Supplementary Information for

An Integrated Structural Model of the DNA Damage Responsive H3K4Me3 Binding WDR76:SPIN1 Complex with the Nucleosome

Xingyu Liu, Ying Zhang, Zhihui Wen, Yan Hao, Charles A.S. Banks, Joseph Cesare, Saikat Bhattacharya, Shreyas Arvindekar, Jeffrey J. Lange, Yixuan Xie, Benjamin A. Garcia, Brian D. Slaughter, Jay R. Unruh, Shruthi Viswanath, Laurence Florens, Jerry L. Workman, and Michael P. Washburn*

*Corresponding author: mwashburn4@kumc.edu

This PDF file includes:

Supplementary Methods
Figs. S1 to S12
Data S1 to S6

Supplementary Methods

Cell Line Construction

Two guide RNAs were used at the same time to target the following sequences in genome: Guide 1: 5'-GGTCGGGCGCGGCGGCTGAG-3'and Guide 2: 5'-AAGAGGGTGCATTCGTTTGG-3'. Approximately 1 kb upstream of guide 1cutting site and about 3 kb downstream of guide 2 cutting site were cloned as homolog arms. The coding sequence of GFP was inserted between two homolog arms as a positive selection marker. In this design, the complete ORF of WDR76 is removed from genome but the endogenous promoter region is preserved to drive the expression of GFP. An RFP expression cassette is placed in reverse direction of GFP as a positive marker for transfection and a negative marker to rule out clones that have random integration of the donor plasmid. Single cells were sorted by flow cytometry and expanded individually for screening. To screen and validate WDR76-/- clones, two pairs of primers targeting exon2 and exon13 of WDR76 were used for PCR from genomic DNA extracted from each clone.

Cell lines that stably expressing Halo-WDR76 or Halo-SPIN1 have been used in previous publications (1, 2). The generation of stable expression cell lines in WDR76 knockout background was similar to the stable transfection of Flp-In™-293 Cell Line described by manual (Invitrogen). All cell lines were maintained in DMEM medium with GlutaMAX, 10% FBS and 100 ug/mL hygromycin B.

Antibodies

Anti-SPIN1 antibodies (ab118784) are purchased from Abcam (Boston, MA, USA) and applied at 1:2000 dilution for Western blot. Anti-H3K4me3 antibodies were purchased from

EMD Millipore (Burlington, MA USA) and applied at 1:5000 dilution for Western blot. Anti-H3 antibodies (05-499) are purchased from EMD Millipore (Burlington, MA USA) and applied at 1:1000 dilution for Western blot. Anti-H3K9ac antibodies (06-942) are purchased from EMD Millipore (Burlington, MA USA) and applied at 1:500 dilution for Western blot.

Serial Capture Affinity Purification (SCAP)

The optimized SCAP protocol is similar to our previously published manuscript (2). For each replicate, cells were collected from 6 confluent 150mm plates and lysed with High Salt Lysis Buffer. The lysate was centrifuged, and supernatant was collected, adjusted to 300mM of NaCl, and then incubated with SNAP-Capture Magnetic Beads (New England BioLabs, Ipswich, MA) at 4°C for 3hr. Beads were washed. Bound proteins were eluted with Elution buffer (containing PreScission Protease) at 4°C for 1hr. 10% of the eluate was aliquoted as E1 and the rest 90% was subjected to further Halo purification step. The eluate obtained above was then incubated with Magne® HaloTag® Beads (Promega) at 4°C overnight. Unbound supernatant was collected as UB2. Beads were washed. Bound proteins were eluted with Elution buffer (containing TEV protease) at room temperature for 1hr. The eluate was collected as E2. Three replicates were performed.

