

MLL1 minimal catalytic complex is a dynamic conformational ensemble susceptible to pharmacological allosteric disruption

Lilia Kaustov^{1*}, Alexander Lemak^{1*}, Hong Wu^{2*}, Marco Faini³, Scott Houliston¹, Lixin Fan⁴, Xianyang Fang⁴, Hong Zeng², Shili Duan¹, Abdellah Allali-Hassani², Masoud Vedadi^{2,5}, Ruedi Aebersold^{3,6}, Yunxing Wang⁴, Cheryl H. Arrowsmith^{1,2#}

¹Princess Margaret Cancer Centre and Department of Medical Biophysics, University of Toronto, ON M5G 2M9, Canada.

²Structural Genomics Consortium, University of Toronto, 101 College Street, MaRS Centre, South Tower, Toronto, ON M5G 1L7, Canada.

³Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland

⁴The Small-Angle X-ray Scattering Core Facility, Center for Cancer Research of National Cancer Institute, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc. Frederick, MD 21702, USA

⁵Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, M5S 1A8, Canada

⁶Faculty of Science, University of Zürich, Zürich, Switzerland

* These authors contributed equally

Corresponding author:

#Cheryl H. Arrowsmith: carrow@uhnresearch.ca

ABSTRACT

Histone H3K4 methylation is an epigenetic mark associated with actively transcribed genes. This modification is catalyzed by the mixed lineage leukaemia (MLL) family of histone methyltransferases including MLL1, MLL2, MLL3, MLL4, SET1A and SET1B. Catalytic activity of MLL proteins is dependent on interactions with additional conserved proteins but the structural basis for subunit assembly and the mechanism of regulation is not well understood. We used a hybrid methods approach to study the assembly and biochemical function of the minimally active MLL1 complex (MLL1, WDR5 and RbBP5). A combination of small angle X-ray scattering (SAXS), cross-linking mass spectrometry (XL-MS), NMR spectroscopy, and computational modeling were used to generate a dynamic ensemble model in which subunits are assembled via multiple weak interaction sites. We identified a new interaction site between the MLL1 SET domain and the WD40 repeat domain of RbBP5, and demonstrate the susceptibility of the catalytic function of the complex to disruption of individual interaction sites.

INTRODUCTION

Post-translational modifications on histone tails are key epigenetic signals for regulation of chromatin structure and gene expression. Histone H3 lysine 4 (H3K4) methylation is the epigenetic mark exclusively associated with transcriptionally active chromatin (1, 2). This modification is mostly catalyzed by the MLL/SET1 family of histone methyltransferases (3, 4), through their evolutionarily conserved SET domain (5, 6). The founding member of this family of H3K4 methyltransferases is the yeast SET1 protein (7, 8). In mammals, methylation of H3K4 is carried out by a family of six proteins: MLL (mixed lineage leukemia protein)1 to MLL4, SET1A and SET1B (9–15). The MLL proteins play crucial roles in embryonic development and hematopoiesis through transcriptional regulation of the clustered homeobox (*Hox*) genes and other genes important for developmental regulation (10, 16–19). Deletion of MLL1 and MLL2 can lead to severe defects in embryonic development in mice (18, 20). The *MLL1* gene is frequently rearranged in human acute leukemia in both adults and children (21–23). MLL3 and MLL4 have also been linked to other human malignancies. Recently studies have identified inactivating mutations in MLL3 and MLL4 in different types of human tumors (24–27), as well as in Kabuki syndrome (28).

The catalytic activity of MLL family members are dependent to varying degrees on the presence of additional conserved protein subunits, RbBP5, WDR5 and ASH2L, and a minimal core enzyme can be reconstituted with the conserved C-terminal SET domain of MLLs and at least two of the other subunits. Interestingly, in studies of these reconstituted core enzymes, MLL1 appears to be unique among the family members in its requirements for, and interactions with other subunits. For example, compared to other family members, the catalytic activity of MLL1 is the most dependent on WDR5 (29–31). Similarly, MLL1 binds with the least affinity to the RbBP5-Ash2L heterodimer, and its catalytic activity is only weakly stimulated by RbBP5-ASH2L compared to WDR5 (32).

Recent crystallographic studies of MLL3 support a model in which the RbBP5-ASH2L heterodimer stabilizes the catalytically active conformations of MLL2,3,4 through interactions with conserved surfaces on their SET domain (32). However, it was suggested that two key variant residues on this surface of MLL1 dramatically weakened the

interaction between MLL1 and RbBP5-ASH2L relative to that of other MLL members, thereby increasing the dependence of MLL1 on the WDR5 subunit. The unique dependence of MLL1 activity on WDR5 may be of therapeutic relevance, as we and others have shown that pharmacological targeting of the MLL interaction site on WDR5 can functionally antagonize MLL1 in cancers that are dependent on MLL1 activity (33–35). While there are several structures of WDR5 bound to MLL and RbBP5 peptides (31, 36–39), as well as a crystal structure of the apo-SET domain (40) of MLL1 and a 24 Å resolution cryo-EM model of the homologous yeast COMPASS complex (41), an atomic level picture of a functional MLL1 catalytic core complex is still lacking. Here, we report a hybrid methods study of MLL1 and its catalytic core components in solution. Using small angle X-ray scattering (SAXS), cross-linking mass spectrometry (XL-MS), NMR spectroscopy, and computational modeling we derived a dynamic ensemble model for MLL1/WDR5/RbBP5 and identify a new interaction site between MLL1-SET domain and the N-terminal WD40 repeat domain of RbBP5. Our data support the notion that the functional MLL1 enzyme comprises a collection of weak but specific interactions, and that the disruption of individual interactions can have significant destabilizing effects on the entire complex. These results highlight the dynamic nature of an important protein complex and the strategy of targeting a weak but druggable protein-protein interaction site to antagonize the function of a larger macromolecular assembly that is dependent on a collection of weak interactions.

MATERIALS AND METHODS

Cloning of MLL1, WDR5 and RbBP5 constructs

The coding regions for human MLL1-SET (aa3785-3969) and MLL1-WIN (aa3745-3969) were PCR-amplified and sub-cloned into the pET28GST-LIC vector (GeneBank ID: EF456739). The mutants, MLL1-SET-7D (Δ 3786-3793) and MLL1-SET_Q3787V/P3788L/-Y3791G, were generated from the wild-type clone using the QuikChange PCR mutagenesis kit (Agilent). RbBP5 constructs of different lengths (aa 10-340, 320-410, 340-538, 1-538) and WDR5 (aa 24-334) were sub-cloned into the pET28-MHL vector (GeneBank ID: EF456738). For expression of dimeric complexes and reconstitution of the trimeric MLL1

complex, full-length WDR5 and MLL1-WIN, WDR5 and RbBP5 were cloned into pFastBac Dual expression vector (Thermo Fisher Scientific), respectively, with MLL1-WIN and RbBP5 tagged with an N-terminal hexa His-tag.

Protein preparation

The individual components of the MLL1 complex used in this study were expressed in *E. coli* and purified using an N-terminal GST-tag (for MLL1) or His-tag (for WDR5 and RbBP5). The dimeric and trimeric complexes of MLL1 used for SAXS and cross-linking studies were expressed in Sf9 cells. The dimeric complex of WDR5-MLL1-WIN and WDR5-RbBP5 were purified using TALON affinity resin (Clontech) followed by gel filtration chromatography. The purified dimeric complexes were incubated on ice for 2 hours together to reconstitute the trimeric complex which was purified and recovered by gel filtration chromatography. Detailed procedures are described in the Supplementary Data section.

SAXS data collection, analysis and modelling

SAXS measurements were carried out at the beamline 12-ID-C of the Advanced Photon Source, Argonne National Laboratory. The energy of the X-ray beam was 18 KeV (wavelength $\lambda=0.6888 \text{ \AA}$), and two setups (small- and wide- angle X-ray scattering, SAXS and WAXS) were used in which the sample to charge-coupled device (CCD) detector (MAR research, Hamburg) distance were adjusted to achieve scattering q values of $0.006 < q < 2.3 \text{ \AA}^{-1}$, where $q = (4\pi/\lambda)\sin\theta$, and 2θ is the scattering angle. Data were analyzed using the program PRIMUS (ATSAS package, EMBL (42)). Detailed descriptions of SAXS data collection and analysis, and modelling protocols, are provided in the Supplementary Data.

Chemical cross-linking mass spectrometry

The reconstituted trimer complex of WDR5, RbBP5 and MLL1-SET was cross-linked at a concentration between 12 and 16 μM , with 1 mM of isotopically coded disuccinimidyl suberate (DSS-d₀, DSS-d₁₂) as described previously (43). Protease digestion was carried out with LysC and trypsin. After acidification, cross-linked peptides were purified on C18 cartridges and enriched by size-exclusion chromatography (SEC). SEC fractions were

analyzed in duplicate on an LC-MS (Easy-nLC 300; Orbitrap LTQ XL). For complete details refer to Supplementary Data.

NMR Spectroscopy

NMR spectra were collected at 25°C on a Bruker Avance(II) 800 MHz spectrometer equipped with a cryoprobe. Samples contained 5% D₂O with protein concentrations ranging from 100 to 350 μM. For the assignment of the backbone resonances of the ΔN-WDR5 construct, a triply-labeled (¹⁵N/¹³C/²H) sample was prepared and conventional triple-resonance backbone spectra were acquired as described previously (44), in order assign backbone and Cβ resonances using the ABACUS approach (45). [¹H-¹⁵N]-TROSY titrations of ¹⁵N-labeled WDR5 was performed by adding aliquots of MLL1-WIN (GSARAEVHLRKS) and RbBP5-WBM (EDEEVDVTSV) peptides at molar ratios ranging from 1:1 to 1:7. The weighted chemical shift displacements were calculated using the following formula: $\Delta\text{ppm}=[\delta_{\text{NH}}]^2+(\delta_{\text{N}}/5)^2]^{1/2}$. Spectra were processed with NMR Pipe (46) and analyzed with SPARKY (47).

GST Pull-down experiments

Recombinant purified MLL1-GST proteins were incubated with RbBP5 constructs in an assay buffer containing 20mM TRIS pH 7.7, 150 mM NaCl, 10μM ZnCl₂, 5mM β-mercaptoethanol, 5 mM DTT, 1 mM PMSF in a 1:2 molar ratio at 4°C for 1h. Proteins were then incubated with 100μl of glutathione-Sepharose beads (GE Healthcare) for an additional 1h. The mixture was transferred to a micro-column and was extensively washed with assay buffer. Bound proteins were eluted with 30mM reduced glutathione and detected by SDS-PAGE and Coomassie staining.

Histone methyltransferase assay

Activity assays were performed in 50 mM Tris-HCl, pH 8.0, 5 mM DTT and 0.01% Triton X-100, using 5 μM ³H-SAM and 5 μM Biotin-H3 1-25. Increasing concentrations of RbBP5 were added to 200 nM of MLL1-WDR5 complexes (with either wild-type or mutant MLL1). All reactions were incubated for 90 minutes at room temperature and the SPA method was

used to determine activities. Experiments were performed in triplicate. To test the effect of OICR-9429 on the MLL1 complex, increasing concentrations of the compound was incubated with 200 nM MLL1-WDR5 complex for 20 min before adding 400 nM RbBP5. The activity of the complex was measured as above.

RESULTS AND DISCUSSION

SAXS data reveal solution ensembles for WDR5, RbBP5 and MLL-SET

To model catalytically active MLL1 complexes, we first collected reference solution data for the individual subunits including the SET domain of MLL1, the WD40 repeat region of WDR5 (Δ N-WDR5), the N-terminus of RbBP5 (RbBP5-NTD) and full-length RbBP5, followed by characterization of dimeric and trimeric complexes. **Fig 1A** shows the protein constructs used in this study. Normalized Kratky plots of Δ N-WDR5 and RbBP5-NTD exhibit a typical bell-shape expected for a globular protein and are nearly superimposable in the q range $0 < qR_g < 3$ (**Fig 1B**). Also, the experimental values of R_g predicted for Δ N-WDR5 and RbBP5-NTD are in agreement with the theoretical values expected for globular proteins (**Table 1 and Fig S2**). The normalized Kratky plot of MLL1-SET also exhibits a bell-shape, but its maximum is shifted with respect to the globular protein position, with poor convergence at high q -values, indicating that MLL1-SET is flexible. The observed flexibility of the MLL1-SET could be attributed to known inherent dynamics of the SET domain in the absence of cofactor (32), and to the disordered N-terminal tail. The calculated solution ensembles for each protein taking into account known or predicted disordered regions (see SI for details) establish good correspondence between our SAXS measurements and the crystal structures of WDR5, the SET domain of MLL1, and our homology model of RbBP5-NTD predicted using ROSETTA (48)(**Fig S1**).

One of the main challenges in modeling the MLL1 complex is the lack of structural information on RbBP5. To better understand its structural arrangement, we collected [^1H - ^{15}N]-TROSY NMR spectra of a full-length construct, as well as constructs corresponding to its C-terminal and N-terminal regions (**Fig 2A**). The data confirm that RbBP5-NTD is a globular, folded domain, consistent with our SAXS analysis and WD40 homology model. The C-terminal region of RbBP5 (RbBP5-CTD) is substantially disordered as evidenced by

the lack of spectral dispersion (**Fig 2A**). Both the gel filtration profile (**Fig S3B**) and the radius of gyration estimated from SAXS data (**Table 1 and Fig S2**) indicate a high degree of disorder in the RbBP5 full-length protein. This is further supported by sequence based secondary structure prediction and order parameters, which predict a rigid globular N-terminus and a flexible coil-like C-terminus (**Fig S3**). Interestingly, the $[^1\text{H}-^{15}\text{N}]$ -TROSY spectrum of full-length RbBP5 is not the superposition of the individual NTD and CTD spectra, and reflects features of both folded and unfolded regions with some apparent conformational broadening, possibly reflecting weak intramolecular interactions (**Fig S3A**).

The NMR data is consistent with the general shape of the normalized Rg-based Kratky plot and the pair distance distribution function $P(r)$ for full-length RbBP5 (**Fig 1B, C**). In particular, the $P(r)$ function of RbBP5 has an asymmetric shape with a long smooth tail at large r values, and the position of its maximum is shifted only slightly ($\sim 4 \text{ \AA}$) with respect to that of RbBP5-NTD. The latter features indicate that full-length RbBP5 has no additional globular content compared to RbBP5-NTD.

Based on the above data we used the Sparse Ensemble Selection (SES) approach (49) to calculate a solution ensemble of RbBP5 that would satisfy the SAXS data. An initial ensemble consisting of 20,000 models of RbBP5 with random conformations of its flexible regions (residues 1-23 and 326-538) did not fit the SAXS data well (the goodness-of-fit $\chi=9.4$). We next generated a solution ensemble that better fits the SAXS data by calculating an optimal weight for each model in the initial ensemble using a multi-orthogonal matching pursuit algorithm (49) (see SI for details). The resulting optimal ensemble fits the SAXS data very well with $\chi=0.38$ (**Figs 2C, S1**). The most populated models in the optimal ensemble are shown in **Figs 2B, S1F, G**. The optimal ensemble displays a much narrower distribution of radius of gyration values than the initial random ensemble, with a major peak at $R_g=37 \text{ \AA}$ (**Fig 2C**). This indicates that RbBP5 adopts a more compact conformation than would be predicted for a fully random CTD, consistent with our NMR data for the full-length protein.

Binary subcomplexes have dynamic non-random solution conformations mediated by WD40 repeat domains

Our SAXS data for the binary complexes of WDR5/MLL1-WIN and WDR5/RbBP5 both suggest the presence of significant disorder, especially for WDR5/MLL1-WIN (**Fig 3A, S3A**). The $P(r)$ functions of WDR5/MLL1-WIN and WDR5/RbBP5 are typical for proteins containing globular domains tethered by long disordered regions (**Fig 3A**). The position of the $P(r)$ major peak for the aforementioned complexes is close to the positions of the major peaks of $P(r)$ of their individual components (**Fig 1C**), indicating that in both complexes the globular domains are not in close contact and may not adopt a unique arrangement in solution. WDR5 is known to interact with RbBP5 and MLL1 through small peptide segments designated as the WDR5 binding motif (WBM) (38) and WDR5 interacting (WIN) (36) motif, respectively (**Fig 1A**). Both interactions with WDR5 have reported dissociation constants on the order of 1-2 μM (30, 36, 38, 39). To calculate solution ensembles of the WDR5 binary complexes, we first used NMR to verify that WDR5's mode of interaction with these two motifs, as observed in the crystal structures, is maintained in solution. We expressed a triply-labeled ($^{15}\text{N}/^{13}\text{C}/^2\text{H}$) ΔN -WDR5 construct, and were able to assign 254 amides (**Fig S6**). We then used chemical shift perturbation (CSP) analysis in $[^1\text{H}-^{15}\text{N}]$ -TROSY titration experiments, to localize the WRD5 binding site for peptides corresponding to the two motifs. As seen in **Fig 3B**, there is excellent agreement between the WDR5 CSP profiles and the WDR5/WIN (PDB:4ESG) and WDR5/RbBP5 (PDB:2XL2) crystal structures. Next, using these two structures to fix each WDR5-peptide interface, we modeled the ensemble of solution conformations for the binary complexes of WDR5 with MLL1-WIN and full-length RbBP5 using the SES method. The arrangement of the globular domains in the most populated models of the optimal ensembles for both WDR5/MLL1-WIN and WDR5/RbBP5 complexes does not support the existence of additional interactions of WDR5 with MLL1-WIN or with RbBP5 other than those described above (**Fig S4**).

There is currently no atomic resolution structural data for the interaction of MLL1 with RbBP5. Recently, the activation segment (AS) of RbBP5 was shown to bind to the SET domain of other MLL family members, but only very weakly to MLL1 (32). In order to determine whether there is a direct interaction between RbBP5 and MLL1, we performed GST pull-down studies of full-length RbBP5, RbBP5-NTD and RbBP5 (320-410) with GST-MLL1-SET and GST-MLL1-WIN (**Fig 3C-E**). Both MLL1 constructs interacted with RbBP5 constructs containing the N-terminal WD40 repeat but did not interact with the C-terminal

residues (aa 320-410) of RbBP5 containing the AS region. These results agree with the lack of conservation of the RbBP5 AS-binding surface on MLL1 (32), and suggest that MLL1-SET may interact with the WD40 domain of RbBP5.

SAXS and cross-linking data suggest a dynamic triangulated ensemble for WDR5/RbBP5/MLL1-WIN

Our SAXS data for the catalytically active WDR5/RbBP5/MLL1-WIN complex showed a substantial amount of flexibility. The shape of the experimental Kratky plots of the complex is typical of partially disordered proteins (**Figs 4A, S5A**). In particular, the Rg-based Kratky plot is a bell-shaped curve with a maximum at (2.26, 1.27) shifted to higher values of the coordinates with respect to its position expected for a globular protein. Also, the presence of a high degree of flexibility is evidenced by the poor convergence of the Kratky plots at high q values. The low maximum value of 0.48 in the Vc-based Kratky plot (**Fig S5A**), as well as the asymmetric shape of the $P(r)$ function (**Fig S5B**), suggests an elongated shape of the complex. This is in agreement with the averaged *ab initio* SAXS-derived molecular envelope, which showed an extended shape with approximate dimensions of 220×105×70 Å (**Fig 4D**).

We note that pair-distance distribution functions of proteins containing several globular domains connected by long disordered regions are characterized by peaks at low r -values, corresponding to the intra-domain distances. Therefore, if the three globular domains of WDR5, MLL1-SET and RbBP5-NTD are not interacting directly with each other within the trimeric MLL1 complex, we would expect the $P(r)$ function of the complex to have peaks at 26-32 Å, reflecting the inter-atomic distances prevailing within these domains (**Figs 1C, S5B**). However, the experimental $P(r)$ function of the trimeric MLL1 complex has its maximum at a much larger distance of ~ 47 Å (**Fig S5B**), suggesting the existence of possible inter-domain contacts in the complex.

