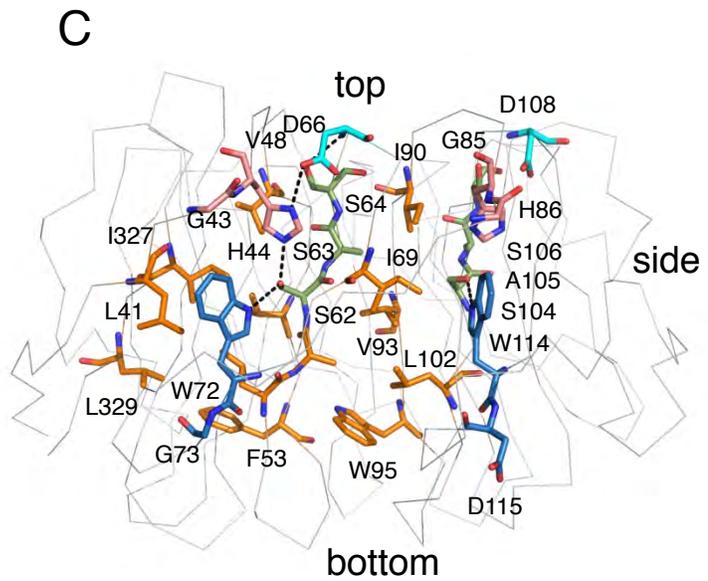
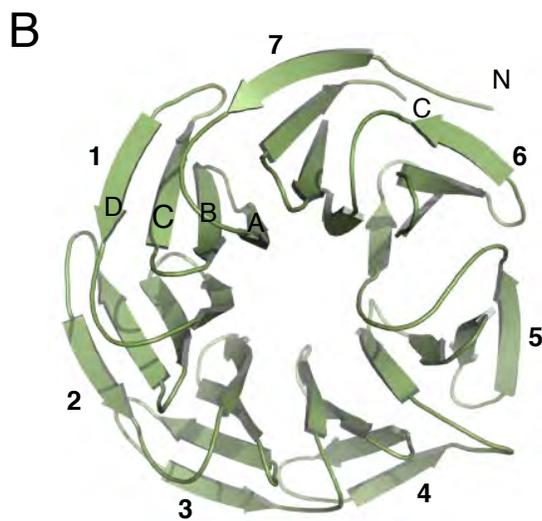
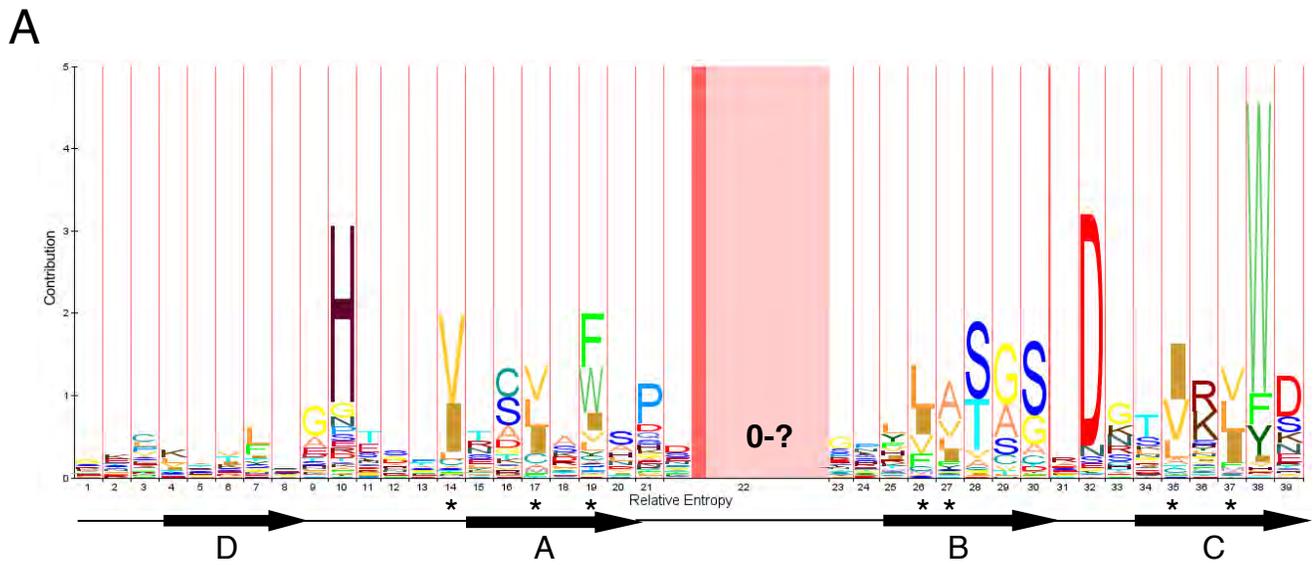
complex with eh1 (red, 2CE8) and WRPW (yellow, 2CE9); (F) G protein beta (red)-gamma (blue) heterodimer in complex with alpha (cyan, PDB code: 1GP2) and phosphatidylinositol 3-kinase (yellow, PDB code: 1A0R).