For histone PTM analysis by western blotting, single-bait expressing cells were collected from 2 confluent 150mm plates for each replicate while SCAP plasmid expressing cells were collected from 3 confluent 850 cm² plastic roller bottles for each replicate. Using similar purification protocols to the ones described above, the purifications of both complexes were performed in parallel with minor adjustments and similar purity was confirmed by SDS-PAGE followed by silver staining. For Western blotting, the input and elution samples were then resolved on the same gel by SDS-PAGE. The gel was then transferred and analyzed with

Western blot using anti-Spindlin1 (Abcam, 1:2000 dilution) and anti-H3K4me3 (EMD Millipore, 1:5000 dilution) for SCAP purifications or anti-Spindlin1 (Abcam, 1:2000 dilution), anti-H3 (EMD Millipore, 1:1000 dilution), and anti-H3K9ac (EMD Millipore, 1:500 dilution) for single-bait purifications. In the SCAP purification, the abundance of the SPIN1 was presented to show that comparable amount of material was used for each purification and loaded on the gel for measuring H3K4me3 pulldown. In the single-bait purification, the abundance of H3 was presented to show the comparable amount of H3 across inputs and bait proteins in measuring H3K9ac pulldown. These experiments were repeated two more times yielding the same results.

Multidimensional Protein Identification Technology

Each digested sample was loaded onto a three-phase column. The column was pulled from capillary (100 µm i.d.) to a 5 µm tip, then packed first with 8 cm of 5 µm C18 RP particles (Aqua, Phenomenex), followed by 3.5 cm of 5 µm Luna SCX (Phenomenex), and last with 2.5 cm of 5 µm Aqua C18. Then the loaded column was washed with Buffer A (5% Acetonitrile, 0.1% Formic Acid) before placed onto instruments. All samples were analyzed using a 10-step MudPIT sequence. Each loaded column was placed in line with an Agilent 1200 quaternary HPLC pump (Palo Alto, CA) and a Velos Orbitrap Elite mass spectrometer (Thermo Scientific, San Jose, CA). MS spray voltage set at 2.5 kV; MS transfer tube temperature set at 275°C; 50 ms MS1 injection time; 1 MS1 microscan; MS1 data acquired in profile mode; 15 MS2 dependent scans; 1 MS2 microscan; and MS2 data acquired in centroid mode. MS1 scans acquired in Orbitrap (OT) at 60000 resolution; full MS1 range acquired from 400 to 1500 m/z; MS1 AGC targets set to 1.00E+06; MS1 charge states between 2-5; MS1 repeat counts of 2; MS1 dynamic exclusion durations of 90 sec; ddMS2 acquired in IT; MS2 collision energy and fragmentation: 35% CID; MS2 AGC targets of 1.00E+05; MS2 max injection times of 150 ms.

Collected MS/MS spectra were searched with the ProLuCID (3) software against a database of 163860 protein sequences combining 81592 non-redundant Homo sapiens proteins (NCBI, 2020-11-23 release), 426 common contaminants, and their corresponding 81930 randomized amino acid sequences. In this manuscript, sequences of Halo tag, SNAP tag, AcTEV protease and PreScission Protease were added to contaminants. All cysteines were considered as fully carboxamidomethylated (+57 Da statically added), while methionine oxidation was searched as a differential modification. The following dynamic modifications were included during analysis: mono-/di-/tri-methylation and acetylation of lysine; mono-/di-methylation of arginine. DTASelect (4) v1.9 and swallow, an in-house developed software, were used to filter ProLuCID search results at given FDRs at the spectrum, peptide, and protein levels. Here, all controlled FDRs were less than 1%. Data sets generated from each experiment were contrasted against their merged data set using Contrast v1.9 and in house developed sandmartin v0.0.1. Our in-house developed software, NSAF7 v0.0.1, was used to generate spectral count-based label free quantitation (5). The contaminants and keratin proteins were eliminated from the total. The dNSAF values generated by NSAF7 were defined in (5). In short, dNSAF stands for normalized spectral abundance factor (NSAF) computed using distributed spectra counts. For plotting the abundance of histone H2A, H2B and H3, distributed spectra among variants of each type of histone were summed. To compare datasets, QSPEC (6) was used to determine whether the abundances of a protein in two samples were statistically significant. Log fold change, FDRs and Z score (Zstatistic) was generated by QSPEC. Only proteins have QSPEC FDR<0.05 and Z score value>2 were considered significantly enriched in one sample over another (or Z score<-2 for proteins enriched in the later sample than in the former one) (**Data S2F**). Full dNSAF tables with QSPEC results can be found in **Data S2A-C**. To compare the Halo-WDR76/SNAP-SPIN1