To aid our modeling of the solution conformations of the trimeric complex we performed cross-linking mass spectrometry studies. We observed many intramolecular cross-links within each of the three proteins. These were highly consistent with the available WDR5 and MLL1-SET crystal structures, and importantly, in agreement with our

RbBP5-NTD homology model, establish that these structural models are reliable representations of the domains within the trimeric complex in solution. We also observed a number of intermolecular cross-links, with the largest number being between MLL1 and RbBP5, suggesting association of these two subunits in solution. **Fig 4B** shows the sequence mapping of both intra- and inter- molecular DSS cross-links observed for the trimeric complex. For the purposes of modeling we used only intermolecular cross-links between lysine residues within the globular subunits (**Table S1**). These six cross-links are shown on **Fig 4B** by solid blue lines.

Using both SAXS and cross-linking data as conformational restraints we utilized the SES approach to calculate solution ensembles of the WDR5/RbBP5/MLL1-WIN complex that would satisfy both sets of experimental data. An initial pool of representative structures was generated by combining rigid-body modeling and molecular dynamics simulations for both all-atomic and coarse-grained models of the complex. It was assumed that the MLL1-SET and RbBP5-NTD domains were tethered to WDR5 by WIN and WBS motifs, respectively, as seen in crystal structures (30, 36–38), along with cross-link derived distance restraints (see SI for details). It should be noted here that individual members of this initial ensemble did not necessary satisfy all inter-molecular cross-links: each satisfies on average 3 to 4 cross-links.

The optimal SES ensemble of WDR5/RbBP5/MLL1-WIN fits the SAXS data as a whole, with $C_{SAXS} = 0.23$ in the q -range $0 < q < 0.23$. While only SAXS data were used to select the optimal ensemble, each inter-domain cross-link is consistent with at least one member of the ensemble, so that the ensemble as a whole is consistent with all cross-links. The most highly populated models of the optimal SES ensemble are shown in **Figs 4 and S5**. The optimal ensemble contains two clear populations; a smaller population (~13%) with a radius of gyration centered at ~37 Å and a more predominant population (~85%) with a more widely distributed radius of gyration centered at ~ 52 Å (**Fig 4C, D**). The relative position of the MLL1-SET and RbBP5-NTD domains is well defined by four inter-domain cross-links and their association is similar in all models within the optimal ensemble. In contrast, the relative position of WDR5 (compared to RbBP5-NTD and MLL1-SET) varies within the ensemble due to flexible linkers between the globular RbBP5-NTD and MLL-SET

domains and their WDR5 interacting sequences (WBS and WIN, respectively) (**Fig 4E, S5C**). There are only two inter-domain cross-links that involve WDR5, and they can only be simultaneously satisfied in the more compact subpopulation corresponding to $R_g \sim 37 \text{ \AA}$. The ensemble distribution of C α -C α distances corresponding to these cross-links showed a large fraction of ensemble members for which the cross-links cannot be simultaneously formed (**Fig 4F**).

RbBP5-NTD has a unique interaction mode with MLL1

A recent crystallographic study revealed an important role for the AS+ABM region of RbBP5 in binding to the SET domain of MLL family proteins, thereby stimulating the latter's catalytic activity (32). This work showed that the catalytic activity of MLL2, 3, 4 and SET1A/B was highly dependent on the RbBP5_{AS+ABM}/ASH2L_{SPRY} dimer, but not WDR5. In contrast, the catalytic activity of MLL1 SET domain was only weakly stimulated by the RbBP5_{AS+ABM}/ASH2L_{SPRY} dimer and instead, its optimal activity was more dependent on WDR5. Our solution model suggests an explanation for these observations.

Highly populated models of the trimeric complex in the optimal ensemble feature a direct interaction between the WD40 domain of RbBP5 and a short peptide sequence of MLL1 located between the WIN motif and the SET domain (**Fig 4E**). We refer to this RbBP5 binding sequence as the RBS region of MLL1 (**Fig 1A**). The RBS binding surface of RbBP5-NTD consists of a number of hydrophobic residues (V249, I283, L286, V287, and I289), and residues Q273 and P253 (**Fig S5D**). The RBS also has an intramolecular association with the SET domain of MLL1 thereby bridging the RbBP5-NTD and the MLL1-SET domain within the complex. The MLL1-SET domain residues comprising the RBS binding interface (K3825, K3828, N3861, R3871, M3897, H3898, G3899, R3903, and F3904) are shown in **Fig 5A**.

Several distinct features of this putative RbBP5-MLL1 association are notable. First, in the trimeric complex the MLL1-SET domain interacts with RBS and RbBP5-NTD via a surface that significantly overlaps with the previously identified ASH2L/RbBP5 interaction surface of the SET domains of the other MLL family members (32). Further mutagenesis and computational studies (32) had suggested that this interaction surface mediates a stabilization of the flexible SET domains of MLL2,3,4 upon interaction with the

ASH2L/RbBP5 dimer, which in turn, stimulated catalytic activity of the respective SET domains. We define this surface of MLL-SET domains (including MLL1) as the activation surface. A second notable feature of our MLL1 solution model is the interaction of residues N3861 and Q3867 of the SET domain activation surface with RBBP5-NTD. These two residues are unique to MLL1 and were previously shown to be incompatible with functional interactions with the ASH2L/RbBP5 dimer (32). This suggests that the MLL1 activation surface may rely on interactions with RBS and RbBP5-NTD as an alternative activation mechanism. Thirdly, the RBS segment is a unique feature of MLL1 and is not conserved in the other MLL proteins that rely more heavily on the ASH2L/RbBP5 dimer for activation (**Fig 5B**). According to our solution structural ensemble the RBS plays a key role in association of MLL1 with RbBP5-NTD. This is supported by both GST-pull down and gel filtration experiments. Indeed, either deletion or mutation of RBS residues of MLL1 to those of MLL2 within the MLL1-SET construct showed no interaction with RbBP5-NTD by pull down and gel filtration experiments (**Fig 5C**). Finally, deletion of the RBS also decreased the ability of RbBP5 to stimulate the catalytic activity of the WDR5/MLL1-WIN complex (**Fig 5D**).

Taken together, our structural model of the RbBP5/WDR5/MLL1-WIN trimeric complex suggests that the activation mechanism of the complex is mediated in part through the unique but likely weak interaction of RbBP5 with MLL1, which in turn stabilizes the SET-I motif of the catalytic SET domain. WDR5 serves as a hub that brings together the MLL1-SET and the WD40 repeat domain of RbBP5, thereby increasing their effective local concentrations and facilitating what is otherwise a weak ($K_D > 1\mu\text{M}$), but specific interaction within the trimeric complex. Hence, we hypothesized that a triumvirate of weak, but specific intermolecular interactions are required to maintain the integrity of the MLL1 minimal complex and that disruption of an individual interaction site may be sufficient to disrupt the entire functional complex. To test this hypothesis, we measured the ability of OICR-9429, a small molecule antagonist of the WDR5-WIN interaction, to disrupt the association of the RbBP5/WDR5/MLL1-WIN complex using gel filtration (**Fig 6A**). Pharmacological disruption of the WDR5-MLL interaction compromised the assembly of the trimeric complex (**Fig 6A,B**). OICR-9429 also inhibited the catalytic activity of the recombinant trimeric complex (**Fig 6C**). These results are consistent with our previous

work showing that OICR-9429 can disrupt assembly and function of the endogenous MLL1 complex in cells (33), and similar results have been reported for MM-401, a peptide-based antagonist of MLL-WIN interaction (34, 50).

Our results provide the first atomic level description of a functional MLL1 complex. Our model reveals that the minimal active MLL1 complex is comprised of a series of three weak (μM) protein-protein interactions, each of which can be a site of vulnerability for disruption of the functional complex. This has important implications for development of pharmacological antagonists of the MLL1 complex and further strengthens this approach to target other multiprotein complexes that are dependent on weak, but druggable interaction sites.

ACKNOWLEDGEMENTS

For the SAXS experiments, we gratefully acknowledge use of the SAXS Core facility of the Center for Cancer Research, National Cancer Institute (NCI). The SAXS data were collected at beamline 12-ID-B. The shared scattering beamline 12-ID-B resource is allocated under the PUP-24152 agreement between the National Cancer Institute and Argonne National Laboratory (ANL). We thank Dr. Xiaobing Zuo (ANL) for his expert support. Use of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility, was operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada through Ontario Genomics Institute, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no.115766], Janssen, Merck & Co., Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and the Wellcome Trust.

FUNDING

This work was supported by the Natural Sciences and Engineering Research Council of Canada. CHA holds a Canada Research Chair in Structural Genomics. MF was supported by

a Long Term Fellowship from the European Molecular Biology Organization (EMBO). MF and RA are supported by ERC grant AdG-670821 Proteomics 4D.

REFERENCES:

1. Rice, J.C. and Allis, C.D. (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.*, **13**, 263–273.
2. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
3. Ruthenburg, A.J., Allis, C.D. and Wysocka, J. (2007) Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell*, **25**, 15–30.
4. Shilatifard, A. (2008) Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.*, **20**, 341–348.
5. Kouzarides, T. (2002) Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.*, **12**, 198–209.
6. Dillon, S.C., Zhang, X., Trievel, R.C. and Cheng, X. (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.*, **6**, 227.
7. 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.
8. Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F. and Allis, C.D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.*, **15**, 3286–3295.
9. Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M. and Canaani, E. (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell*, **10**, 1119–1128.
10. Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L. and Roeder, R.G. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell*, **121**, 873–885.
11. Hughes, C.M., Rozenblatt-Rosen, O., Milne, T.A., Copeland, T.D., Levine, S.S., Lee, J.C., Hayes, D.N., Shanmugam, K.S., Bhattacharjee, A., Biondi, C.A., *et al.* (2004) Menin associates with a trithorax family histone methyltransferase complex and with the *hoxc8* locus. *Mol. Cell*, **13**, 587–597.
12. Goo, Y.-H., Sohn, Y.C., Kim, D.-H., Kim, S.-W., Kang, M.-J., Jung, D.-J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T., *et al.* (2003) Activating signal cointegrator 2 belongs to a novel

steady-state complex that contains a subset of trithorax group proteins. *Mol. Cell. Biol.*, **23**, 140–149.

13. Wysocka, J., Myers, M.P., Laherty, C.D., Eisenman, R.N. and Herr, W. (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev.*, **17**, 896–911.

14. Lee, J.-H. and Skalnik, D.G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.*, **280**, 41725–41731.

15. Lee, J.-H., Tate, C.M., You, J.-S. and Skalnik, D.G. (2007) Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J. Biol. Chem.*, **282**, 13419–13428.

16. Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G. and Hess, J.L. (2005) MLL associates specifically with a subset of transcriptionally active target genes. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 14765–14770.

17. Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D. and Hess, J.L. (2002) MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol. Cell*, **10**, 1107–1117.

18. Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A. and Korsmeyer, S.J. (1995) Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*, **378**, 505–508.

19. Terranova, R., Agherbi, H., Boned, A., Meresse, S. and Djabali, M. (2006) Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 6629–6634.

20. Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K.R., Aasland, R., Anastassiadis, K., Ang, S.-L. and Stewart, A.F. (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Dev. Camb. Engl.*, **133**, 1423–1432.

21. Ayton, P.M. and Cleary, M.L. (2001) Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene*, **20**, 5695–5707.

22. Hess, J.L. (2004) Mechanisms of transformation by MLL. *Crit. Rev. Eukaryot. Gene Expr.*, **14**, 235–254.

23. Krivtsov, A.V. and Armstrong, S.A. (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nat. Rev. Cancer*, **7**, 823–833.

24. Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., *et al.* (2011) Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*, **476**, 298–303.

25. Dalglish, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., *et al.* (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, **463**, 360–363.
26. Pleasance, E.D., Stephens, P.J., O’Meara, S., McBride, D.J., Meynert, A., Jones, D., Lin, M.-L., Beare, D., Lau, K.W., Greenman, C., *et al.* (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*, **463**, 184–190.
27. Parsons, D.W., Li, M., Zhang, X., Jones, S., Leary, R.J., Lin, J.C.-H., Boca, S.M., Carter, H., Samayoa, J., Bettegowda, C., *et al.* (2011) The genetic landscape of the childhood cancer medulloblastoma. *Science*, **331**, 435–439.
28. Li, Y., Bögershausen, N., Alanay, Y., Simsek Kiper, P.O., Plume, N., Keupp, K., Pohl, E., Pawlik, B., Rachwalski, M., Milz, E., *et al.* (2011) A mutation screen in patients with Kabuki syndrome. *Hum. Genet.*, **130**, 715–724.
29. 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.
30. 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. *Struct. Lond. Engl.* **19**, 101–108.
31. Zhang, P., Lee, H., Brunzelle, J.S. and Couture, J.-F. (2012) The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res.*, **40**, 4237–4246.
32. Li, Y., Han, J., Zhang, Y., Cao, F., Liu, Z., Li, S., Wu, J., Hu, C., Wang, Y., Shuai, J., *et al.* (2016) Structural basis for activity regulation of MLL family methyltransferases. *Nature*, **530**, 447–452.
33. Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Grover, A., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Smil, D., *et al.* (2015) Pharmacological targeting of the Wdr5-MLL interaction in C/EBP α N-terminal leukemia. *Nat. Chem. Biol.*, **11**, 571–578.
34. Cao, F., Townsend, E.C., Karatas, H., Xu, J., Li, L., Lee, S., Liu, L., Chen, Y., Ouillet, P., Zhu, J., *et al.* (2014) Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol. Cell*, **53**, 247–261.
35. Zhu, J., Sammons, M.A., Donahue, G., Dou, Z., Vedadi, M., Getlik, M., Barsyte-Lovejoy, D., Al-awar, R., Katona, B.W., Shilatifard, A., *et al.* (2015) Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature*, **525**, 206–211.
36. Patel, A., Dharmarajan, V. and Cosgrove, M.S. (2008) Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J. Biol. Chem.*, **283**, 32158–32161.

37. Song, J.-J. and Kingston, R.E. (2008) WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J. Biol. Chem.*, **283**, 35258–35264.
38. 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.
39. Dharmarajan, V., Lee, J.-H., Patel, A., Skalnik, D.G. and Cosgrove, M.S. (2012) Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *J. Biol. Chem.*, **287**, 27275–27289.
40. 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.
41. Takahashi, Y., Westfield, G.H., Oleskie, A.N., Trievel, R.C., Shilatifard, A. and Skiniotis, G. (2011) Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 20526–20531.
42. Konarev, P.V., Petoukhov, M.V., Volkov, V.V. and Svergun, D.I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **39**, 277–286.
43. Leitner, A., Walzthoeni, T. and Aebersold, R. (2014) Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protoc.*, **9**, 120–137.
44. Lemak, A., Gutmanas, A., Chitayat, S., Karra, M., Farès, C., Sunnerhagen, M. and Arrowsmith, C.H. (2011) A novel strategy for NMR resonance assignment and protein structure determination. *J. Biomol. NMR*, **49**, 27–38.
45. Lemak, A., Steren, C.A., Arrowsmith, C.H. and Llinás, M. (2008) Sequence specific resonance assignment via Multicanonical Monte Carlo search using an ABACUS approach. *J. Biomol. NMR*, **41**, 29–41.
46. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
47. Goddard, T. and Kneller, D.G. (2004) SPARKY 3.
48. Song, Y., DiMaio, F., Wang, R.Y.-R., Kim, D., Miles, C., Brunette, T., Thompson, J. and Baker, D. (2013) High-resolution comparative modeling with RosettaCM. *Struct. Lond. Engl. 1993*, **21**, 1735–1742.
49. Berlin, K., Castañeda, C.A., Schneidman-Duhovny, D., Sali, A., Nava-Tudela, A. and Fushman, D. (2013) Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J. Am. Chem. Soc.*, **135**, 16595–16609.

50. Alicea-Velázquez, N.L., Shinsky, S.A., Loh, D.M., Lee, J.-H., Skalnik, D.G. and Cosgrove, M.S. (2016) Targeted Disruption of the Interaction between WD-40 Repeat Protein 5 (WDR5) and Mixed Lineage Leukemia (MLL)/SET1 Family Proteins Specifically Inhibits MLL1 and SETd1A Methyltransferase Complexes. *J. Biol. Chem.*, **291**, 22357–22372.
51. Svergun, D., Barberato, C. and Koch, M.H.J. (1995) *CRY SOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
52. Rambo, R.P. and Tainer, J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477–481.
53. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.

FIGURE LEGENDS:

Figure 1. Individual components of the WDR5-RbBP5-MLL1 core complex.

(A) Schematic representation of MLL1, RbBP5 and WDR5 domain organization and constructs used in this study. For clarity, only the C-terminal region is shown for MLL1. WIN: WDR5 interacting motif as previously defined (36); RBS: RbBP5 binding site as defined in this study; SET: catalytic methyltransferase domain; NTD: N-terminal domain; AS+ABM: activation segment and ASH2L binding motif as defined in (32); WBM: WDR5 binding motif (38). (B) Rg-based Kratky plots of experimental SAXS data for MLL1-SET (green), Δ N-WDR5 (blue), RbBP5-NTD (magenta), and full-length RbBP5 (red). (C) Normalized pair distance distribution function $P(r)$ calculated from experimental SAXS data with GNOM. See also Table 1 and Figs S1, S2.

Figure 2. SAXS model for RbBP5 full-length protein.

(A) [^1H - ^{15}N]-TROSY spectra of RbBP5-NTD (left) and RbBP5-CTD (right) constructs. For the CTD, amide resonances are clustered between 7.5 and 8.5 ppm in the ^1H dimension indicating that it is unfolded, while spectra for the NTD is consistent with a structured domain. (B) A representative member of the most populated SES ensemble model reflects a structured N-terminal domain and a flexible, but non-random, C-terminus. (C) The clear difference in Rg distribution profiles for the initial pool of 30,000 models (with random conformations of the C-terminal part (dashed)) vs. the SAXS-derived SES ensemble (solid) indicate that the CTD, in the context of full-length RbBP5, is not randomly disordered. Pair distance $P(r)$ distribution function calculated for the experimental data (black circles) and for the SAXS-derived SES ensemble (red line) are displayed in the inset. See also Figs 1C, S1-S3.

Figure 3. WDR5/MLL1 and RbBP5/MLL1 Binary Complexes.

(A) Rg-based Kratky plots of experimental SAXS data for binary complexes of WDR5/MLL1-WIN (orange) and WDR5/RbBP5 (maroon). (B) WDR5 residues (pink) that show substantial CSPs in [^1H - ^{15}N]-TROSY spectra when titrated with WIN and RbBP5 peptides (green). The perturbation patterns were mapped onto, and are in excellent agreement the WDR5/WIN (PDB: 4ESG) and WDR5/RbBP5 (PDB:2XL2) structures. (C-E) Interaction between RbBP5 constructs and MLL1-WIN/MLL1-SET. GST-pull down experiments show direct interaction between (C) RbBP5-FL and (D) RbBP5-NTD with MLL1 constructs along with the control GST experiments. (E) There was no interaction detected between RbBP5₃₂₀₋₄₁₀ and MLL1-SET. Lanes correspond to L=loaded protein mixture; FT=flow through; W=wash; E=eluate. See also Figs S3, S4.

Figure 4. Dynamic model of trimeric WDR5/RbBP5/MLL1-WIN complex derived from SAXS and cross-link data.