**Figure 4. Utilization of insertions and extensions inside WD40 repeat domains in ligand binding.** The WD40 protein is colored in gray except the insertion or extension (A) Bub3 (green) in complex with Mad3 (salmon, 2I3T); (B) RBBP7 (blue) in complex with H4 (yellow, 3CFS) and RBBP4 (blue) in complex with Fog-1 (salmon, 2XU7).

**Figure 5. Utilization of the inter-blade binding grooves of WD40 domain in ligand binding.** (A) Crystal structure of yeast Sro7 protein. The C-terminal tail of Sro7 is colored in yellow, and the insertion motif from the second WD40 domain of Sro7 is colored in salmon. (B) The first WD40 domain of Sro7 is displayed in surface representation and the two  $\alpha$ -helices from the C-terminal tail of Sro7 reside in two hydrophobic pockets formed between the second and third blades, and the third and fourth blades, respectively. (C) Crystal structure of the C-terminal WD40 domain of PALB2 and the BRCA2 peptide (aa 21-39). The BRCA2 peptide is colored in yellow. The short  $\alpha$ -helix from the BRCA2 peptide is bound in a pocket formed between the fourth and fifth blades of the PALB2 WD40  $\beta$ -propeller.

**Figure 6. Recognition of post-translational modification marks by WD40 domains.**

(A) FBW7 (gray) in complex with cyclin E degron (yellow). The residues involved in binding with phosphorylated peptide are shown in cyan sticks. (B) WDR5 (gray) in complex with dimethylated-H3K4 (red). Residues involved in interaction with H3R2 are shown in green sticks. (C) EED (gray) in complex with trimethylated-H3K27 (yellow). Residues involved in directly interaction with the peptide are shown in blue sticks.



**D**

RMSD/ Identity	WDR5	WDR39	WDR61	WDR92	G $\beta$	Overall
WDR5	0.7-1Å/23-44%	0.3-0.6Å/11-38%	0.5-0.9Å/22-38%	0.4-4.6Å/5-16%	0.6-1.0Å/13-25%	0/100%
WDR39		0.3-0.8Å/23-47%	0.4-1.2Å/7-25%	0.3-2.9Å/2-13%	0.5-4.8Å/11-19%	1.0Å/18%
WDR61			0.3-0.8Å/8-20%	0.4-2.3Å/4-29%	0.4-0.7Å/11-17%	1.3Å/23%
WDR92				0.8-3.2Å/6-9%	0.9-2.6Å/5-17%	2.4Å/10%
G $\beta$					0.6-1.0Å/6-22%	1.4Å/24%