SCAP-MS data to published Halo-SPIN1/SNAP-SPINDOC SCAP-MS data (2) (**Data S2D**), the new datasets were also analyzed using the same pipeline. The analysis method was the same as described above except for the use of a different database (Homo sapiens proteins, NCBI, 2016-06-10 release). For the post translational modifications of histone H3, the percentages of modified H3K9 were calculated from the spectral counts of peptides containing modified H3K9 and total H3K9-containing peptides (**Data S2E**). To confirm the result, we also calculated the percentages using the intensity of peptides, the results are consistent with the spectral counts based quantification method.

Mass Spectrometry Analysis of Histone Post-Translational Modifications

Total histone extraction, chemical derivatization and trypsin digestion for bottom-up LC/MS analysis were performed as described in previous protocol (7, 8). In brief, cells were lysed with Nuclear Isolation Buffer (15 mM Tris, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, and 250 mM sucrose with 0.2% NP-40 Alternative) with protease inhibitors and stabilizing reagents (DTT, AEBSF, microcystin and sodium butyrate). Nuclei were pelleted and washed with Nuclear Isolation Buffer without NP-40 Alternative. Histones were extracted from the nuclei with 0.2M sulfuric acid and cleaned with TCA precipitation. Purified histone samples were then subjected to chemical derivatization. The lysine residues were derivatized with an anhydride reagent containing acetonitrile and propionic anhydride, and the solution pH was adjusted to 8.0 using ammonium hydroxide. The derivatized histones were then digested with trypsin in 50 mM ammonium bicarbonate buffer at 37 °C overnight. The N-termini of histone peptides were derivatized with the anhydride reagent, and the peptides were desalted with a self-packed C18 stage tip. The purified peptides were then dried in a centrifugal vacuum concentrator. The dried peptides were reconstituted in 0.1% formic acid. For analysis of WDR76

or SPIN1 co-purified histones, Halo purifications were performed as described in the above single-bait affinity purification method section with additionally added sodium butyrate in buffers for inhibition of HDACs. Halo-AP elution samples were processed the same as the purified histone samples after TCA precipitation.

For LC/MS analysis, peptides were separated with an Easy-Spray PepMap Neo C18, 100 Å, 2 µm, 75 µm x 150 mm column (Thermo Fisher Scientific) for the analysis on an Orbitrap Ascend instrument (Thermo Fisher Scientific) coupled with a Vanquish Neo HPLC System (Thermo Fisher Scientific). Buffer A is 0.1% formic acid in water and Buffer B is 0.1% formic acid in acetonitrile. The LC gradient consisted of 2 to 32% Buffer B in 48 min and then 32 to 42% Buffer B in 7min. Data-independent acquisition (DIA) was performed with the following settings. A full MS1 scan from 290 to 1200 m/z was acquired with a resolution of 120k, an automatic gain control (AGC) target of 8×10^5, and a maximum injection time of 50 ms. A series of DIA scans were acquired across the same mass range with sequential isolation windows of 24 m/z with a normalized collision energy of 25%, a Orbitrap resolution of 15k, an AGC target of 5×10^5, and a maximum injection time of 60 ms.

All raw DIA data files were analyzed with in-house developed software EpiProfile (7, 8). Label-free quantification was performed at the MS1 level. In the EpiProfile result output, a total of 186 isomers of 30 histone peptides were quantified, and the ratio of each isomer was also calculated (**DataS3A**). Additionally, a total of 45 single PTM ratios representing the abundance of each PTM over all the peptides containing the same site of lysine were reported by EpiProfile (**DataS3B**). For each histone mark, the single PTM ratio in histones co-purified with SPIN1 or WDR76 was compared to the ratio in total histones extracted from the corresponding cell line. A two-tailed t-test was performed to determine if a histone mark was enriched by Halo-AP over

7

total histones (**DataS3C**). The same t-test was also performed to compare the total histones

extracted from Halo-WDR76 or Halo-SPIN1 stable-expression cell lines used for corresponding

purifications (**DataS3C**). We did not find any differences exceeding 2-fold to be significant at a

0.05 p-value cutoff between the histones extracted from the two cell lines.