(A) Rg-based Kratky plot of experimental SAXS data for WDR5/RbBP5/MLL1-WIN indicates a high degree of flexibility within the complex. (B) Sequence mapping of intra-protein and inter-protein cross-links. Intra-protein cross-links are indicated with purple arcs. Inter-protein cross-links between globular domains are indicated with green lines and inter-protein cross-links within flexible regions are indicated with black dashed lines. (C) Rg distribution for the optimal ensemble of trimeric complex models is shown by a solid black line. Experimental SAXS

profiles (black circles) plotted with the theoretical profiles (red line) averaged over the ensemble are shown in the inset. (D) Surface representation of the four most populated models in the optimal ensemble models of WDR5/RbBP5/MLL1-WIN overlaid with average *ab-initio* SAXS-predicted molecular envelope (gray mesh). WDR5, MLL1 and RbBP5 are colored in blue, green and red, respectively. (E) The most populated (~40%) model of the complex is shown by a surface representation of the globular regions and a backbone trace of the flexible regions. For clarity, the C-terminus of RbBP5 (i.e. RbBP5₃₈₂₋₅₃₈) is not shown and the residues from the AS+ABM region are in dark orange, WIN in green, WBM in violet and RBS in red. The cross-links between globular domains are shown by solid black lines and the cross-linked lysine residues are shown in cyan. (F) The ensemble distribution of the distance between C α atoms of Lys residues involved in cross-links between MLL1-SET and WDR5, RbBP5-NTD and WDR5 are shown by solid blue and black lines, respectively. See also Fig S5.

Figure 5. MLL1-SET domain binding interface between the RBS and RbBP5-NTD.

(A) Surface representation of MLL1 SET domain (residues 3815-3969). The helix of the SET-I sub-domain is shown in cyan. Residues involved in the interface with RbBP5-NTD and the RBS are shown in pink and red. Conserved residues that are involved in the interface with the AS+AMB and ASH2L_{SPRY}/RbBP5_{AS+ABM} in the MLL3 complex (32) are shown in yellow/red. Two GLY residues, that serve as the hinge points of SET-I rotation (32), are shown in blue. (B) Sequence alignment for MLL1, MLL2 and MLL3 linker residues between the WIN motif and the SET domain showing the residues that are not conserved. The WIN motif is highlighted in a green bracket. The residues within the MLL1 RBS segment are shown in a red bracket. (C) No interaction was detected between MLL1-SET-7D-GST (aa 3786-3793 deletion) and MLL1-SET-Q3787V/P3788L/Y3791G-GST (MLL1 to MLL2 mutation) mutants and RbBP5-NTD by GST pulldown experiments. Lanes are labeled as in Figure 3. (D) Deletion of the RBS region of MLL1 (MLL1-SET-7D) attenuates the catalytic activity of the trimeric complex. See also Fig S5.

Figure 6. OICR-9429 attenuates the assembly of a functional trimeric complex.

(A) Size exclusion chromatography of RbBP5/WDR5/MLL1-WIN (at concentrations of ~ 13 μ M/12 μ M/6 μ M respectively) in the absence (navy) or presence of (cyan) of ~ 5-fold molar excess of OICR-9429. This compound binds to WDR5 ($K_D = 93$ nM) (33) and competes with the MLL1 WIN motif for binding to WDR5. (B) SDS-PAGE of elution fractions (L=loaded protein mixture). Fractions containing the trimeric and WDR5/MLL1-WIN (shoulder at 15.2 ml) complexes are not recovered from the column when run in the presence of OICR-9429. (C) OICR-9429 inhibits the catalytic activity of the trimeric complex.

Table 1. SAXS parameters for data validation and interpretation

	Rg^a (Å)	Rg^b (Å) (real)	D_{max}^c (Å)	V_c^d	Mw^e (kDa)	NSD^f (SAXS envelope)
<i>Individual</i>						
MLL1-SET^g	20.2	20.8	73	207	22.0 (21.6)	0.58
ΔN-WDR5^h	19.8	19.9	70	251	25.9 (34.1)	0.60
RbBP5-NTDⁱ	21.5	21.7	76	276	28.7 (36.8)	0.72
RbBP5^j	32.1	33.0	113	489	60.6 (59.1)	0.64
<i>Dimeric complexes</i>						
WDR5/ MLL1-WIN^k	32.5	33.6	120	372	56.5 (64.7)	0.79
WDR5/RbBP5^l	39.8	41.1	140	669	91.2 (96.5)	0.79
<i>Trimeric complex</i>						
WDR5/RbBP5/ MLL1-WIN^m	49.1	51.8	183	929	135.5 (124.6)	0.69
Oi						

a) Radius of gyration calculated using Guinier fit

b) Radius of gyration calculated using GNOM (51)

c) Maximum distance between atoms from GNOM

d) Volume of correlation (52)

e) Molecular weight estimated from SAXS data using V_c (52). The MW expected from the sequence is shown in the parentheses

f) NSD: Normalized spatial discrepancy; The values given are the average and standard deviation from fifteen runs of DAMMIF (53)

g) MLL1₃₇₈₅₋₃₉₆₉

h) WDR₅₂₄₋₃₃₄

i) RbBP5₁₀₋₃₄₀

j) RbBP5 (full-length)

k) full-length WDR5 and MLL1₃₇₄₅₋₃₉₆₉

l) full-length WDR5 and RbBP5

m) full-length WDR5, full-length RbBP5, and MLL1₃₇₄₅₋₃₉₆₉

A

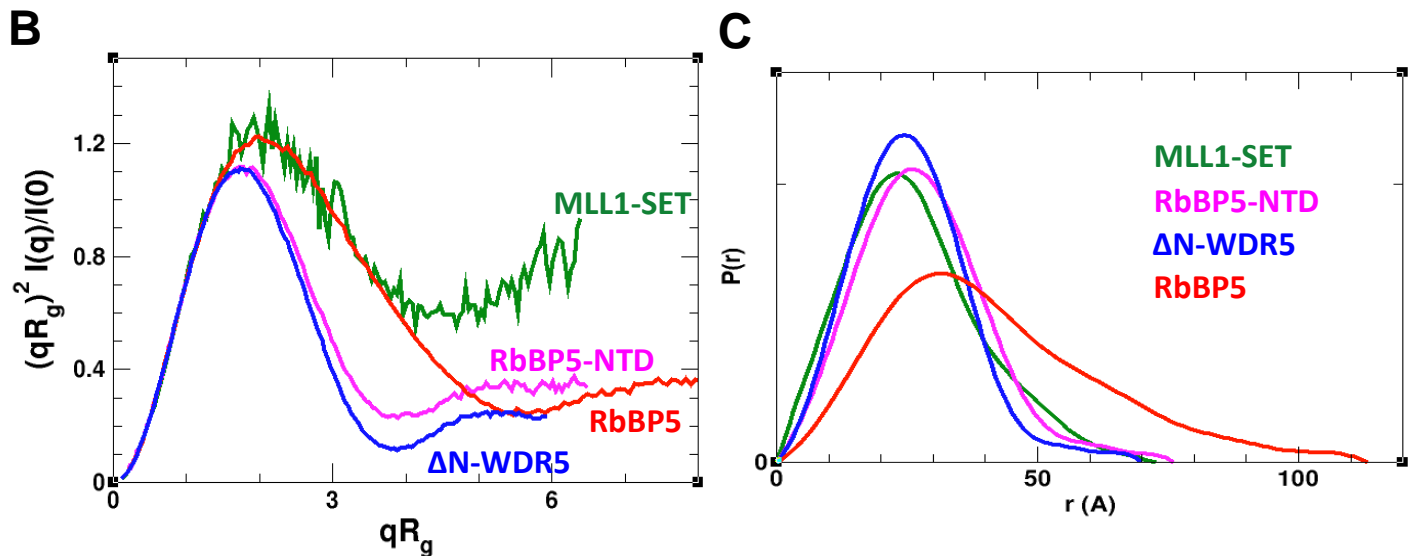
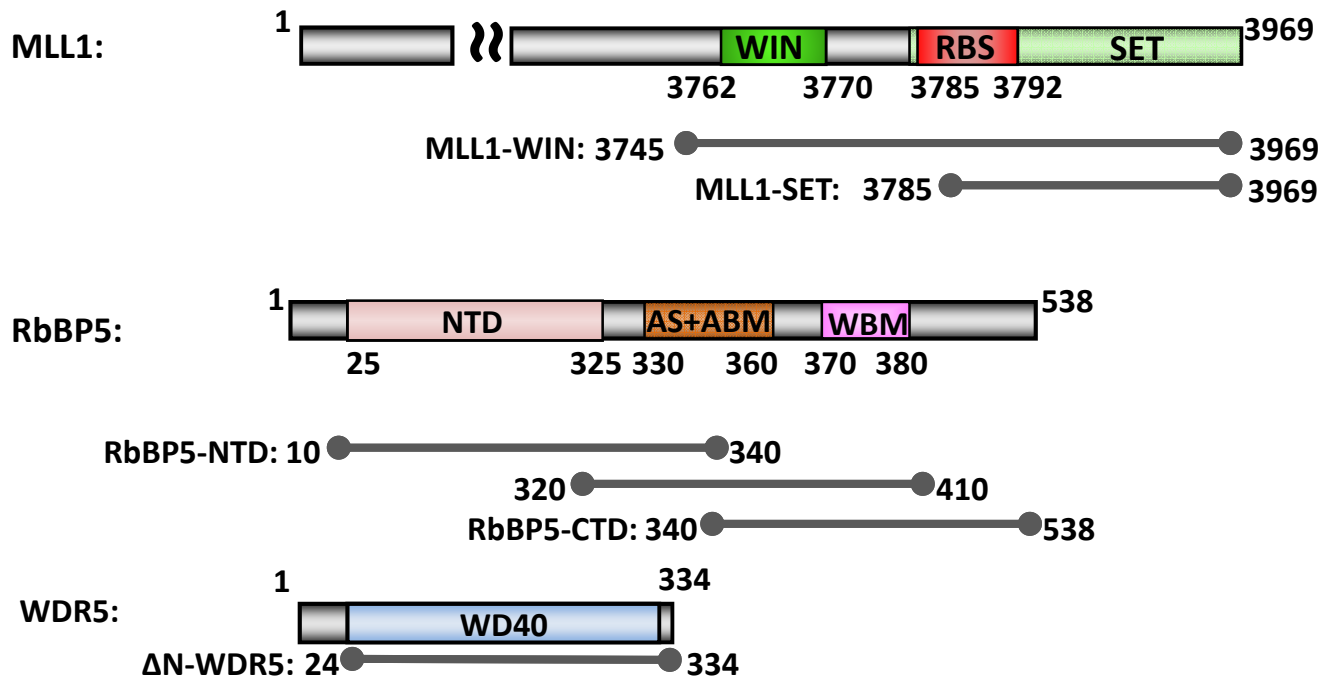


Figure 1.

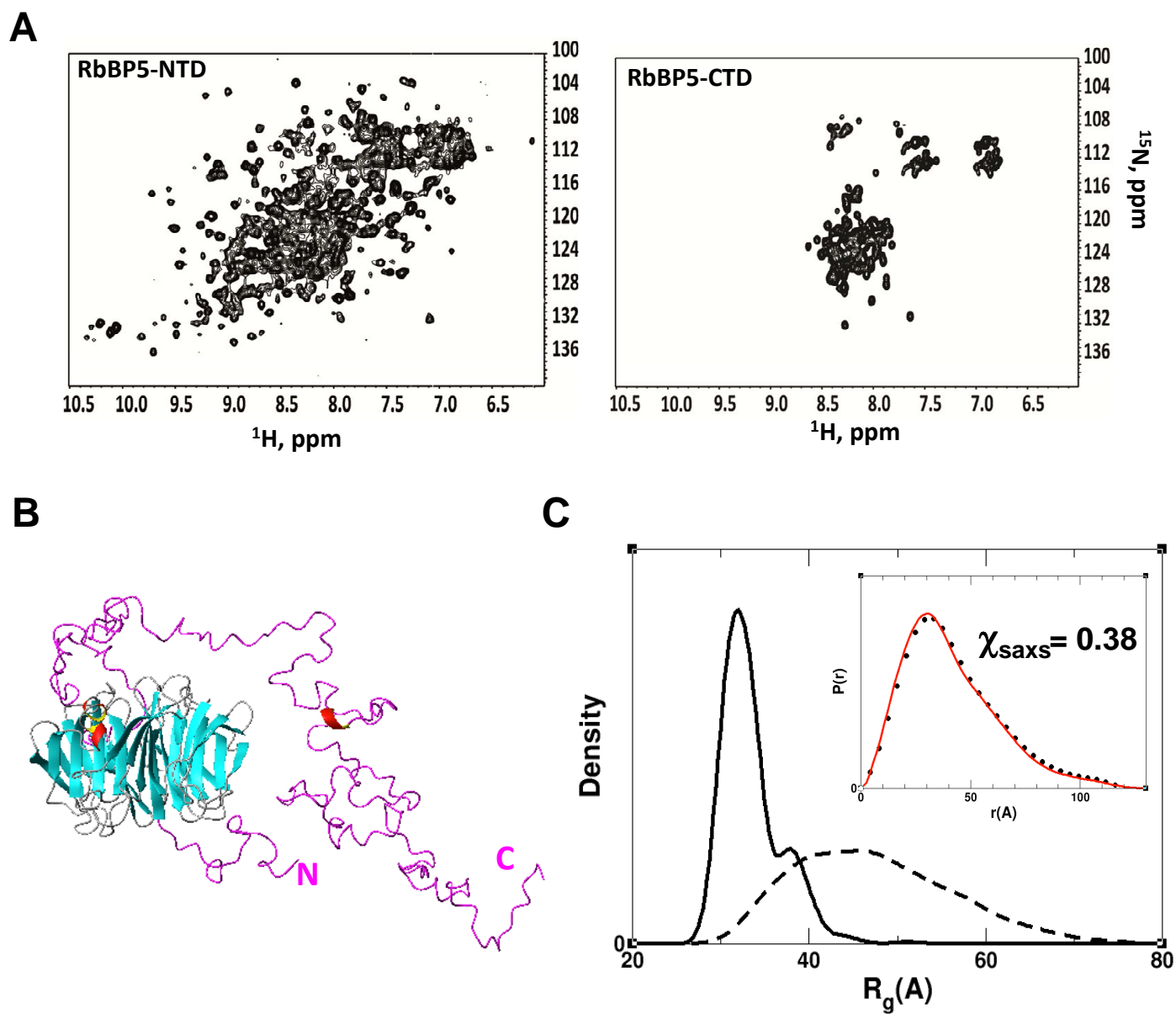


Figure 2.

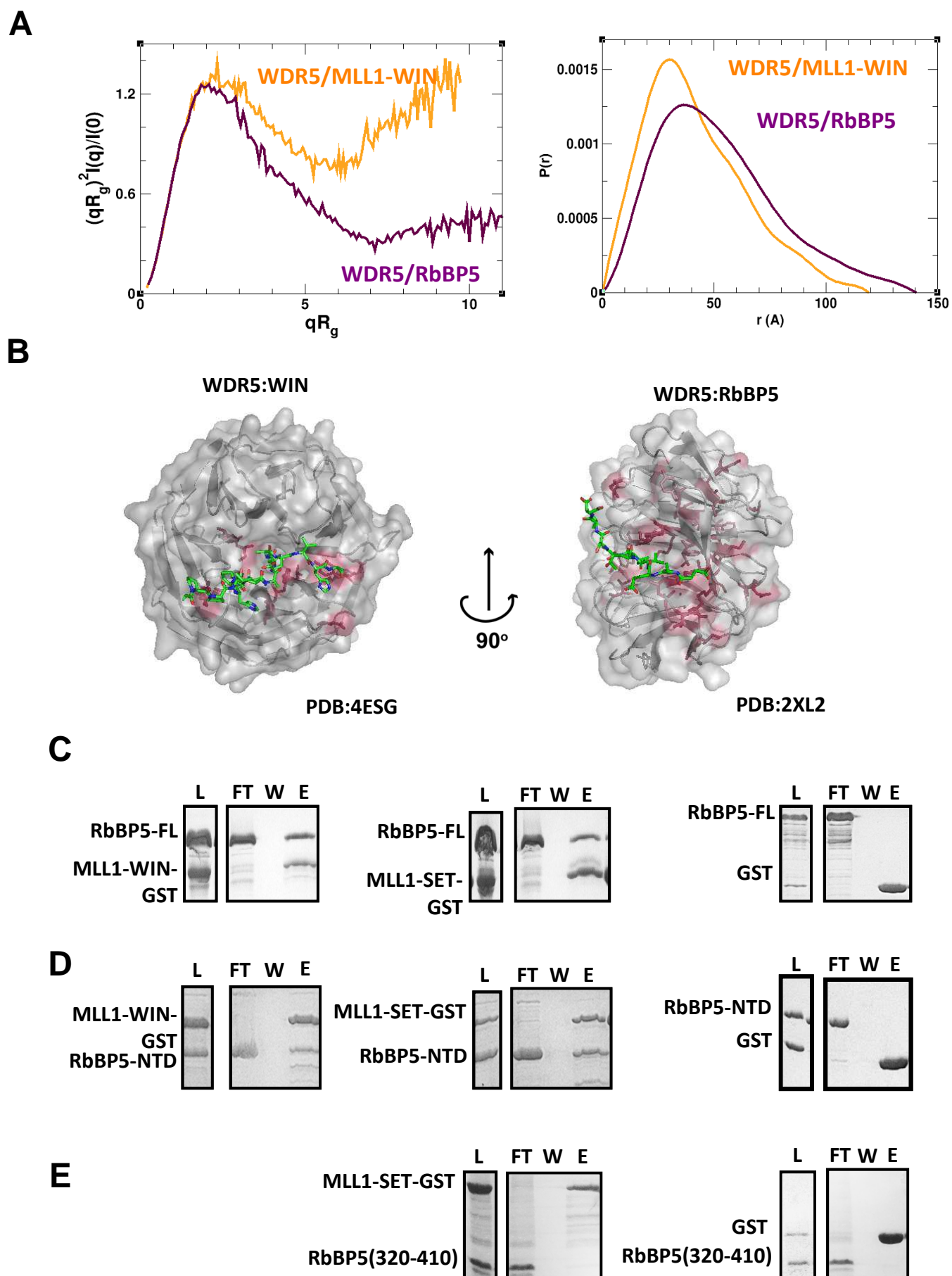
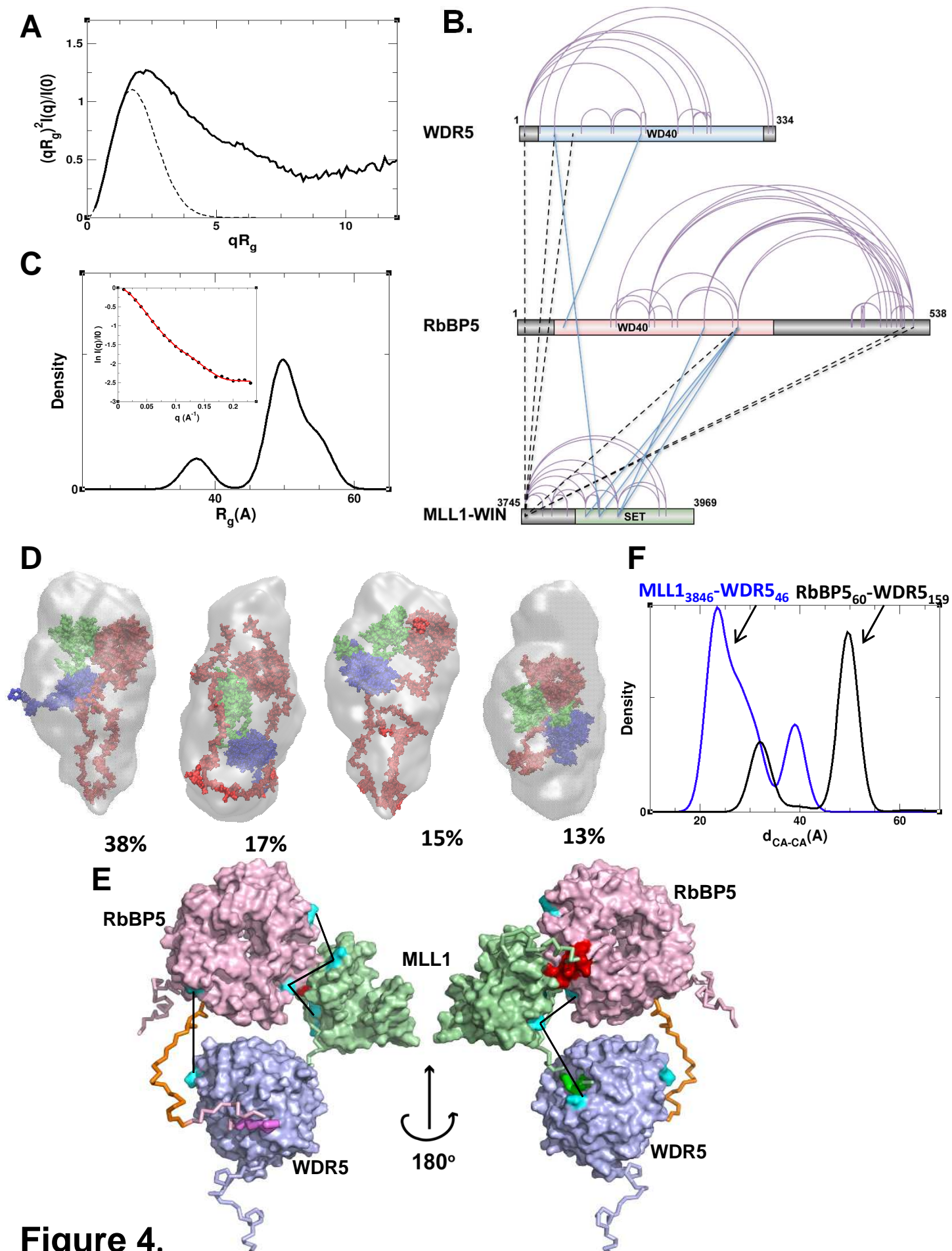


Figure 3.