**Fig. 1**

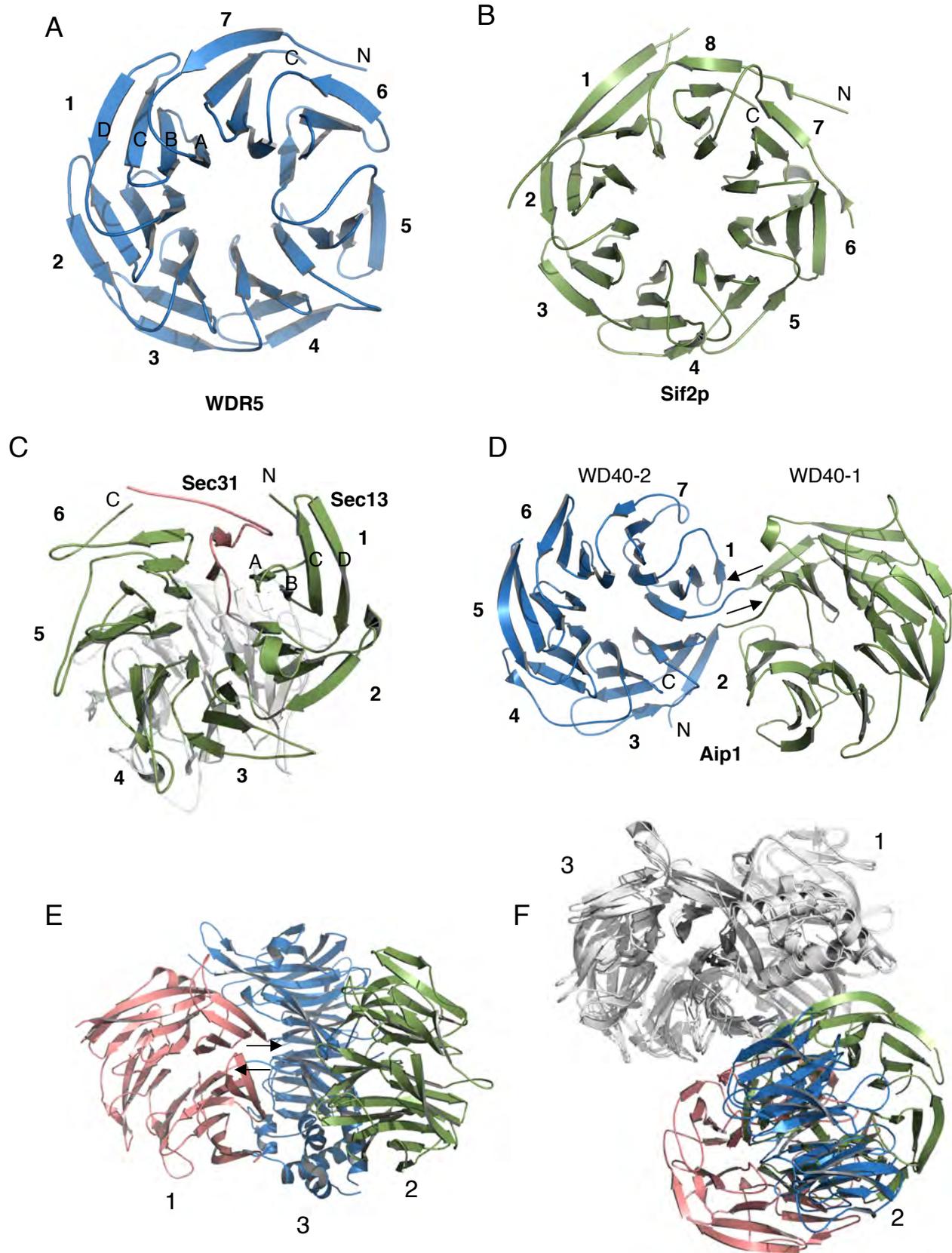


Fig. 2

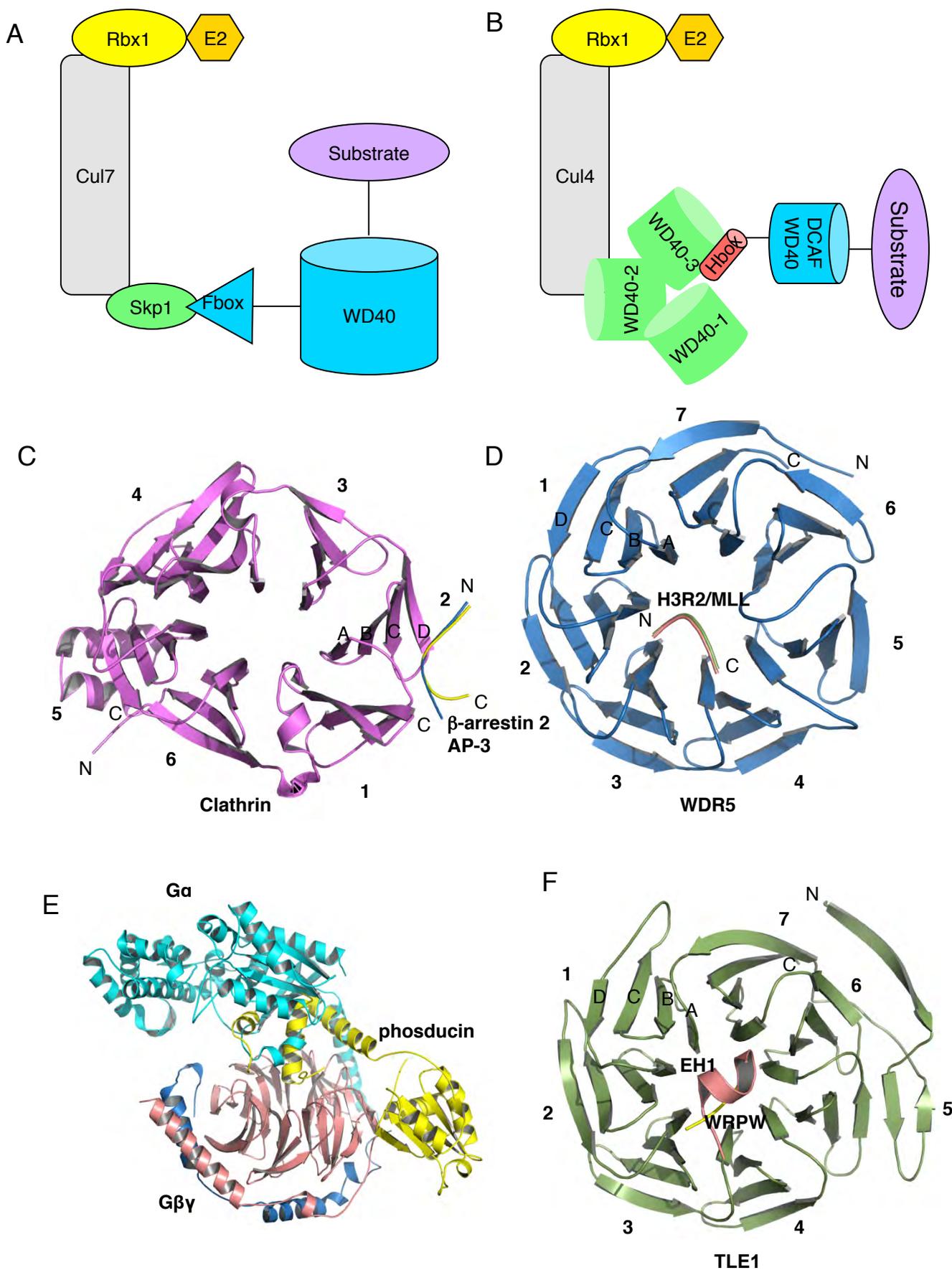