Cross Linked Peptide Analysis

Crosslinked peptides were analyzed on an Orbitrap Fusion Lumos mass spectrometer (Thermo

Scientific, San Jose, CA) coupled to a Dionex UltiMate 3000 RSCLnano System. Peptides were

loaded on the Acclaim™ PepMap™ 100 C18 0.3 mm i.d. 5 mm length trap cartridge (Thermo

Scientific, San Jose, CA) with loading pump at 2 µl/min via autosampler. Analytical column

with 50 µm i.d. 150 mm length, was packed in-house with ReproSil-Pur C18-AQ 1.9 µm resin

(Dr. Masch GmbH, Germany). The organic solvent solutions were water/acetonitrile/formic acid

at 95:5:0.1 (v/v/v) for buffer A (pH 2.6), and at 20:80:0.1 (v/v/v) for buffer B. When peptides

were analyzed, the chromatography gradient was a 20 min column equilibration step in 2% B; a

10 min ramp to reach 10% B; 120 min from 10 to 40 % B; 5 min to reach 95% B; a 14 min wash

at 95% B; 1 min to 2% B; followed by a 10 min column re-equilibration step at 2% B. The nano

pump flow rate was at 180 nL/min. An MS3 method was made specifically for the analysis of

DSSO cross-linked peptides. Full MS scans were performed at 60,000 m/z resolution in the

orbitrap with 1.6 m/z isolation window, and the scan range was 375-1500 m/z. Top 3 peptides

with charge state 4 to 8 were selected for MS2 fragmentation with 20% CID energy. MS2 scans

were detected in orbitrap with 30,000 m/z resolution and dynamic exclusion time is 40 s. Among

MS2 fragments, if two peptides with exactly mass difference of 31.9720 with 20 ppm mass

tolerance, both were selected for MS3 fragmentation at CID energy 35% respectively. MS3

scans were performed in the ion trap at rapid scan with isolation window of 3 m/z, maximum ion

injection time was 200 ms. Each MS2 scan was followed by a maximum of 4 MS3 scans. For the analysis of crosslinked samples in this work, a FAIMS pro module (Thermo Scientific, San Jose, CA) was installed, and compensation voltage (CV) was set to alternate between -40, -60 and -80V during acquisition.

Acceptor Photobleaching Förster Resonance Energy Transfer (apFRET)

Data was acquired with a PerkinElmer Life Sciences UltraVIEW VoX spinning disk microscope controlled by Velocity software. The microscope is equipped with Yokogawa CSU-X1 Spinning disk scanner, ORCA-R2 camera (Hamamatsu C10600-10B), EMCCD (Hamamatsu C9100-23B) and PhotoKinesis accessory. The base of the microscope is Carl Zeiss Axiovert 200M. The Dichroic passes 405, 488, 561 and 640nm. HaloTag® Ligands TMRDirect was excited by 561nm laser and collected through 445 (W60), 615 (W70) filter. SNAP-Cell® 505-Star was excited by 488nm laser and collected through 525 (W50) filter. To collect AP-FRET data, time lapse movies were recorded to collect at least 10 timepoints before and after acceptor photobleaching. The movies were recorded at a speed of one image per second. Images were recorded using ORCA-R2, and objective was 40x (Oil, NA=1.3).

For each cell, accepters were bleached by 561nm laser at 100% laser power for 10 cycles. The donor intensity before (I_{Before}) and after (I_{After}) were, separately, averaged over time. FRET efficiency (E) was represented as: $E = 1 - (I_{Before} / I_{After})$. E values were calculated in batch using in house imageJ (National Institutes of Health) plugin (accpb FRET analysis jru v1). For statistically comparing the FRET efficiency between different protein pairs, the measurements for each pair were first tested for normal distribution, and then were compared using two-tailed t-test in OriginPro (OriginLab, Northampton, MA). Measurements and statistical results can be found in **Data S6A**.