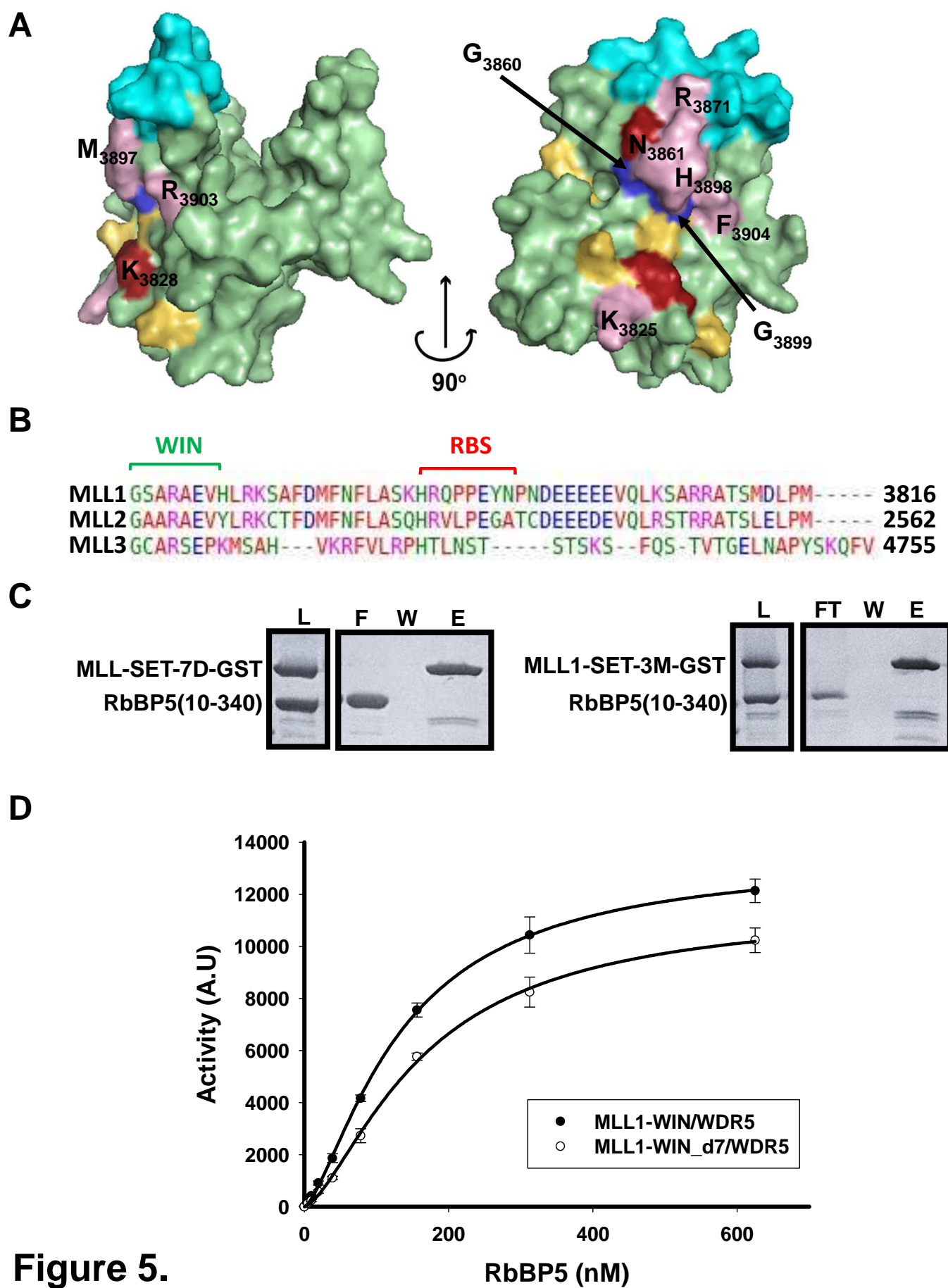


Figure 5.

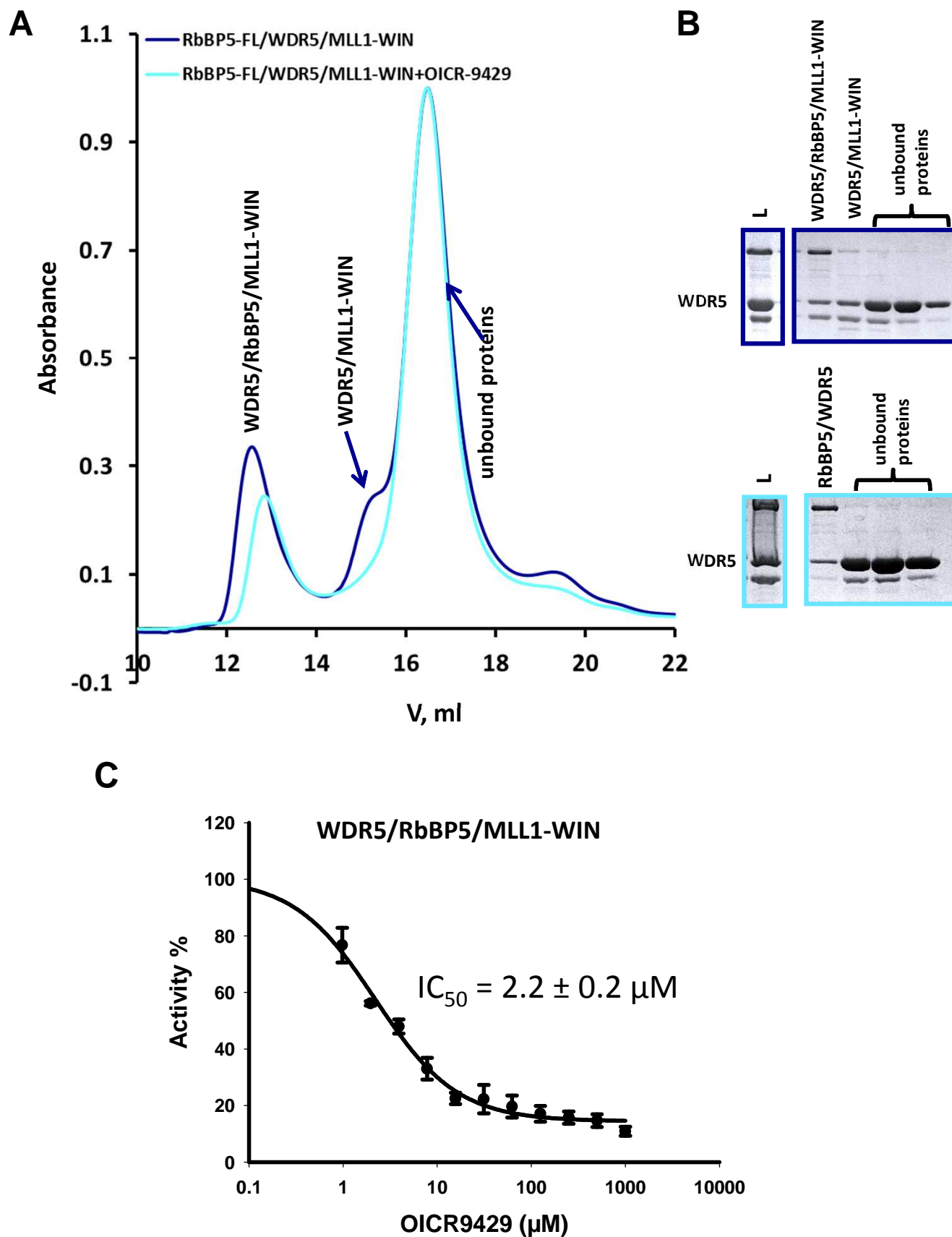


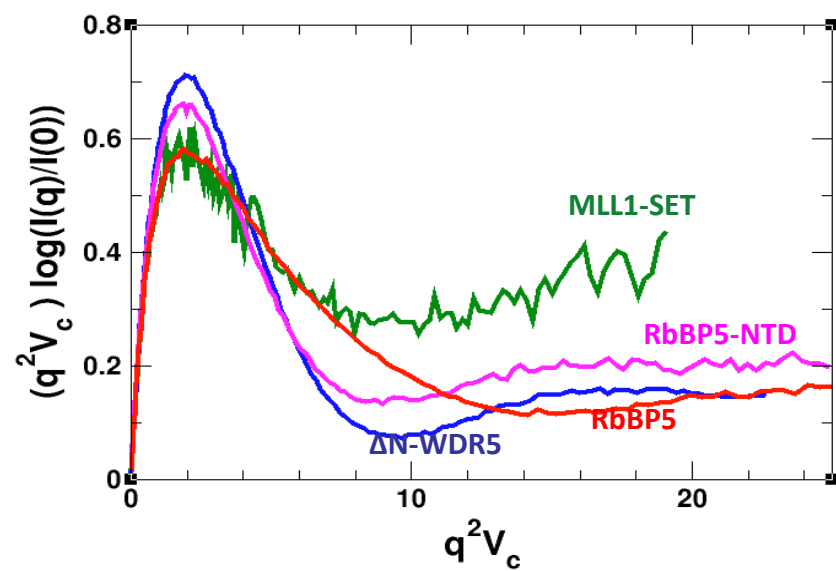
Figure 6.

Supplementary Data for:

MLL1 minimal catalytic complex is a dynamic conformational ensemble susceptible to pharmacological allosteric disruption

Lilia Kaustov, Alexander Lemak, Hong Wu, Marco Faini, Scott Houliston, Lixin Fan, Xianyang Fang, Hong Zeng, Shili Duan, Abdellah Allali-Hassani, Masoud Vedadi, Ruedi Aebersold, Yunxing Wang, Cheryl H. Arrowsmith

A



B

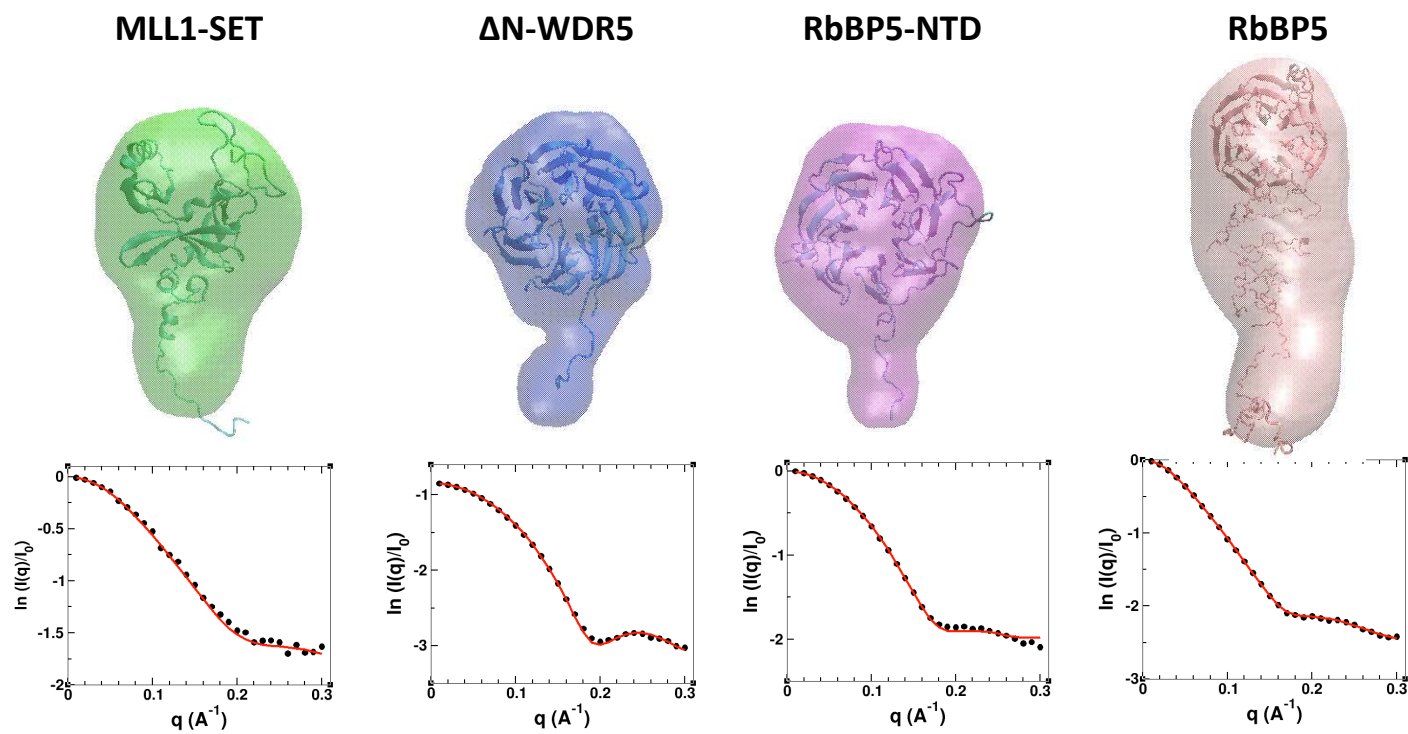


Figure S1

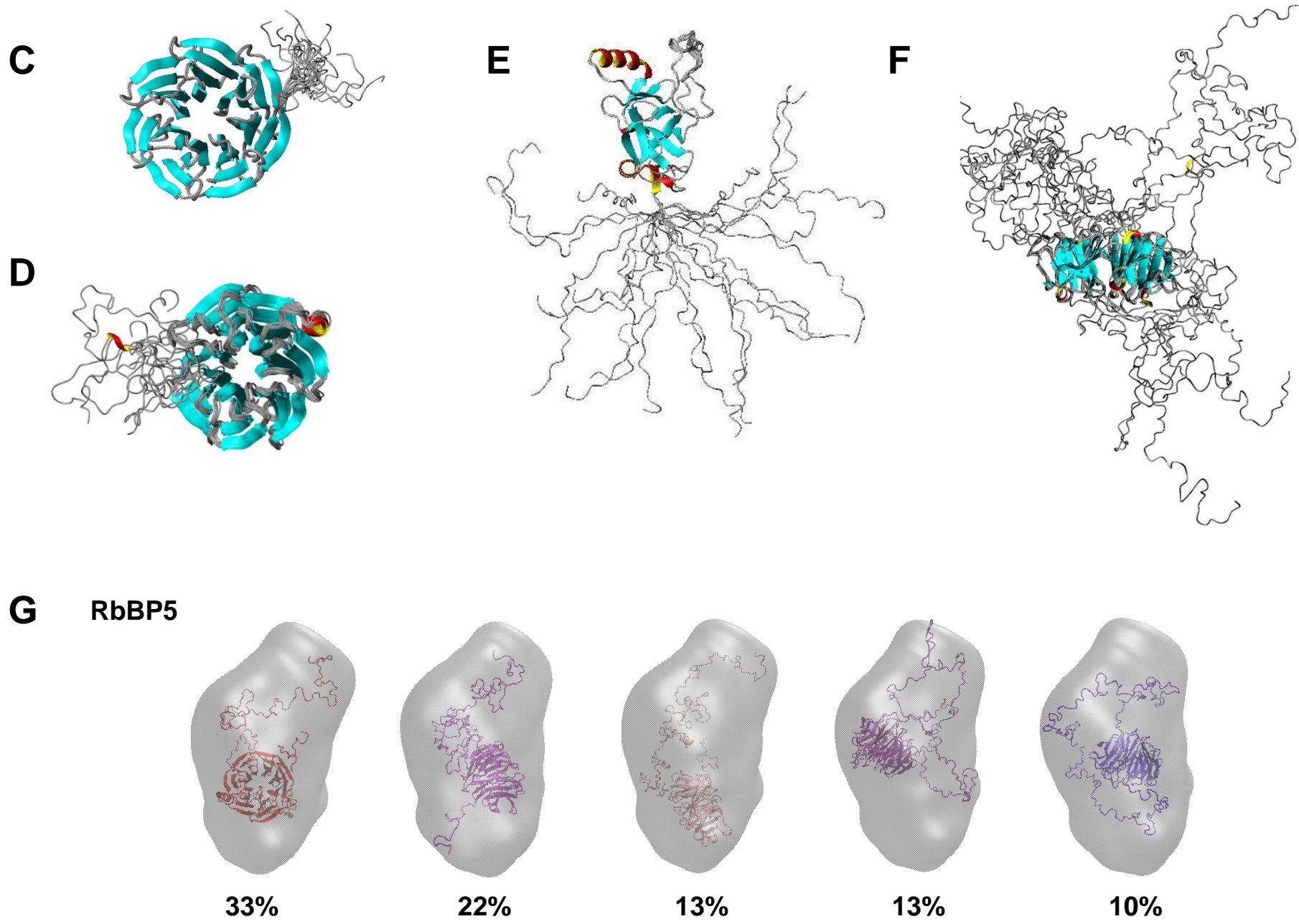


Figure S1. (related to Figures 1, 2, and Table 1)

Individual components of Wdr5-RbBP5-MLL1 core complex exhibit different degrees of flexibility. (A) V_c -based Kratky plots of experimental SAXS data for MLL1-SET (green), Δ N-WDR5 (blue), RbBP5-NTD (magenta), and full-length RbBP5 (red). (B) *Top row*: *Ab-initio* SAXS-predicted molecular envelopes with fitted molecular models. Filtered envelopes were calculated from fifteen runs of DAMMIF. The fitted models for MLL1-SET and Δ N-WDR5 are the crystal structures (PDBID: 2W5Y and 2H9M, respectively). For RbBP5-NTD, a homology model obtained with ROSETTA is shown. For full-length RbBP5, the five most populated models of the dynamic ensemble calculated using the SES method are shown (see also panel G). *Lower row*: Experimental SAXS profiles (black circles) superimposed with theoretical profiles (red lines) calculated for dynamical models of MLL1-SET, Δ N-WDR5, RbBP5-NTD, and full-length RbBP5 (for details see the Materials and Methods section). Ribbon representation for the optimal ensembles of Δ N-WDR5 (C), RbBP5-NTD (D), MLL1-SET with flexible N-terminal tail (E), and RbBP5 (F). For all ensembles shown, the structured parts from different members are superimposed. (G) Ribbon diagrams of the five most populated species in the optimal ensemble models of RbBP5, overlaid with the average *ab-initio* SAXS-predicted molecular envelope (gray mesh).