Fig 3

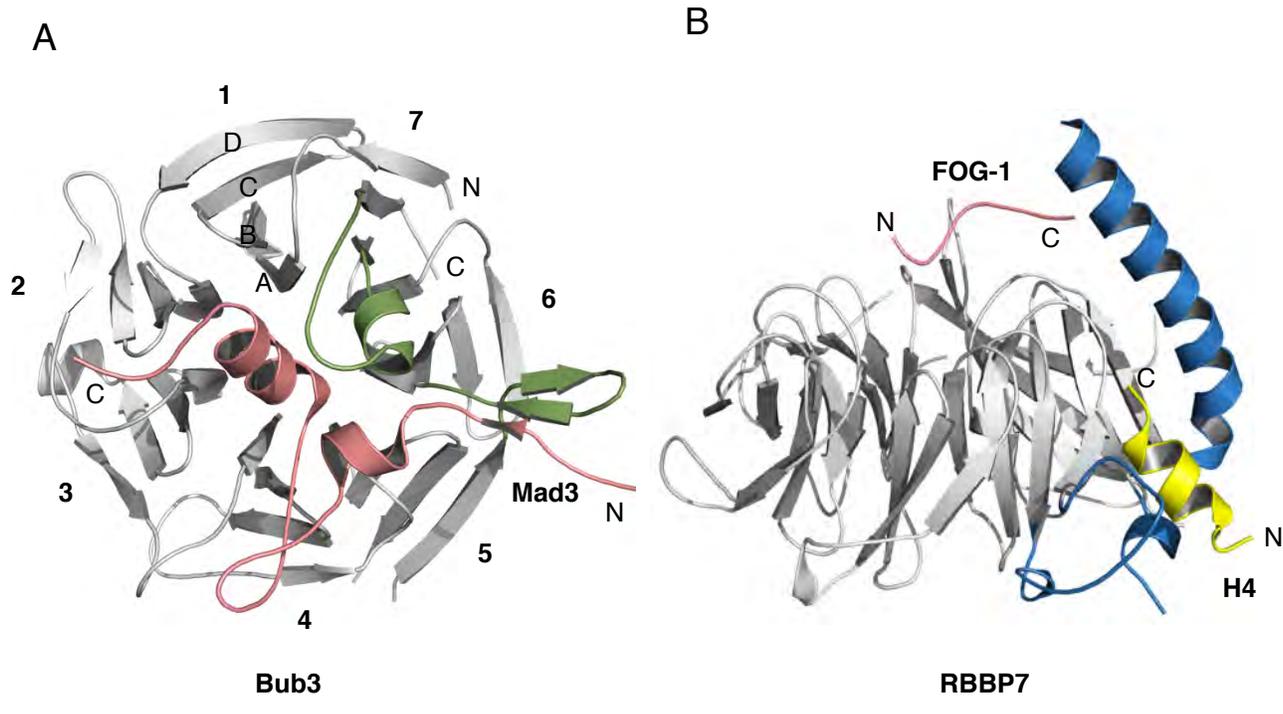


Fig. 4