Fluorescence Cross-Correlation Spectroscopy (FCCS)

FCCS data was acquired using a ConfoCor 3 with an LSM-510 (Zeiss) microscope. Cells were imaged through a C-Apochromat 40x (NA=1.2) objective. The coverslip thickness correction collar was adjusted to achieve maximum signal in the first Mat-Tek dish. In all cases, the pinholes for all channels were aligned just before acquisition. It was left at that position for the rest of the acquisitions. SNAP-Cell® 505-Star was excited at 488nm, and its fluorescence was collected through a 505-540 bandpass filter. HaloTag® Ligands TMRDirect was excited at 561nm, and its fluorescence was collected through a 580 longpass filter. To prevent crosstalk between the fluorophores, Pulsed Interleaved Excitation (PIE) laser switching was employed. The 488 and 561nm lasers were alternately switched on for 100µs total period (50µs for each laser) with a 2.5µs illumination delay. For each pair of proteins, 10 curves were measured and used for data analysis. Files were analyzed in Fiji (https://fiji.sc/) using in-house written plugins (analysis cross corr jru v2). More detailed results from the analysis can be found in **Data S6B**.

UV-Laser Microirradiation

The UV-laser microirradiation assay was also performed on UltraVIEW VoX spinning disk microscope. The excitation and emission filter settings for each fluorophore have been stated above. Time lapse movies were recorded using ORCA-R2 at speed of one image per second. Objective was 40x (Oil, NA=1.3). The assay has been used in Gilmore et al. (1) to test the recruitment of Halo-WDR76. Similarly, each cell was irradiated in a 1-pixel (~0.17µm) width stripe area near the center of the nucleus. The microirradiation was induced with 405nm laser at 100% laser power (20 cycles). The fluorescence intensities at each timepoint (t) of the stripe (I_t) and the whole nucleus (T_t) were measured cell by cell using imageJ. The timepoint starting microirradiation was defined as t=0. Recruitment (R_t) was defined as: $R_t=(I_t/T_t)/(I_0/T_0)$.

For each experiment, multiple cells were measured and averaged. Rt values were calculated in batch using in-house written plugin (dna damage analysis multicolor jru v1). The highest RT (MaxRT) values were also obtained from the plugin. Detailed values of RT and MaxRT are available in **Data S6C and D**.

To compare the kinetics of the recruitments, the averaged RT values from timepoint 0 to 60 of Halo-WDR76/SNAP-SPIN1 and Halo-WDR76/SNAP-XRCC5 were fitted to a modified Hill function in OriginPro (OriginLab, Northampton, MA):

$$y = START + (END - START)\frac{x^n}{k^n + x^n}$$

Here, START stands for the y value at the beginning of the curve and END stands for the y value when it reaches the maximum and the curve turns flat. k is the Michaelis constant, and n is the Hill coefficient. The time for the recruitment of each protein taken to reach its half MaxRT ($t_{MaxRT/2}$) was used to estimate the rate of recruitment. $t_{MaxRT/2}$ values were obtained either directly from the k values provided by OriginPro or manually measured from the fitted curves. Fitting results can be found in **Data S6E**.

HADDOCK Docking

Inter crosslinks of WDR76 and SPIN1 were used as unambiguous restraints. The center of mass restraints option was switched on for all docking calculations. We used the same sampling and clustering parameters for all analyses. HADDOCK performed three steps for docking. (9) First, randomization and rigid body energy minimization were performed ("it0"). 180 degrees rotated solutions were sampled during the rigid body energy minimization step. Up to 10,000 models were sampled and 1,000 models were saved at this stage. The best 200 models were subjected to semi-flexible refinements ("it1"), then refined in water as the final step ("water"). The refined

models were clustered using the Fraction of Common Contacts (FCC) (10) method with a

fraction of contacts cutoff of 0.6. For the docking of WDR76 model3 and SPIN1 model2,