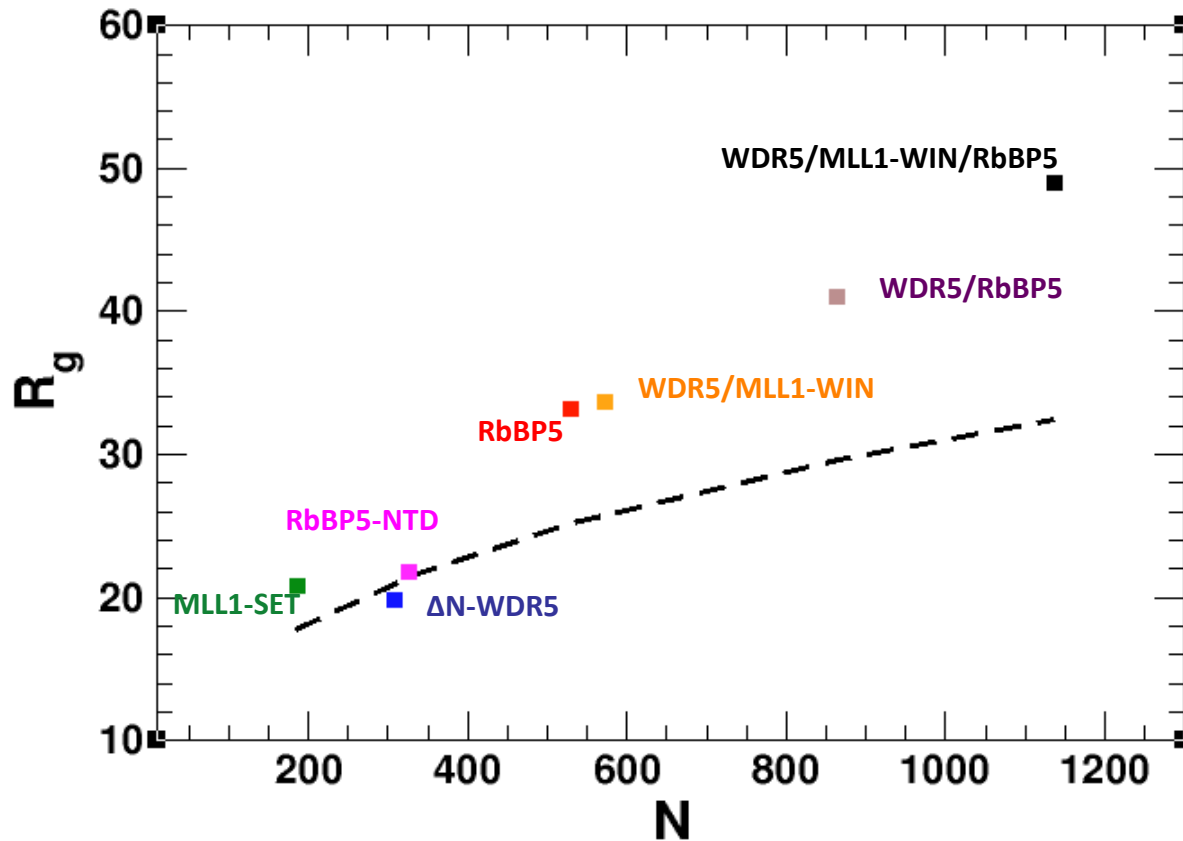


Figure S2. (related to Figures 1,3,4)

Comparison of the experimental R_g with the expected R_g for globular proteins. Experimental R_g derived from SAXS data vs. the number of residues (N) for MLL1-SET (green), Δ N-WDR5 (blue), RbBP5-NTD (magenta), full-length RbBP5 (red), WDR5/MLL1-WIN complex (orange), WDR5/RbBP5 complex (maroon), and WDR5/MLL1-WIN/RbBP5 complex (black). Theoretical R_g expected for a globular protein with the same molecular mass is shown by the dashed black line.

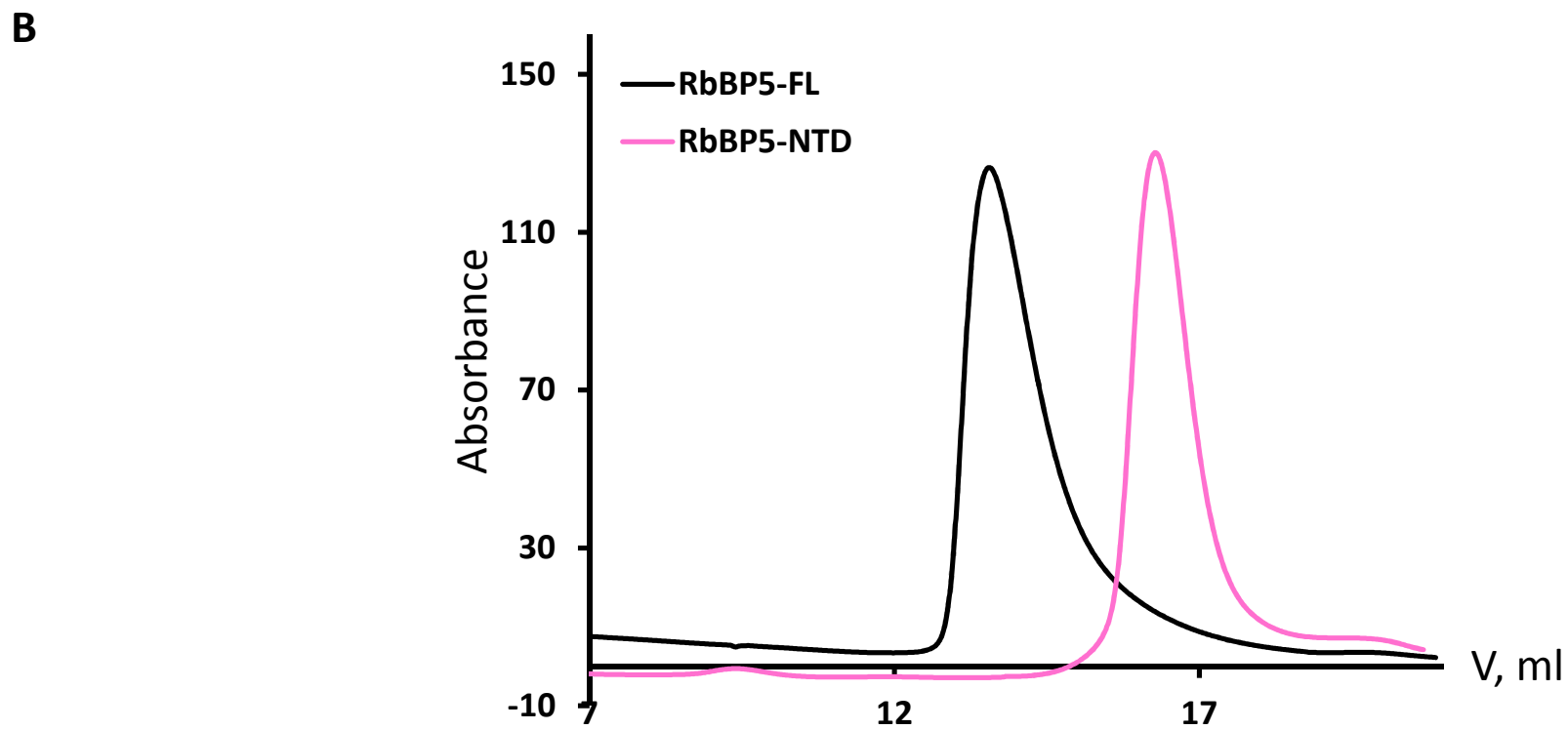
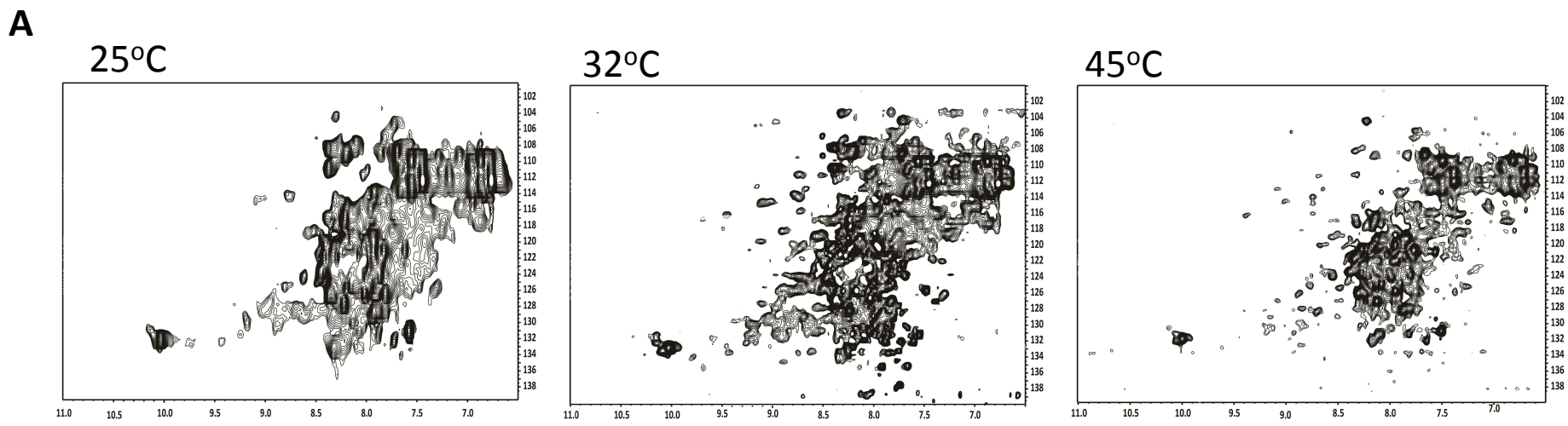


Figure S3.

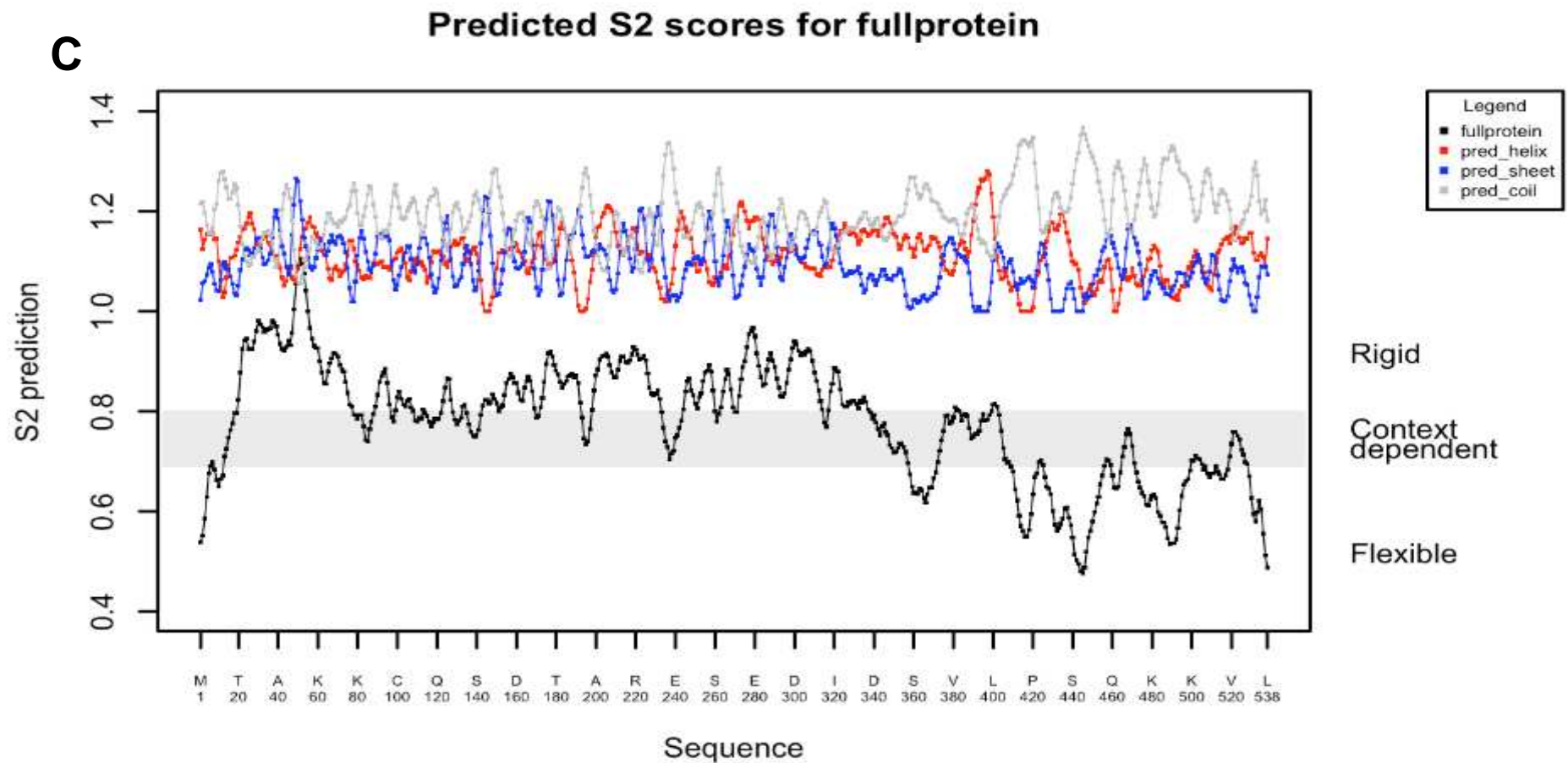
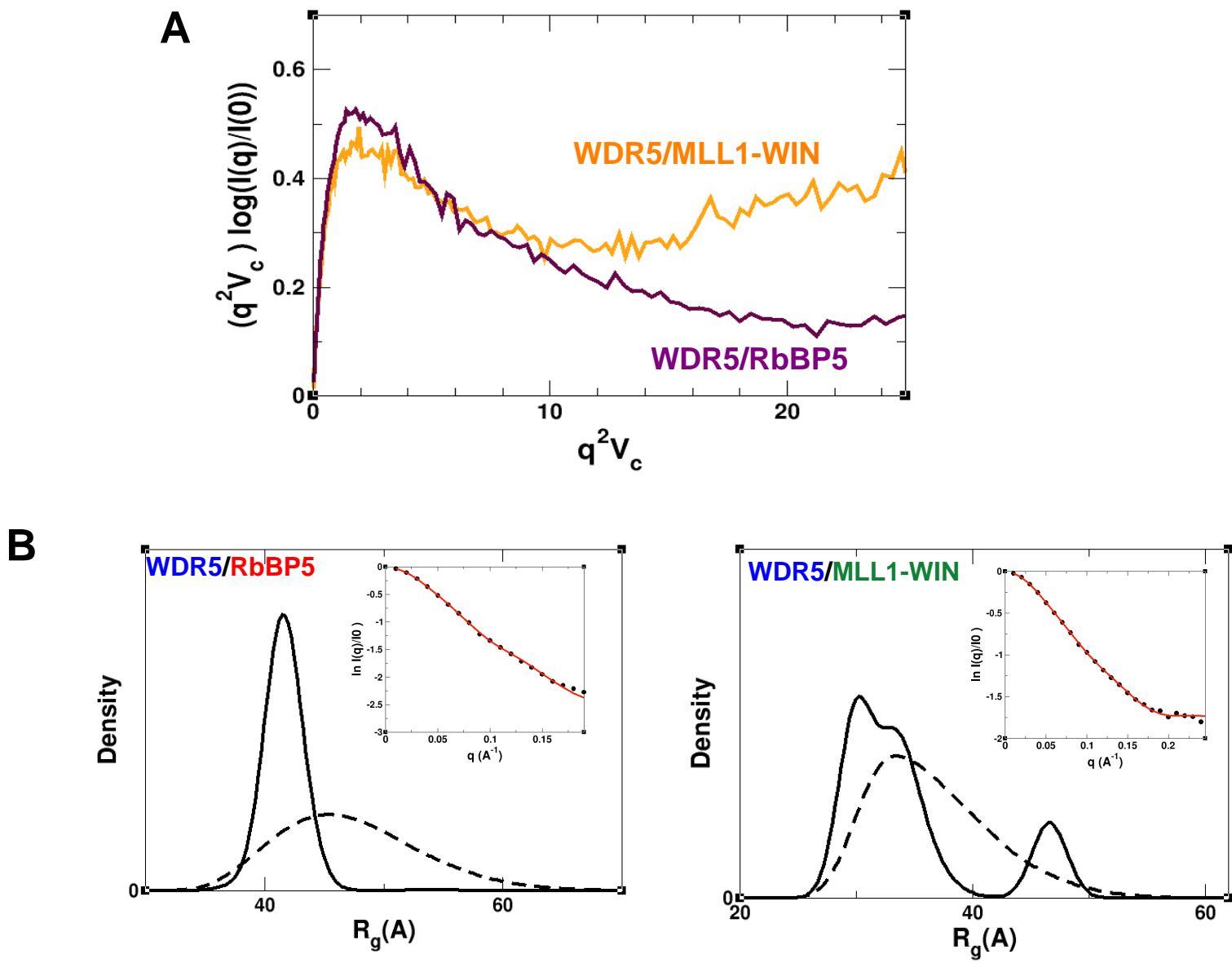


Figure S3. (related to Figure 2)

Structural characterization of full-length RbBP5. (A) $[^1\text{H}-^{15}\text{N}]$ -TROSY spectra of RbBP5 collected at 25°C, 32°C and 45°C. (B) Gel filtration profiles of RbBP5 full-length (**black**) and RbBP5-NTD (**pink**). RbBP5-FL runs with a higher molecular mass (MW=112.2 kDa, V_e =13.3ml) than its calculated value (MW=59.2kDa) indicating a high degree of disorder. RbBP5-NTD runs as a globular, folded domain with an elution volume (V_e =16.2 ml) corresponding to its expected molecular mass (MW=36.8kDa). (C) S2 order parameter for RbBP5 predicted from the protein sequence using DYNAMINE.