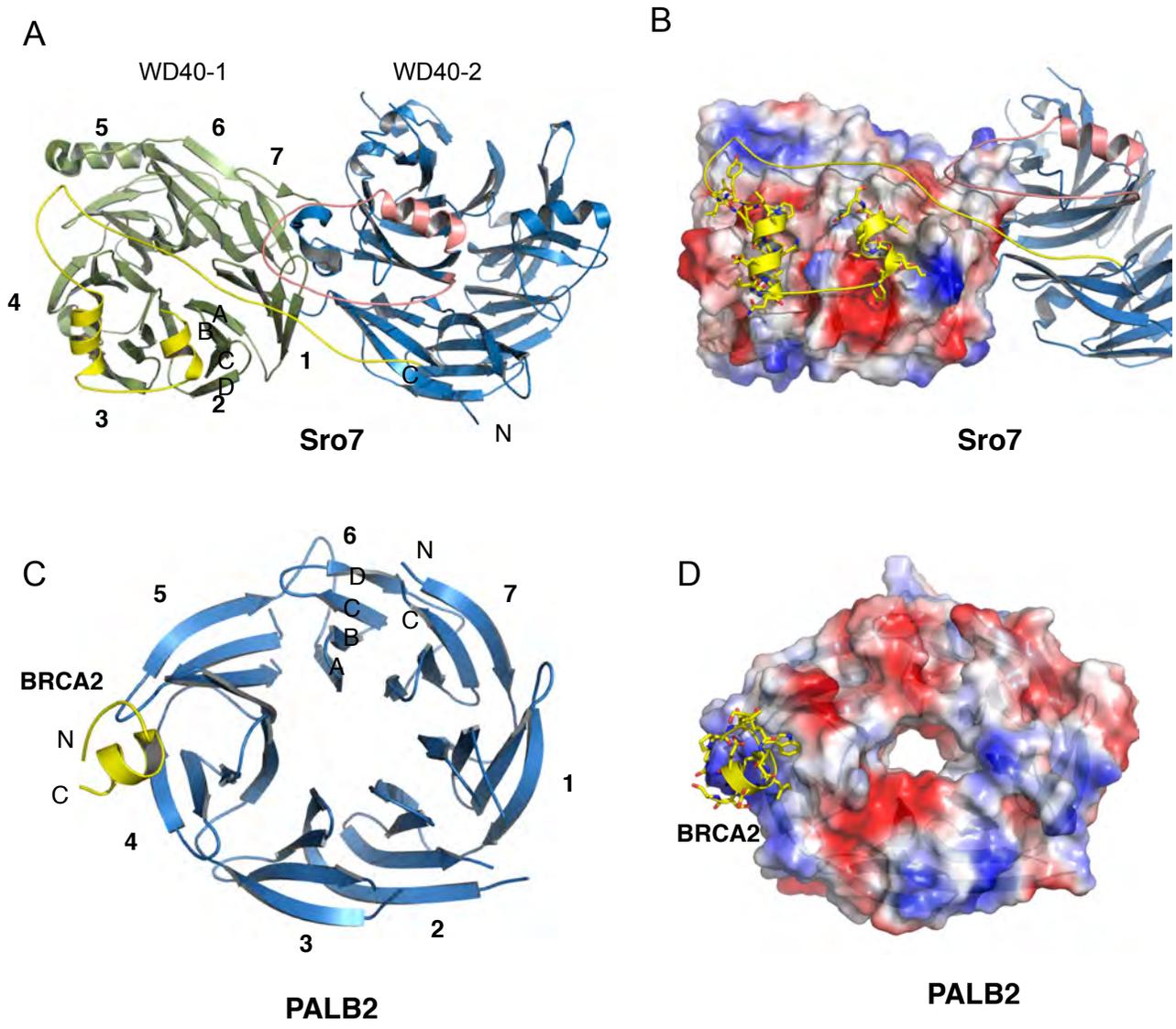


Fig 5

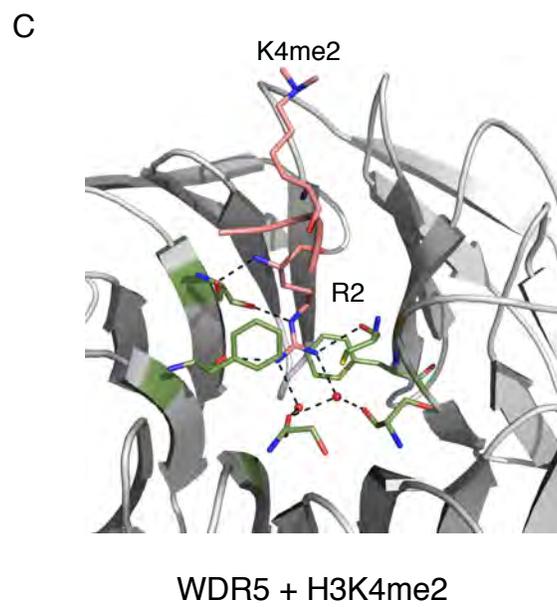