HADDOCK clustered 127 structures in 13 clusters and reported the best 4 structures for the top

10 clusters in their result page (**Data S5C**). RMSD values were provided in the downloaded

result package from HADDOCK. The average RMSD values were calculated by cluster for

either all the structures or the best 4 structures in each cluster. The rmsd values were computed

by comparing to the best structure in each cluster; the rmsd-Emin values were computed by

comparing to the structure with the overall lowest energy. From the HADDOCK stats, cluster6

has the lowest HADDOCK score and this score is significantly better than the next cluster. The

best 4 structures in cluster6 have a small rmsd of 1.3Å. The Cα-Cα distances between

crosslinked lysine residues in the best 4 structures were measured (**Data S5D**). Using 35 Å as a

cutoff, all 3 crosslinks are satisfied by these 4 structures. Therefore, we picked the structure

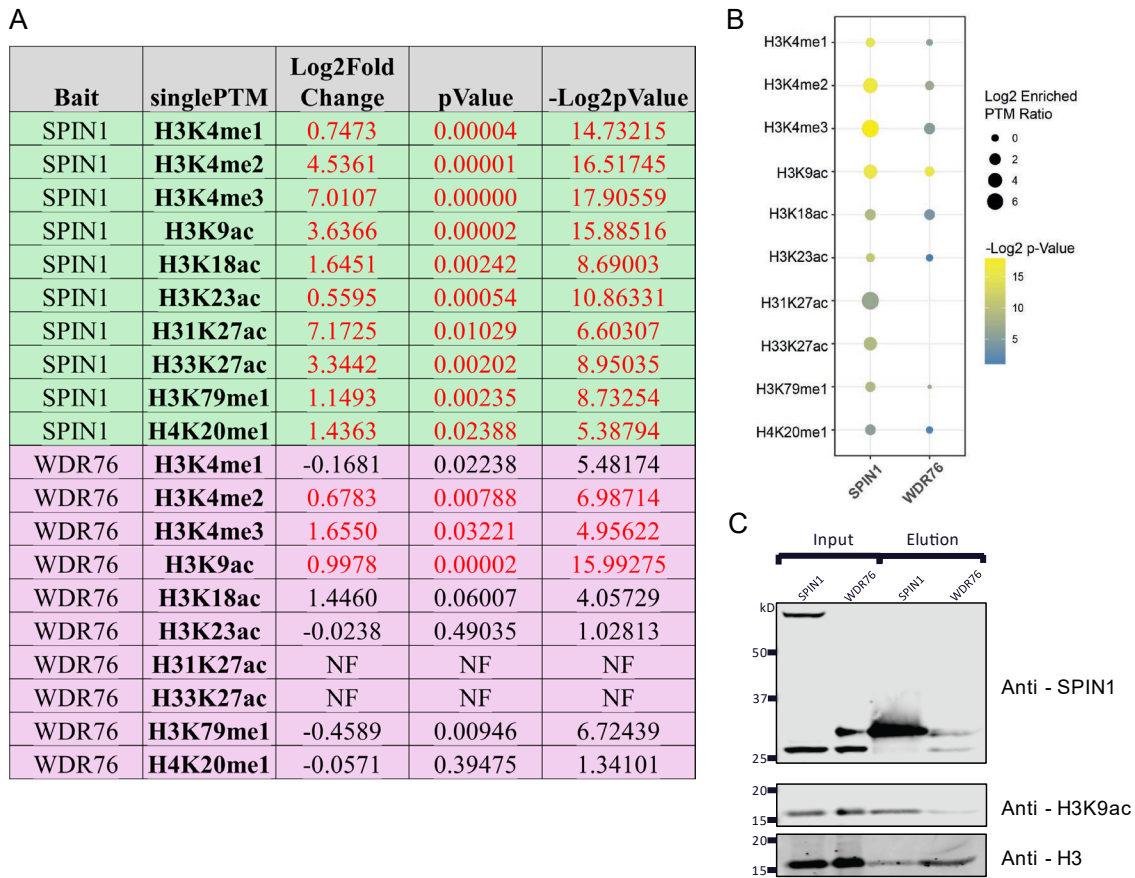cluster6-1 with the lowest HADDOCK score as our best model for visualization.