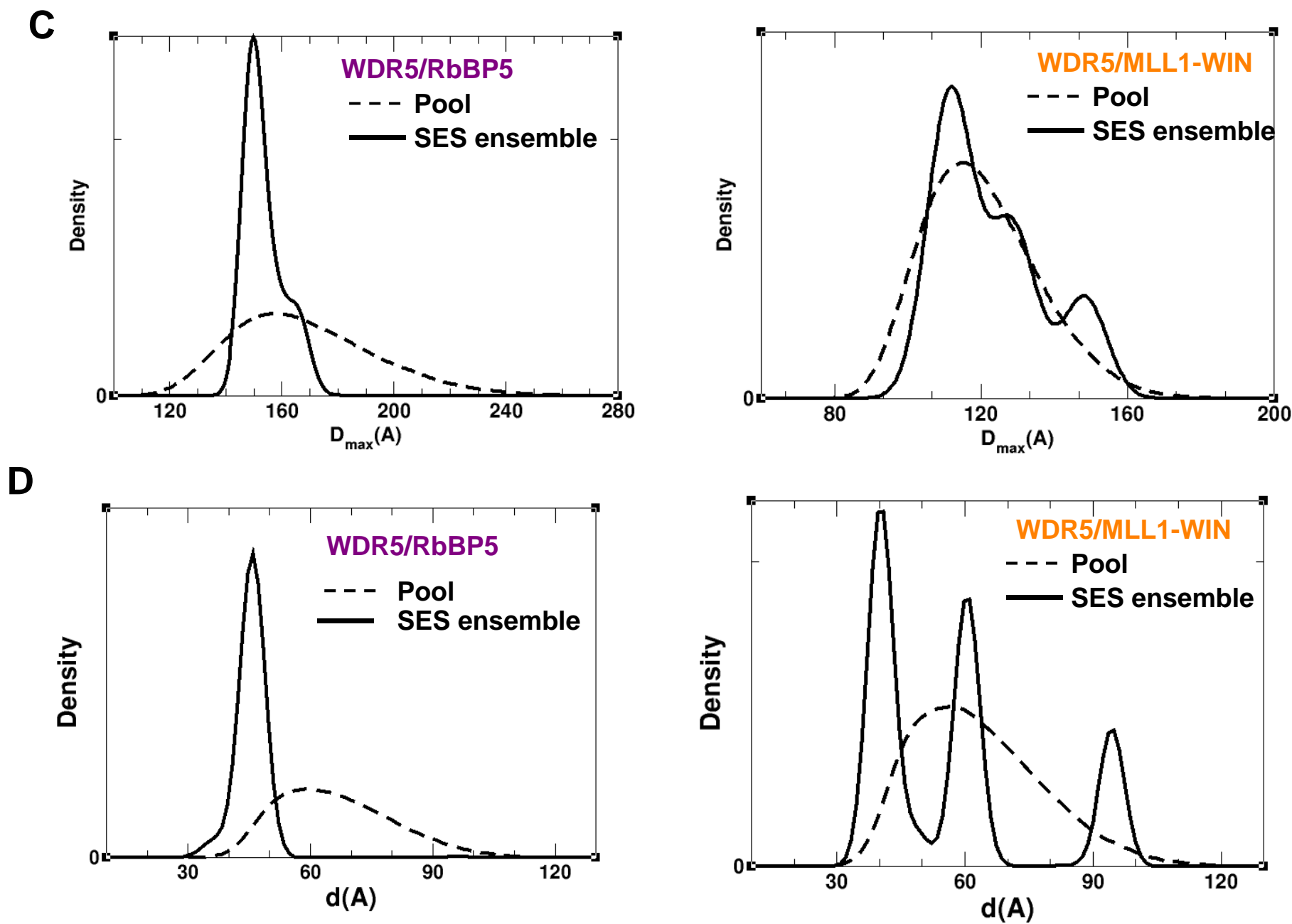


Figure S3.

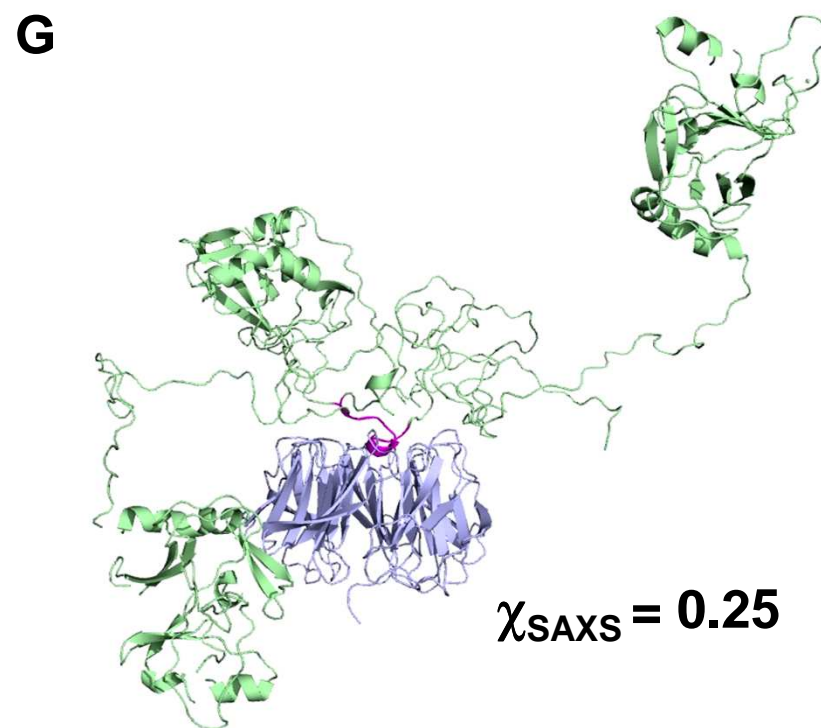
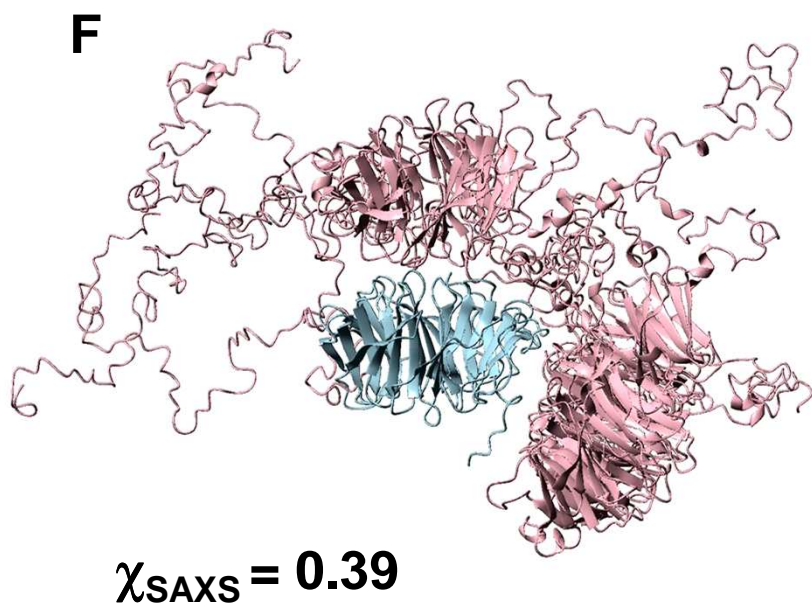
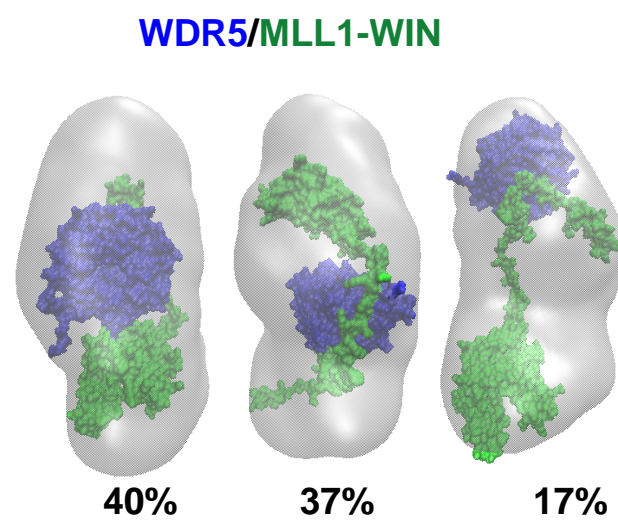
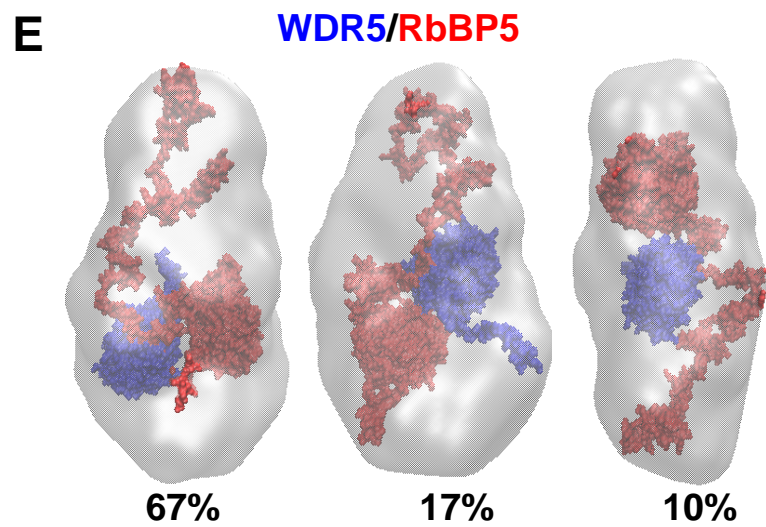


Figure S4. (related to Figure 3)

Structural characterization of WDR5/MLL1-WIN and WDR5/RbBP5 binary complexes. (A) V_c -based Kratky plots derived from experimental SAXS data for WDR5/MLL1-WIN (orange) and WDR5/RbBP5 complexes (maroon). (B) R_g distributions for the initial pool of random structures (dashed), and for the selected SES ensemble (solid). Experimental SAXS profiles (black circles) in a match with theoretical profiles (red lines) averaged over the SES ensemble are shown (inset). (C) D_{\max} distribution for WDR5/MLL1-WIN (left) and WDR5/RbBP5 (right). The distribution for the initial pool of random structures (dashed lines) and for the selected SES ensemble (solid lines) are shown. WDR5/RbBP5 favors structures that are more condensed than fully extended, while WDR5/MLL1-WIN has a higher degree of flexibility and exhibits a population of fully extended structures. (D) Distribution of the distance between centers of mass of WDR5 and globular RbBP5 (right panel), and WDR5 and MLL1-SET (left panel) domains in the optimal SES ensembles of the binary complexes. (E) Surface representation of models for the binary complex overlaid with the *ab-initio* SAXS-predicted molecular envelopes (gray mesh) for WDR5/RbBP5 (left) and WDR5/MLL1-WIN (right). The three most populated models of the optimal SES ensemble are shown. WDR5, MLL1, and RbBP5 are colored in blue, green and red, respectively. (F-G) Ribbon diagram of the representative models of the optimal ensemble of WDR5/RbBP5 (F) and WDR5/MLL1-WIN (G). WDR5, RbBP5 and MLL1 are colored in blue, pink, and green, respectively. The WIN motif of MLL1 is shown in magenta. WDR5 structured domain of different members of an ensemble are superimposed. The goodness-of-fit of the optimal ensemble to SAXS data is shown by the χ_{SAXS} value.

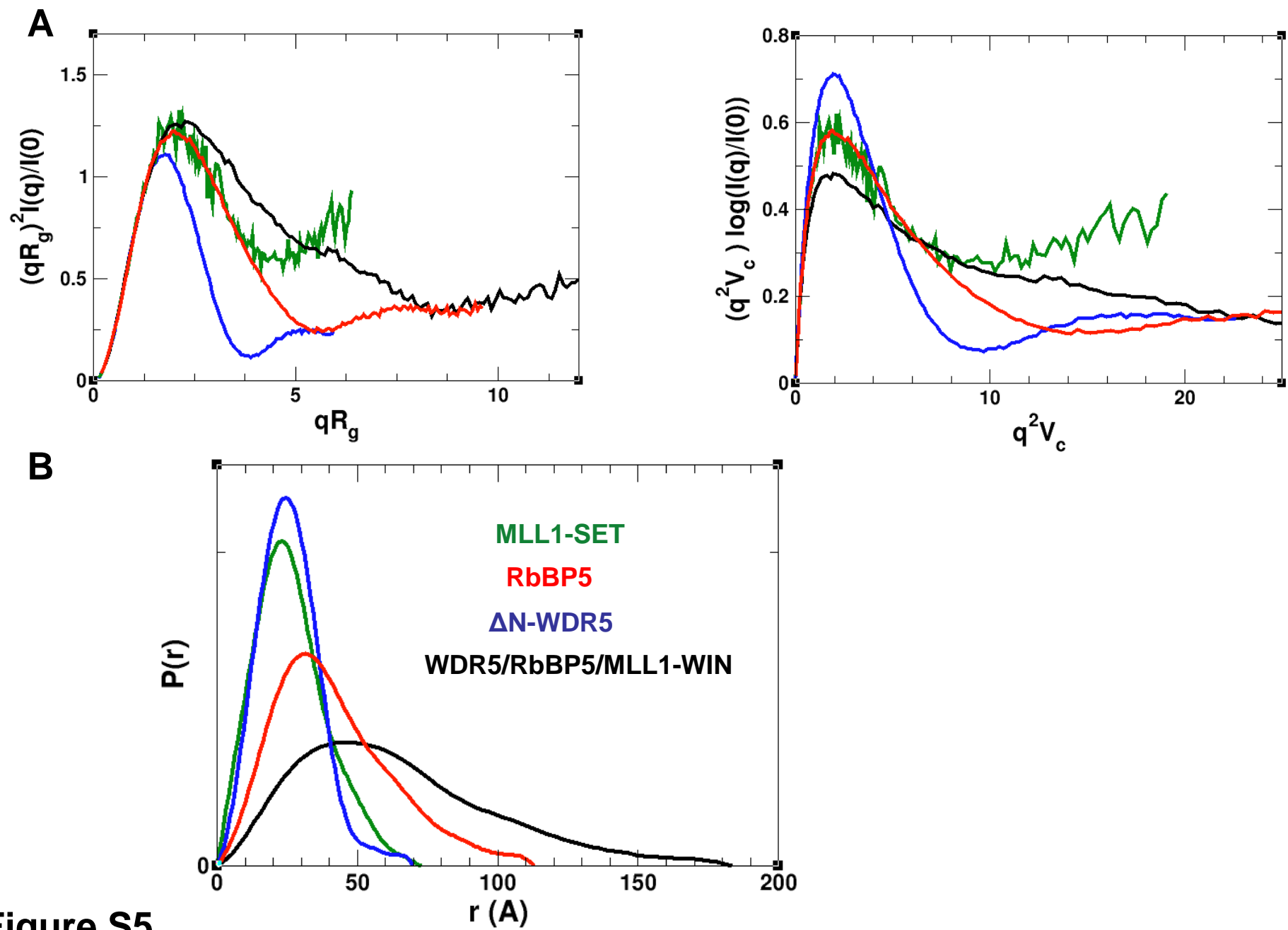
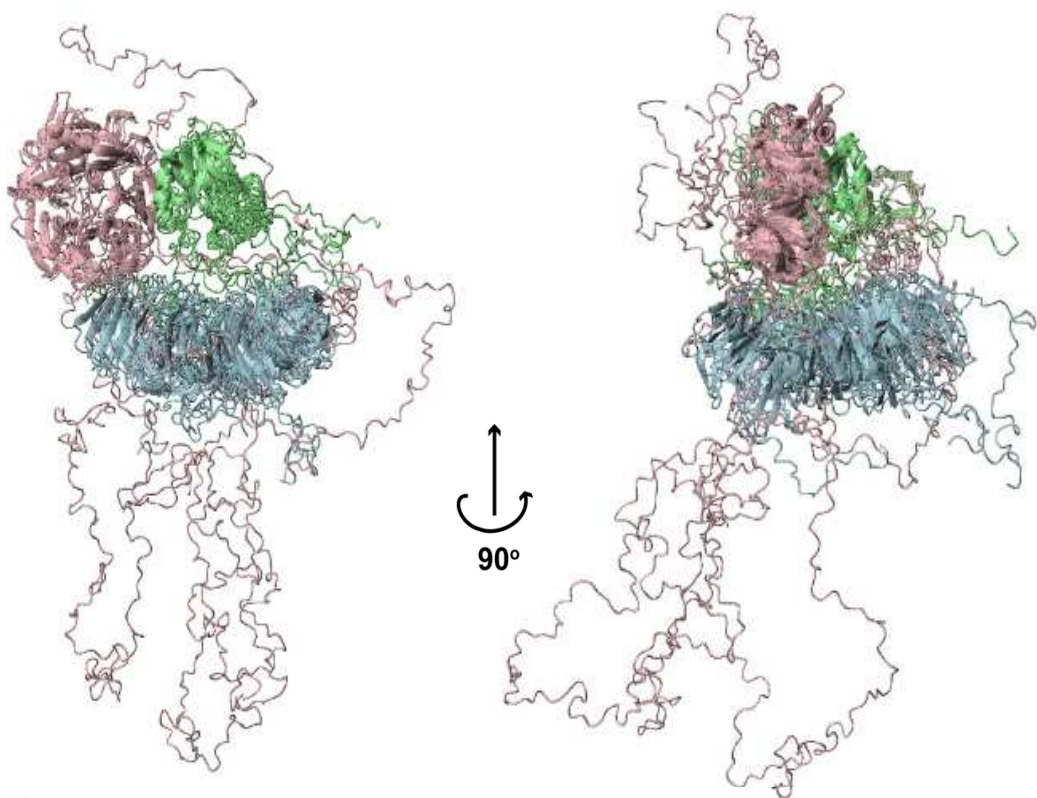


Figure S5.

C



D

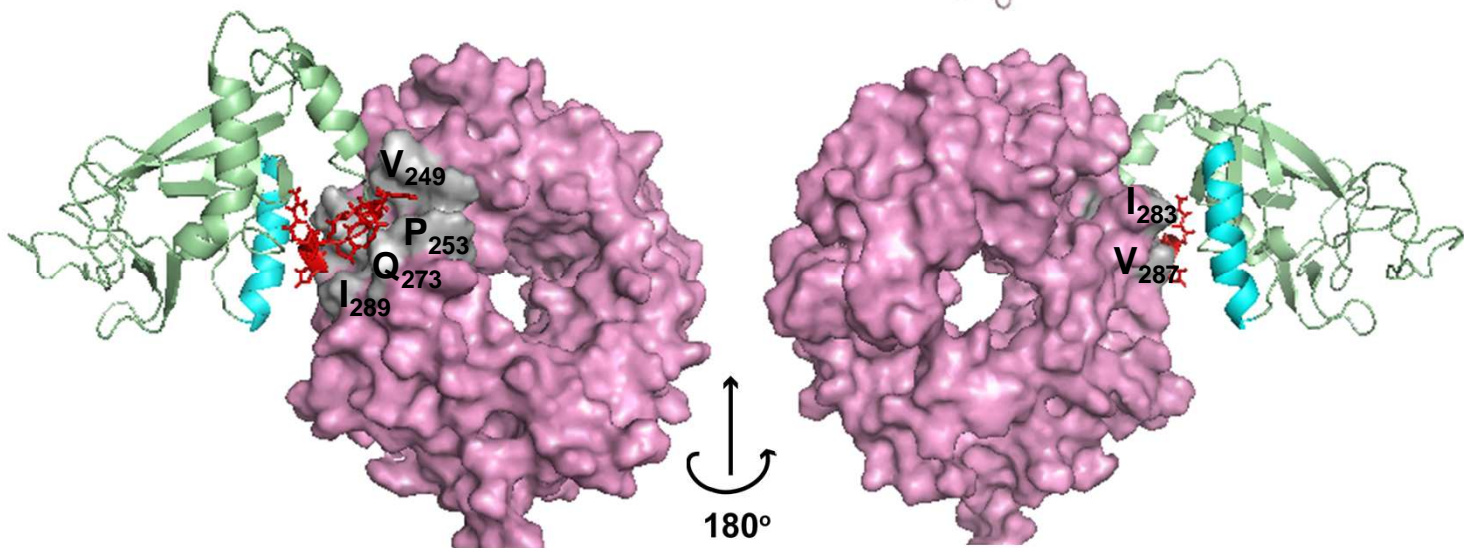


Figure S5.