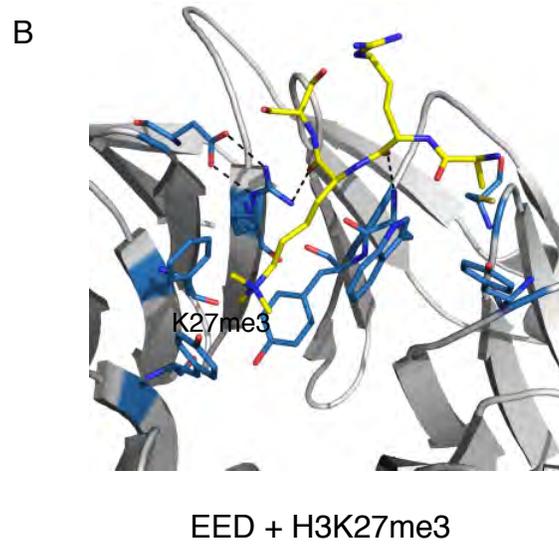
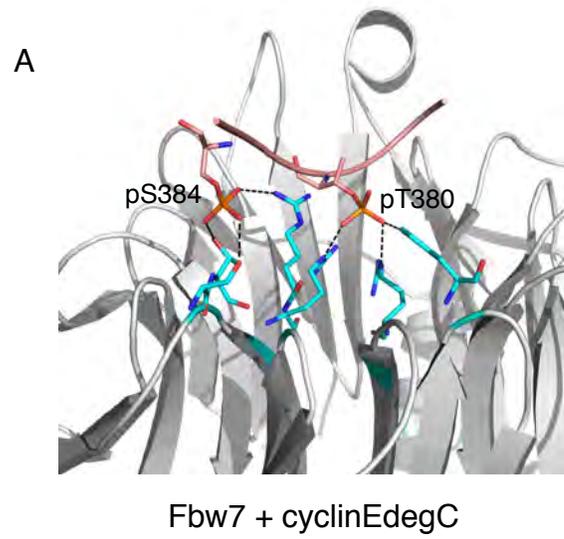


Fig. 6