A

Bait	singlePTM	Log2Fold Change	pValue	-Log2pValue
SPIN1	H3K4me1	0.7473	0.00004	14.73215
SPIN1	H3K4me2	4.5361	0.00001	16.51745
SPIN1	H3K4me3	7.0107	0.00000	17.90559
SPIN1	H3K9ac	3.6366	0.00002	15.88516
SPIN1	H3K18ac	1.6451	0.00242	8.69003
SPIN1	H3K23ac	0.5595	0.00054	10.86331
SPIN1	H31K27ac	7.1725	0.01029	6.60307
SPIN1	H33K27ac	3.3442	0.00202	8.95035
SPIN1	H3K79me1	1.1493	0.00235	8.73254
SPIN1	H4K20me1	1.4363	0.02388	5.38794
WDR76	H3K4me1	-0.1681	0.02238	5.48174
WDR76	H3K4me2	0.6783	0.00788	6.98714
WDR76	H3K4me3	1.6550	0.03221	4.95622
WDR76	H3K9ac	0.9978	0.00002	15.99275
WDR76	H3K18ac	1.4460	0.06007	4.05729
WDR76	H3K23ac	-0.0238	0.49035	1.02813
WDR76	H31K27ac	NF	NF	NF
WDR76	H33K27ac	NF	NF	NF
WDR76	H3K79me1	-0.4589	0.00946	6.72439
WDR76	H4K20me1	-0.0571	0.39475	1.34101