E

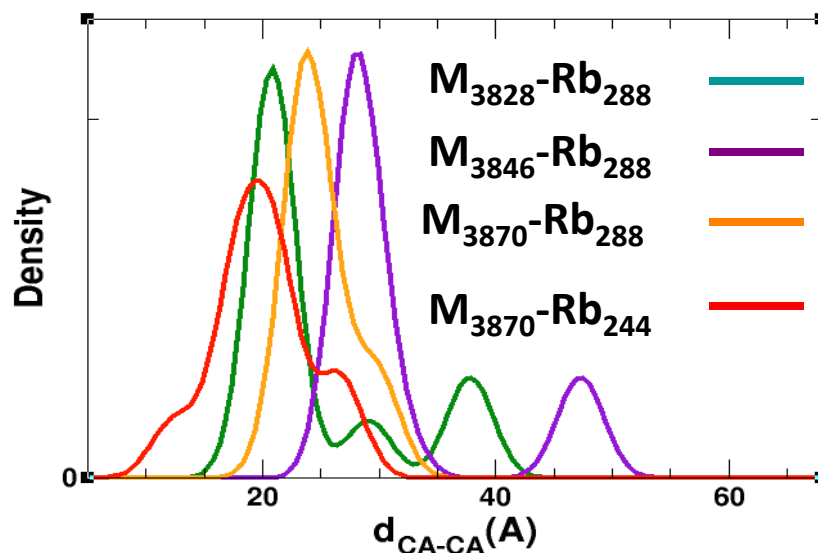


Figure S5. (related to Figure 4)

Structural characterization of trimeric WDR5/RbBP5/MLL1-WIN complex. (A) Rg-based (left panel) and Vc-based (right panel) Kratky plots of experimental SAXS data for the WDR5/RbBP5/MLL1-WIN complex (black), and its components, Δ N-WDR5 (blue), MLL1-SET (green), and full-length RbBP5 (red line). SAXS data indicate a high degree of flexibility for the complex. (B) Normalized pair distance distribution function $P(r)$ calculated from experimental SAXS data with GNOM. (C) Ribbon diagram of six representative models of the optimal ensemble of the complex. These models comprise 86% of the ensemble. WDR5, RbBP5 and MLL1 are colored in blue, pink, and green, respectively. Structured N-terminal domain of RbBP5 and SET domain of MLL1 from different members of the ensemble are superimposed. (D) Molecular model of RbBP5-NTD/MLL1-SET complex. RbBP5-NTD surface is colored in pink. MLL1-SET is shown as a ribbon diagram in pale green. The helix of the SET-I sub-domain is highlighted in cyan. The RBS (MLL1₃₇₈₅₋₃₇₉₂) residues are shown in red. The RbBP5 residues that are in contact with MLL1-SET are shown in gray. (E) Ensemble distribution of four X-links between RbBP5-NTD and MLL1-SET domains.

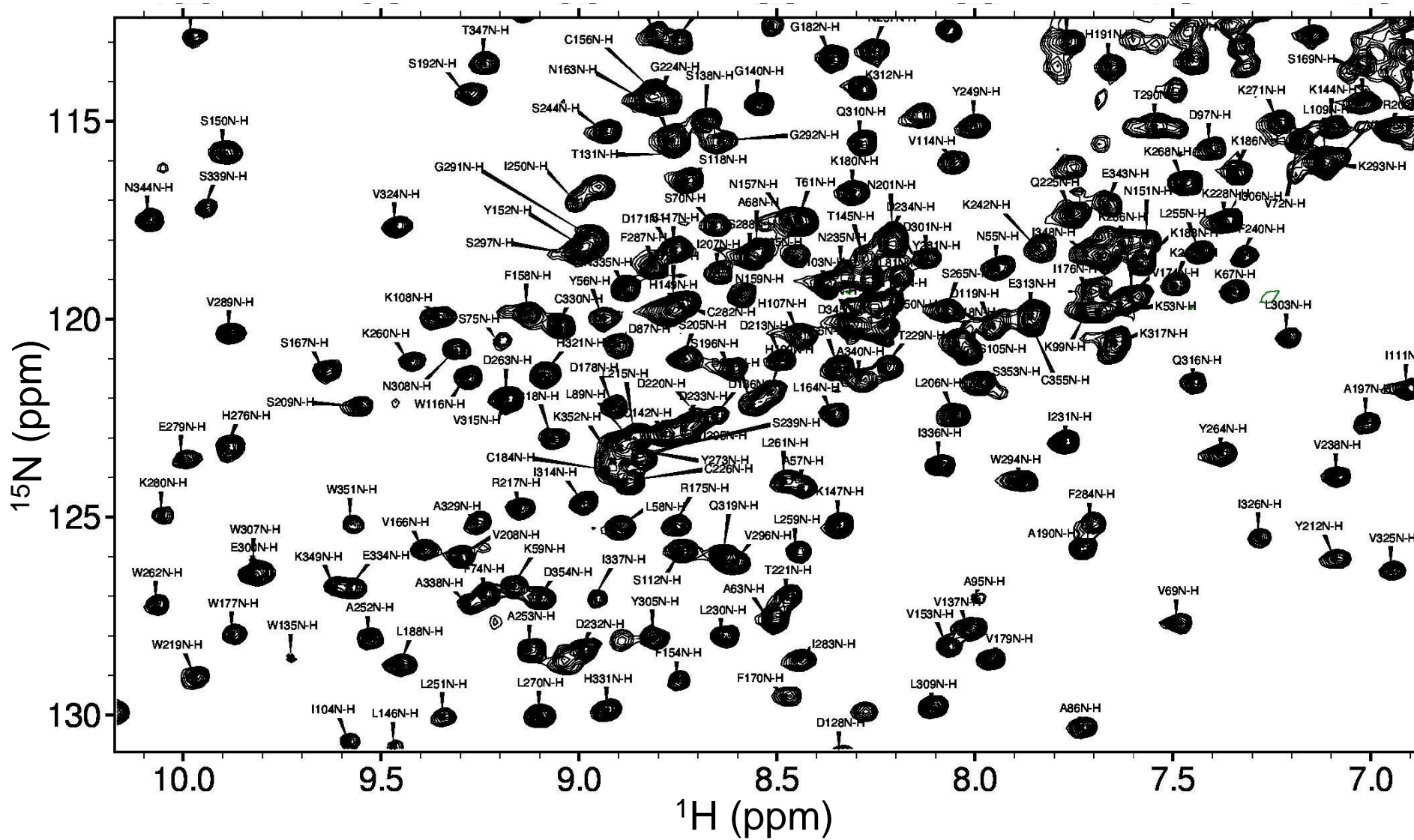


Figure S6.

Figure S6. (related to Figure 3)

Portion of an [^1H - ^{15}N]-TROSY spectrum of ΔN -WDR5.

Peaks are labeled with resonance assignments – 254 backbone amides were assigned in the construct.

1

2

3 **Supplementary Data for:**

4 **MLL1 minimal catalytic complex is a dynamic conformational ensemble susceptible to**
5 **pharmacological allosteric disruption**

6 Lilia Kaustov, Alexander Lemak, Hong Wu, Marco Faini, Scott Houliston, Lixin Fan, Xianyang
7 Fang, Hong Zeng, Shili Duan, Abdellah Allali-Hassani, Masoud Vedadi, Ruedi Aebersold, Yunxing
8 Wang, Cheryl H. Arrowsmith

9

10

1 **Detailed Materials and Methods:**

2 ***Cloning of MLL1, WDR5 and RbBP5 constructs:***

3 The coding regions for human MLL1-SET (aa3785-3969) and MLL1-WIN (aa3745-3969) were
4 PCR-amplified and sub-cloned into the pET28GST-LIC vector (GeneBank ID: EF456739). The mutants,
5 MLL1-SET-7D (Δ 3786-3793) and MLL1-SET_Q3787V/P3788L/Y3379G, were generated from the wild-
6 type clone using the QuikChange PCR mutagenesis kit (Agilent). RbBP5 of different lengths (aa 10-340,
7 320-410, 340-538, 1-538) and WDR5 (aa 24-334) were sub-cloned into pET28-MHL vector (GeneBank ID:
8 EF456738). For expression of dimeric complexes and reconstitution of trimeric MLL1 complex, full-
9 length WDR5 and MLL1-WIN, WDR5 and RbBP5 were cloned into pFastBac Dual expression vector
10 (Thermo Fisher Scientific), respectively, with MLL1-WIN and RbBP5 tagged with N-terminal hexa His-tag.

11 ***Protein expression and purification.***

12 Human MLL1 constructs containing residues 3745-3969 (MLL1-WIN) and 3785-3969 (MLL1-SET),
13 as well as the WD-40 repeat region of WDR5 (residues 24-334), the N- and C-terminus of RbBP5 (RbBP5-
14 NTD, RbBP5-CTD), and full-length RbBP5 were individually expressed in *E. Coli* (BL21(DE3) codon plus RIL,
15 Agilent). The MLL1 proteins were expressed as N-terminal GST fusions and purified on a GST-bind
16 (Novagen) column according to manufacturer's instruction. The GST tag was cleaved off by incubating
17 the resin-bound fusion protein with thrombin (Sigma) at 4 C overnight. The eluted MLL1 proteins were
18 passed through a gel filtration column (Superdex 200, GE Healthcare) pre-equilibrated with 20 mM Tris-
19 HCl (pH 7.5), 500 mM NaCl. WDR5 and RbBP5 proteins were expressed as N-terminal Hexa-His fusions
20 and purified on Nickle-chelating column (GE Healthcare). After elution and removing His-tag by
21 incubation with TEV protease at 4 °C overnight, the proteins were subjected to a gel filtration column
22 Superdex 200, GE Healthcare) pre-equilibrated with 20 mM Tris-HCl (pH 8.0), 500 mM NaCl (for RbBP5)
23 and 20 mM PIPES (pH 6.5), 250 mM NaCl (for WDR5), respectively.

24 The dimeric and trimeric complexes of MLL1 used for SAXS and cross-linking studies were
25 expressed in Sf9 cells. The dimeric complex of WDR5-MLL1-WIN and WDR5-RbBP5 were purified by
26 TALON affinity column (Clontech) followed by a gel filtration column (Superdex 200, GE Healthcare) pre-
27 equilibrated with 20 mM BisTris propane (pH 7.0), 250 mM NaCl. The fractions containing the two
28 dimeric complexes were collected separately and used for SAXS data collection. The two dimeric
29 complexes were mixed and incubated on ice for 2 hours and further purified on a gel filtration column

1 (Superdex 200, GE Healthcare) pre-equilibrated with 20 mM BisTris propane (pH 7.0), 250 mM NaCl.
2 The fractions containing the trimeric complex were collected and used for SAXS data collection and
3 cross-linking experiments.

4 **SAXS data collection and analysis**

5 SAXS measurements were carried out at the beamline 12-ID-B of the Advanced Photon Source,
6 Argonne National Laboratory. The energy of the X-ray beam was 14 Kev (wavelength $\lambda=0.8856 \text{ \AA}$), and
7 two setups (small- and wide- angle X-ray scattering, SAXS and WAXS) were used simultaneously in which
8 the sample to Pilatus 2M detector distance were adjusted to achieve scattering q values of $0.006 < q <$
9 2.6 \AA^{-1} , where $q = (4\pi/\lambda)\sin\theta$, and 2θ is the scattering angle. Thirty two-dimensional images were
10 recorded for each buffer or sample solutions using a flow cell, with the accumulated exposure time of
11 0.8-2 seconds to reduce radiation damage and obtain good statistics. No radiation damage was
12 observed as confirmed by the absence of systematic signal changes in sequentially collected X-ray
13 scattering images. The 2D images were corrected and reduced to 1D scattering profiles using the
14 Matlab software package at the beamlines. The 1D SAXS profiles were grouped by sample and
15 averaged. The scattering profile of the protein was calculated by subtracting the background buffer
16 contribution from the sample-buffer profile using the program PRIMUS (ATSAS package, EMBL) (1).
17 Concentration series measurements for each sample were carried out to remove the scattering
18 contribution due to inter-particle interactions and to extrapolate the data to infinite dilution. The
19 protein concentration ranges used for MLL1-WIN/WDR5, RbBP5/WDR5, and MLL1-WIN/WDR5/RbBP5
20 were 14-28 μM , 10-36 μM and 9.6-46.5 μM , respectively. These concentrations are >10 fold of the
21 dissociation constants for each binary interaction of the dimers. Guinier analysis and the experimental
22 radius of gyration (R_g) estimation from the data of infinite dilution were performed using PRIMUS. The
23 pair distance distribution function (PDDF), $p(r)$, and the maximum dimension of the protein, D_{max} , in real
24 space was calculated with the indirect Fourier transform using program GNOM (2). To avoid under-
25 estimation of the molecular dimension and consequent distortion in low resolution structural
26 reconstruction, the parameter D_{max} , the upper end of distance r , was chosen such that the resulting
27 PDDF has a short, near zero-values tail at large r . The R_g from $P(r)$ analysis was also reported. The
28 Volume of correlation (3), V_c , was calculated using in-house script. The molecular weights were
29 estimated using V_c (3) in q range of $0 < q < 0.3 \text{ \AA}^{-1}$. Fifteen *ab-initio* shape reconstructions (molecular
30 envelopes) were generated using DAMMIF (4) and averaged with DAMAVER (5). The structural models

1 were superimposed and overlaid with the averaged envelop using SUPCOMP (6). The theoretical
2 scattering intensity of a structural model was calculated and fitted to the experimental scattering
3 intensity using CRY SOL (7) and FoXS (8) programs.

4 ***Chemical cross-linking mass spectrometry***

5 The reconstituted trimer complex of WDR5, RbBP5 and MLL1-SET was cross-linked at a
6 concentration between 12 and 16 μ M with 1 mM of isotopically coded disuccinimidyl suberate (DSS-
7 d_0 , DSS- d_{12}) for 30 minutes at 37 °C while shaking at 500 rpm on a Thermomixer (Eppendorf) as
8 previously described (9). Samples were quenched with 50 mM NH_4HCO_3 for 20 minutes at 37 °C and
9 evaporated to dryness in a vacuum centrifuge. The dried pellets were then dissolved in 50 μ L of 8 M
10 urea, reduced with 2.5 mM Tris(2-carboxyethyl)phosphine (Pierce) for 30 minutes at 37 °C and alkylated
11 with 5 mM iodoacetamide (Sigma-Aldrich) for 30 minutes at room temperature, in the dark. Digestion
12 was carried out after diluting urea to 5 M with 50 mM NH_4HCO_3 and adding 1% (w/w) LysC protease
13 (Wako Chemicals) for 2 hours at 37 °C and subsequently diluting to 1 M urea with 50 mM NH_4HCO_3 and
14 further adding 2% (w/w) trypsin (Promega) for 14 hours at 37 °C. Protein digestion was stopped by
15 acidification with 1% (v/v) formic acid. Digested peptides were purified using Sep-Pak C18 cartridges
16 (Waters) according to the manufacturer's protocol. Cross-linked peptides were enriched by peptide size-
17 exclusion chromatography (SEC) as previously described (9). SEC fractions were then reconstituted in 5%
18 acetonitrile and 0.1% formic acid and analysed in duplicates on a LC (Easy-nLC 300) coupled to a mass
19 spectrometer (Orbitrap LTQ XL). Analytes were separated on self- packed New Objective PicoFrit
20 columns (11 cm x 0.075 mm I.D.) containing Magic C18 material (Michrom, 3 μ m particle size, 200 Å
21 pore size) over a 60-min gradient from 7% to 35% acetonitrile at a flow rate of 300 nL/min. The mass
22 spectrometer was operated in data-dependent acquisition (DDA) mode with MS acquisition in the
23 Orbitrap analyzer at 60,000 resolution and MS/MS acquisition in the linear ion trap at normal resolution
24 after collision-induced dissociation. DDA was set up to select up to five most abundant precursors with a
25 charge state of +3 or higher (9). MS data were converted to mzXML format with msConvert (10) and
26 searched with xQuest/xProphet (11) against a database containing the FASTA sequences of the analysed
27 proteins and relative decoy sequences. Cross-linked peptides were identified with a minimal length of 5
28 amino acids and at least four bond cleavages or three adjacent ones per peptide. Validated cross-linked
29 peptides had a total ion current of total ion current explained higher than 0.1 and xQuest score higher
30 than 20. Figures were prepared with xiNET (12).

1 **Structural characterization using SAXS data**

2 The SAXS data indicate that the trimeric complex and its sub-complexes, as well as individual
3 molecules MLL1-SET and RbBP5, are flexible molecular systems in solution. Thus we take an ensemble
4 approach for structural characterization of these systems by utilizing SES protocol (13). The strategy on
5 which SES method is based consists of two main steps: 1) generate the initial ensemble of
6 conformations in order to approximate the conformational space available for a system in solution; 2)
7 find optimal weight w_k for each conformation k from the initial ensemble that minimizes discrepancy

$$8 \quad \chi^2 = \sum_{i=1}^{N_q} \left[\frac{\sum_{k=1}^{N_{ens}} I_{calc}^k(q_i) \cdot w_k - I_{exp}(q_i)}{\sigma(q_i)} \right]^2$$

9 were $I_{exp}(q)$ is the experimental scattering intensity, N_q is number of experimental points, $\sigma(q)$ is
10 the experimental error, and $I_{calc}^k(q)$ is scattering intensity predicted for k th conformation, and of N_{ens}
11 is number of conformations in the initial ensemble.

12 Multi-orthogonal matching pursuit (13) is used to find possible ensembles on step 2, and optimal
13 ensemble size was select using l -curve. The optimal weights w_k were then obtained by averaging over
14 top solutions with similar χ^2 .

15 **Generation of the structural ensembles**

16 *MLL1-SET*. The high degree of flexibility observed for MLL1-SET sample originate from inherent
17 flexibility of SET domain and 28 aa long disordered N-terminal tail. We used all-atom molecular
18 dynamics simulations to generate initial ensemble of conformers. We used all atom MD simulations to
19 generate a trajectory started from the known crystal structure of the MLL1 SET domain with the
20 cofactor product AdoHcy (14) (PDB id: 2W5Y). After minimization and equilibration a productive run
21 was continued for 70 ns. Theoretical scattering profiles in the q range $0 < q < 0.3 \text{ \AA}^{-1}$ for 7,000 frames
22 taken from the trajectory were calculated using CRY SOL.

23 *RbBP5-NTD*. The homology model of the N-terminal domain of RBBP5, RbBP5₂₄₋₃₄₀, generated by
24 ROBETTA (15), was used as starting structure for all atom MD simulations of RbBP5-NTD domain. MD

1 trajectory of 20ns was generated and theoretical scattering profiles in the q range $0 < q < 0.3 \text{ \AA}^{-1}$ for
2 2000 frames taken from the trajectory were calculated using FoXs (8). The domain keeps its structure
3 along the trajectory within 3.5 \AA of backbone r.m.s.d. to the initial homology model. The calculated
4 scattering curves were averaged over the entire ensemble of structures using the optimal weights for
5 each ensemble member obtained with SES method, and this average profile was compared with the
6 experimental scattering data.

7 *RbBP5*. The initial ensemble for SES analysis of full length RbBP5 was generated using RANCH (16)
8 program using homology model for the WD40 domain and assuming the N-terminal RbBP5₁₋₂₄ and C-
9 terminal RbBP5₃₂₅₋₅₃₈ regions to be disordered. The ensemble consists of 20,000 models with random
10 conformation of the flexible regions. Theoretical scattering profile for each member of the ensemble
11 was calculated in the q range $0 < q < 0.3 \text{ \AA}^{-1}$ using CRY SOL.

12 *WDR5/MLL1-WIN*. A pool of possible conformations of the dimeric complex was generated
13 assuming the interaction between WDR5 and WIN peptide of MLL1 and utilizing known crystal structure
14 of the WDR5/WIN complex (17)(PDB id: 3EMH). 30,000 random configurations of rigid SET domain
15 tethered to WDR5 by WIN motif and flexible linker consisting of 46 amino acids, MLL1₃₇₇₁₋₃₈₁₇, was
16 generated using RANCH. Theoretical scattering profiles were calculated in the q range $0 < q < 0.25 \text{ \AA}^{-1}$
17 using CRY SOL.

18 *WDR5/RbBP5*. We used the model of full length RbBP5 described above to generate ensemble
19 of possible conformations of the WDR5/RbBP5 complex with RANCH. We assumed that RbBP5 interacts
20 with WDR5 by WBM motif as in the known crystal structure of sub-complex WDR5/WBM (18)(PDB
21 ID:2XL2), so that the WD40 domain of RbBP5 and the WDR5/WBM sub-complex, both considered to be
22 rigid in the simulations, are connected by flexible linker consisting of 48 residues, RbBP5₃₂₅₋₃₇₂.
23 Theoretical scattering profile for each of 30,000 generated random conformations of the WDR5/RbBP5
24 complex was calculated in the q range $0 < q < 0.20 \text{ \AA}^{-1}$ using CRY SOL.

25 *WDR5/RbBP5/MLL1-WIN*. The initial ensemble of possible conformations for the trimeric
26 complex was obtained in three steps. (i) On the first step, a rigid-body modeling of the complex using
27 CORAL (19) was performed. The known ordered regions of the complex, which are assumed to be rigid
28 on this step of simulations, consist of the known structures of MLL1 SET domain (14), sub-complex
29 WIN/WDR5/WBM (20)(PDB id: 3P4F), and a homology model of WD40 domain of RbBP5. The rest of the
30 complex segments (~31% of all residues) are assumed to be flexible and are modeled by chains of

1 dummy residues. The experimental inter-domain cross-links data were taken into account in the
2 CORAL calculations by introducing six C_{α} - C_{α} distance restraints (see Table S1) with upper bound of 30 Å.
3 CORAL tries to build a single conformation of the complex that fits SAXS data under the imposed
4 constraints. Performing multiple CORAL runs we generated a number of different conformations of the
5 complex that fit SAXS data with $\chi_{SAXS} \sim 0.9$. Although the obtained conformations have different inter-
6 domain arrangements the relative position of the SET domain and WD40 domain of RbBP5 is well
7 defined and suggests the interaction of these domains in the complex. (ii) On the second step we
8 “refined” the best CORAL models by carrying out all-atom molecular dynamic simulations. The initial
9 conformation for MD simulations was constructed from CORAL model by building an all-atom
10 reconstruction model using PULCHRA (21). A 20 ns MD trajectory was generated at $T = 300$ K. (iii) On
11 the third step, we used coarse-grained MD simulations to generate a pool of possible conformations of
12 trimetric complex that are consistent with known inter-molecular binary interactions and cross-links.
13 The structures from the step 2 were used to derive native contact map of quasi-rigid regions of the
14 complex, which determines the nonbonded part of the Go-like potential. The quasi-rigid regions include
15 residues WDR5₃₈₋₃₃₀, RbBP5₂₉₋₃₂₀, MLL1₃₈₁₆₋₃₉₆₉, MLL1₃₇₆₁₋₃₇₆₇, MLL1₃₇₈₅₋₃₇₉₂, and RbBP5₃₇₄₋₃₇₉ that
16 correspond to the WD40 domains of WDR5 and RbBP5, SET domain of MLL1, WIN motif, RBS, and WBM,
17 respectively. We have generated eight 300ns MD trajectories at $T=300$ K. Then 24,000 conformations
18 were saved and used as initial ensemble for fitting to SAXS data by SES method. Theoretical scattering
19 profiles for each conformation in the ensemble were calculated in the q range $0 < q < 0.23 \text{ \AA}^{-1}$ using FoXS.

20 ***All-atom molecular dynamics simulations***

21 A modified Generalized Born implicit solvent model (22) was exploited in the MD simulations in
22 order to accelerate sampling of the conformational space for each of the systems. All simulations used
23 an integration step of 2 fs with fixed bonds between hydrogen atoms and heavy atoms. Temperature
24 was controlled by carrying out Langevin dynamics with damping coefficient set to 2 ps^{-1} . The cut-off
25 for non-bonded Lennard-Jones and electrostatic interactions were set to 18 Å. Ionic strength was set to
26 0.15M. All simulations were performed using NAMD 2.9 code (23) with the AMBER Parm99SB parameter
27 set (24). For residues that coordinate Zn ions a Zinc AMBER Force Field (25) was used.

28 ***Coarse-grained molecular dynamics simulations***

29 We used a coarse-grained model of RbBP5/MLL1-WIN/WDR5 protein complex in order to
30 enhance the sampling efficiency in the conformational space of the complex. In this model, amino acid

1 residues in the proteins are represented as single beads located at their C_α positions and interacting via
2 appropriate bonding, bending, torsion-angle, and non-bonding potential. A Gō-like model of Clementi
3 and Onuchic (26) was employed to maintain the structured, globular domains as quasi-rigid in the
4 simulation. For flexible regions, we adopt simple model in which adjacent amino acids beads are joined
5 together into a polymer chain by means of virtual bond and angle interactions with a quadratic potential.

$$6 \quad V_b = K_b(b - b_0)^2 ; \quad V_\alpha = K_\alpha(\alpha - \alpha_0)^2$$

7 with the constants $K_b = 50 \text{ kcal/mol}$ and $K_\alpha = 1.75 \text{ kcal/mol}$ and the equilibrium values $b_0 = 3.8 \text{ \AA}$ and
8 $\alpha_0 = 112^\circ$ for bonds and angles, respectively. The excluded volume between nonbonded beads was
9 treated with pure repulsive potential

$$10 \quad V_R = \varepsilon_R(\sigma_R / r_{ij})^{12}$$

11 where r_{ij} is the inter-bead distance, $\sigma_R = 4 \text{ \AA}$, and $\varepsilon_R = 2.0 \text{ kcal/mol}$.

12 The interaction between quasi-rigid domains is modeled with the residue-specific pair interaction
13 potentials that combine short-range interactions with the long-range electrostatics as it described (27,
14 28). The short-range interaction is given by a Lennard-Jones 12-10-6 -type potential and simple Debye-
15 Hückel-type potential is used for the electrostatics interaction (27). In this study we used the dielectric
16 constant of 80 and the Debye screening length of 10 \AA , which corresponds to a salt concentration of
17 about 100 mM.

18 To account for the experimentally observed cross-links we introduced used in the force field additional
19 distance restraints term given by a potential

$$20 \quad V_{XL} = \sum_{k=1}^{N_{XL}} \delta_{\xi(t)}^k V_l^k ; \quad V_l^k = \begin{cases} K_l(l_k - l_0)^2, & l \geq l_0 \\ 0, & l < l_0 \end{cases}$$

21 Here sum is over all cross-links, N_{XL} is the number of cross-links, l_k is C_α - C_α distance for residues
22 involved in k-th cross-link, $l_0 = 32 \text{ \AA}$ is upper bound, $K_{XL} = 10 \text{ kcal/mol}$ is force constant, δ_k^i is
23 Kronecker delta, and $\xi(t)$ is random digital number selected from the interval $[1, N_{XL}]$. $\xi(t)$ is a
24 number that is randomly changed every $\tau_{XL} = 10 \text{ ns}$ during the MD simulation.

1 In-house software was developed and used for performing constant temperature molecular
2 dynamics simulations of the coarse-grained model described above. Andersen method (29) was used to
3 control temperature.

4 ***NMR Spectroscopy.***

5 All TROSY NMR spectra were collected at 25°C on Bruker Avance(II) 800 MHz spectrometer
6 equipped with ¹H/¹³C/¹⁵N cryoprobe. All NMR samples were prepared at pH 7.7 with 20 mM TRIS, 250
7 mM NaCl, 2mM β-mercaptoethanol, 2 mM DTT, 1 mM PMSF. The final NMR samples contained 5% D₂O
8 with a protein concentration ranging between 100 μM and 350 μM. The spectra were processed with
9 NMR Pipe software and analyzed with SPARKY (T.D. Goddard and D.G. Kneller, SPARKY 3, University of
10 California, San Francisco). ¹H-¹⁵N-ΔN-WDR5 backbone chemical shifts were assigned using ABACUS
11 approach(30) combined with manual analysis using NMR data collected at high resolution from
12 nonlinearly sampled spectra (HNCO, CBCA(CO)NH, HBHA(CO)NH, HNCA, HNCACB), and processed using
13 multidimensional decomposition(31, 32). Figures were prepared using PyMol (DeLano Scientific).

14 ***NMR Titration Experiments.***

15 NMR samples were prepared in buffer containing 20 mM TRIS pH 7.4, 150 mM NaCl, 2mM DTT,
16 1 mM TCEP, 0.5 mM PMSF and 5% D₂O. Aliquots of MLL1-WIN (GSARAEVHLRKS) and RbBP5
17 (EDEEVDVTSV) peptides were titrated into the labeled WDR5 protein in molar ratios of 1:1, 1:3, 1:5 and
18 1:7, until no further changes in chemical shifts were detected in the ¹H-¹⁵N TROSY spectra. OICR-9429
19 added to WDR5 in molar ratios 1:1 and 1:2. The weighted chemical shift perturbations were calculated
20 using following formula: $\Delta\text{ppm}=[(\delta_{\text{NH}})^2+(\delta_{\text{N}}/5)^2]^{1/2}$.

21 ***Gel Filtration experiments.***

22 A calibrated Superdex 200 column was equilibrated with 20 mM Tris pH 7.7, 150 mM NaCl, 10uM ZnCl₂,
23 5mM β-mercaptoethanol, 5 mM DTT, 1 mM phenylmethylsulphonyl fluoride (PMSF). For Fig 6A, WDR5,
24 MLL1-WIN and RbBP5-FL proteins were combined in 2:1:2 molar ratios with protein concentrations 12
25 μM, 6 μM and 13 μM, respectively and loaded onto the Superdex 200 column (Figs 6A navy blue trace
26 and S3B). For OICR-9429 competition studies (Fig 6A, cyan), the column was also pre-equilibrated with
27 5-fold excess of OICR9429 and the WDR5/MLL1-WIN/RbBP5-FL complex was pre-incubated with OICR-
28 9429 at 1:5 molar ratio.

1 ***GST Pull-down experiments***

2 Recombinant purified MLL1-GST proteins were incubated with RbBP5 fragments at a 1:2 molar
3 ratio in an assay buffer containing 20mM Tris pH 7.7, 150 mM NaCl, 10 μ M ZnCl₂, 5mM β -
4 mercaptoethanol, 5 mM DTT, 1 mM PMSF at 4 °C for 1 hour, followed by incubation with 100 μ L of
5 glutathione-Sepharose beads (GE Healthcare) for an additional 1 hour. Protein concentrations were 19-
6 30 μ M, 9-19 μ M, and 10-15 μ M for RbBP5fl, RbBP5-NTD, and MLL wild type/mutant constructs,
7 respectively. The mixture was transferred to a micro-column and extensively washed with assay buffer.
8 Bound proteins were eluted with 30 mM reduced glutathione and detected by SDS-PAGE and Coomassie
9 staining.

10 ***Histone methyltransferase assay:***

11 Activity assays were performed in 50 mM Tris-HCl, pH 8.0, 5 mM DTT and 0.01% Triton X-100,
12 using 5 μ M ³H-SAM and 5 μ M Biotin-H3 1-25. Increasing concentrations of RbBP5 were added to 200
13 nM of MLL1-WDR5 complexes (with either wild-type or mutant MLL1). All reactions were incubated for
14 90 minutes at room temperature and SPA method was used to determine the activities. Experiments
15 were performed in triplicate. To test the effect of OICR9429 on MLL1 complex, increasing concentration
16 of OICR9429 was incubated with 200 nM MLL1-WDR5 complexes for 20 min before adding 400 nM
17 RbBP5. The activity of the complex was measured as above.

18

19 **SI references:**

- 20 1. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI. **ATSAS 2.1, a program package for small-**
21 **angle scattering data analysis.** *Journal of Applied Crystallography.* 2006 Apr;39:277-86. PubMed PMID:
22 **WOS:000235991200021.**
- 23 2. Feigin LA, Svergun DI. **Structure Analysis by Small-Angle X-Ray and Neutron Scattering.** N.Y.:
24 **Springer US; 1987.**
- 25 3. Rambo RP, Tainer JA. **Accurate assessment of mass, models and resolution by small-angle**
26 **scattering.** *Nature.* 2013 Apr 25;496(7446):477-81. PubMed PMID: 23619693. Pubmed Central PMCID:
27 **3714217.**
- 28 4. Franke D, Svergun DI. **DAMMIF, a program for rapid ab-initio shape determination in small-**
29 **angle scattering.** *J Appl Crystallogr.* 2009 Apr 01;42(Pt 2):342-6. PubMed PMID: 27630371. Pubmed
30 **Central PMCID: 5023043.**
- 31 5. Volkov VV, Svergun DI. **Uniqueness of ab initio shape determination in small-angle scattering.**
32 *Journal of Applied Crystallography.* 2003;36(3 Part 1):860-4.
- 33 6. Kozin MB, Svergun DI. **Automated matching of high- and low-resolution structural models.**
34 *Journal of Applied Crystallography.* 2001 Feb;34:33-41. PubMed PMID: WOS:000166532800007.

- 1 7. Svergun D, Barberato C, Koch MHJ. CRY SOL - A program to evaluate x-ray solution scattering
2 of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*. 1995
3 Dec;28:768-73. PubMed PMID: WOS:A1995TP80300013.
- 4 8. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and
5 fitting of SAXS profiles. *Nucleic acids research*. 2010 Jul;38(Web Server issue):W540-4. PubMed PMID:
6 20507903. Pubmed Central PMCID: 2896111.
- 7 9. Leitner A, Walzthoeni T, Aebersold R. Lysine-specific chemical cross-linking of protein
8 complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software
9 pipeline. *Nature Protocols*. 2014 Jan;9(1):120-37. PubMed PMID: WOS:000329186400012.
- 10 10. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-
11 platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*. 2012 Oct;30(10):918-
12 20. PubMed PMID: WOS:000309965500011.
- 13 11. Walzthoeni T, Claassen M, Leitner A, Herzog F, Bohn S, Forster F, et al. False discovery rate
14 estimation for cross-linked peptides identified by mass spectrometry. *Nature methods*. 2012
15 Sep;9(9):901-+. PubMed PMID: WOS:000308497800022.
- 16 12. Combe CW, Fischer L, Rappsilber J. xiNET: Cross-link Network Maps With Residue Resolution.
17 *Molecular & Cellular Proteomics*. 2015 Apr;14(4):1137-47. PubMed PMID: WOS:000352194400026.
- 18 13. Berlin K, Castaneda CA, Schneidman-Duhovny D, Sali A, Nava-Tudela A, Fushman D.
19 Recovering a representative conformational ensemble from underdetermined macromolecular
20 structural data. *Journal of the American Chemical Society*. 2013 Nov 06;135(44):16595-609. PubMed
21 PMID: 24093873. Pubmed Central PMCID: 3902174.
- 22 14. Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. Structural basis for the requirement of
23 additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Molecular cell*.
24 2009 Jan 30;33(2):181-91. PubMed PMID: 19187761.
- 25 15. Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, et al. High-resolution comparative
26 modeling with RosettaCM. *Structure*. 2013 Oct 08;21(10):1735-42. PubMed PMID: 24035711. Pubmed
27 Central PMCID: 3811137.
- 28 16. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization
29 of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*. 2007
30 May 02;129(17):5656-64. PubMed PMID: 17411046. Epub 2007/04/07. eng.
- 31 17. Song JJ, Kingston RE. WDR5 interacts with mixed lineage leukemia (MLL) protein via the
32 histone H3-binding pocket. *The Journal of biological chemistry*. 2008 Dec 12;283(50):35258-64.
33 PubMed PMID: 18840606. Pubmed Central PMCID: 2596411.
- 34 18. Odho Z, Southall SM, Wilson JR. Characterization of a novel WDR5-binding site that recruits
35 RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage
36 leukemia protein-1. *The Journal of biological chemistry*. 2010 Oct 22;285(43):32967-76. PubMed PMID:
37 20716525. Pubmed Central PMCID: 2963364.
- 38 19. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New
39 developments in the ATSAS program package for small-angle scattering data analysis. *J Appl*
40 *Crystallogr*. 2012 Apr 01;45(Pt 2):342-50. PubMed PMID: 25484842. Pubmed Central PMCID: 4233345.
- 41 20. Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, et al. Structural and
42 biochemical insights into MLL1 core complex assembly. *Structure*. 2011 Jan 12;19(1):101-8. PubMed
43 PMID: 21220120.
- 44 21. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from
45 reduced representations. *Journal of computational chemistry*. 2008 Jul 15;29(9):1460-5. PubMed
46 PMID: 18196502. Pubmed Central PMCID: 2692024.

- 1 22. Onufriev A, Bashford D, Case DA. Modification of the generalized Born model suitable for
2 macromolecules. *Journal of Physical Chemistry B*. 2000 Apr;104(15):3712-20. PubMed PMID:
3 WOS:000086655600048.
- 4 23. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular
5 dynamics with NAMD. *Journal of computational chemistry*. 2005 Dec;26(16):1781-802. PubMed PMID:
6 16222654. Pubmed Central PMCID: 2486339.
- 7 24. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple
8 Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006 Nov
9 15;65(3):712-25. PubMed PMID: 16981200. Pubmed Central PMCID: 4805110.
- 10 25. Peters MB, Yang Y, Wang B, Fusti-Molnar L, Weaver MN, Merz KM, Jr. Structural Survey of Zinc
11 Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *Journal of chemical
12 theory and computation*. 2010 Sep 14;6(9):2935-47. PubMed PMID: 20856692. Pubmed Central PMCID:
13 2941202.
- 14 26. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the
15 structural details of the transition state ensemble and "en-route" intermediates for protein folding?
16 An investigation for small globular proteins. *Journal of molecular biology*. 2000 May 19;298(5):937-53.
17 PubMed PMID: 10801360. Epub 2000/05/10. eng.
- 18 27. Kim YC, Tang C, Clore GM, Hummer G. Replica exchange simulations of transient encounter
19 complexes in protein-protein association. *Proceedings of the National Academy of Sciences of the
20 United States of America*. 2008 Sep 02;105(35):12855-60. PubMed PMID: 18728193. Pubmed Central
21 PMCID: 2529066.
- 22 28. Kim YC, Hummer G. Coarse-grained models for simulations of multiprotein complexes:
23 application to ubiquitin binding. *Journal of molecular biology*. 2008 Feb;375(5):1416-33. PubMed
24 PMID: WOS:000253098100020.
- 25 29. Andersen HC. MOLECULAR-DYNAMICS SIMULATIONS AT CONSTANT PRESSURE AND-OR
26 TEMPERATURE. *Journal of Chemical Physics*. 1980;72(4):2384-93. PubMed PMID:
27 WOS:A1980JK06800026.
- 28 30. Lemak A, Gutmanas A, Chitayat S, Karra M, Fares C, Sunnerhagen M, et al. A novel strategy for
29 NMR resonance assignment and protein structure determination. *Journal of biomolecular NMR*. 2011
30 Jan;49(1):27-38. PubMed PMID: 21161328. Pubmed Central PMCID: 3715383.
- 31 31. Gutmanas A, Jarvoll P, Orekhov VY, Billeter M. Three-way decomposition of a complete 3D
32 15N-NOESY-HSQC. *Journal of biomolecular NMR*. 2002 Nov;24(3):191-201. PubMed PMID: 12522307.
- 33 32. Orekhov VY, Ibraghimov I, Billeter M. Optimizing resolution in multidimensional NMR by
34 three-way decomposition. *Journal of biomolecular NMR*. 2003 Oct;27(2):165-73. PubMed PMID:
35 12913413.

36

37

38 **Table S1. Experimental inter-protein cross-links^a collected for**
39 **trimeric WDR5/RbBP5/MLL1-WIN complex.**

40

Protein 1	Protein 2

Cross-links used in the modeling:	
MLL1₃₈₂₈	RbBP5₂₈₈
MLL1₃₈₄₆	RbBP5₂₈₈
MLL1₃₈₇₀	RbBP5₂₈₈
MLL1₃₈₇₀	RbBP5₂₄₄
MLL1₃₈₄₆	WDR5₄₆
RbBP5₆₀	WDR5₁₅₉
Cross-links within flexible regions:	
MLL1₃₇₄₉	WDR5₇
MLL1₃₇₄₉	WDR5₇₀
MLL1₃₇₄₉	WDR5₄₆
MLL1₃₇₄₉	RbB5₂₈₈
MLL1₃₇₄₉	RbB5₅₀₅
MLL1₃₇₄₉	RbB5₅₁₇

- 1
- 2 a) Residues in disordered regions of a protein are colored in blue.
- 3