B

C

Fig. S1. Differential enrichment of histone PTMs by WDR76 and SPIN1 (**A**) Table displaying differential enrichment of histone PTMs for each bait compared to the total histone extraction in the respective cell line by LC-DIA-MS analysis and EpiProfile (PTM ratio in Halo elution/ PTM ratio in total histone extraction). Significant enrichment was determined by a two-tailed t-test with three replicates. (**B**) Dot plot displays histone marks significantly enriched by SPIN1 or WDR76 purification. The size of dots represents the fold change of each PTM enriched by Halo purification while the color represents the p-value. Consistent with the table, all marks shown were significantly enriched by Halo-SPIN1 purification at a p-value cutoff of 0.05. Only H3K4me2, H3K4me3, and H3K9ac were significantly enriched by Halo-WDR76 purification. H3K27ac was not detected in Halo-WDR76 co-purified histones. (**C**) Comparison of lysine 9 acetylation on histone H3 copurified by Halo-WDR76 or Halo-SPIN1 purifications by Western blotting. Histone H3 was probed to correct for differences in total histone H3 between purifications. An antibody for tagged and endogenous SPIN1 was probed given it is shared in both purifications as reference. Increased enrichment of H3K9ac despite lower amount of histone H3 was found in the Halo-SPIN1 purification as compared to Halo-WDR76 consistent with histone PTM analysis by LC-DIA-MS and EpiProfile (7, 8).
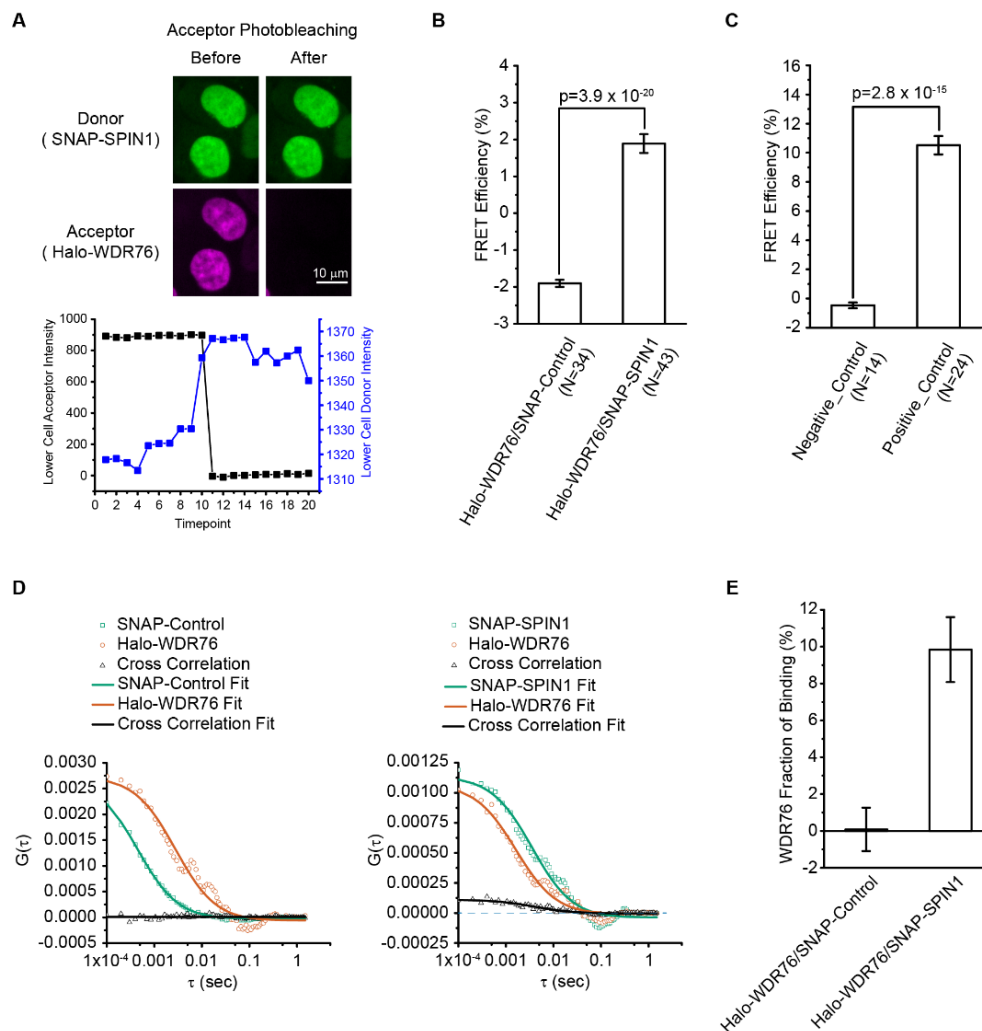
Fig. S2. Imaging approaches to assess the interaction between WDR76 and SPIN1. The interaction between WDR76 and SPIN1 was characterized using Acceptor-Photobleaching Fluorescence Resonance Energy Transfer (apFRET) (**A-C**) and Fluorescence Cross-Correlation Spectroscopy (FCCS) (**D-E**). (**A**) Example image and intensity measurement of apFRET. Halo-WDR76 was labeled with HaloTag TMRDirect ligands and SNAP-SPIN1 was labeled with SNAP-Cell 505-Star ligands. (**B**) Averaged FRET efficiencies measured for Halo-WDR76 and SNAP-SPIN1 in live HEK293FRT cells. SNAP tag was used as donor in the control experiment to be tested with Halo-WDR76. Error bars stand for standard error of mean and p-values were calculated using two-tailed t-test. (**C**) FRET efficiencies measured for control proteins in live HEK293FRT cells. Separate Halo tag and SNAP tag were used as a negative control pair and a fusion protein of Halo tag and SNAP tag was used as positive control. Error bars stand for standard error of mean and p-values were calculated using two-tailed t-test. (**D**) Auto and cross-correlation curves for Halo-WDR76 and SNAP-SPIN1, and for Halo-WDR76 and SNAP tag as control experiment. (**E**) The average percentage of Halo-WDR76 binding to SNAP tag or SNAP-SPIN1 was calculated from the y amplitudes of correlation curves. Error bars stand for standard error of mean. Values was obtained based on the fitted correlation curves, t-test was not performed for this data.
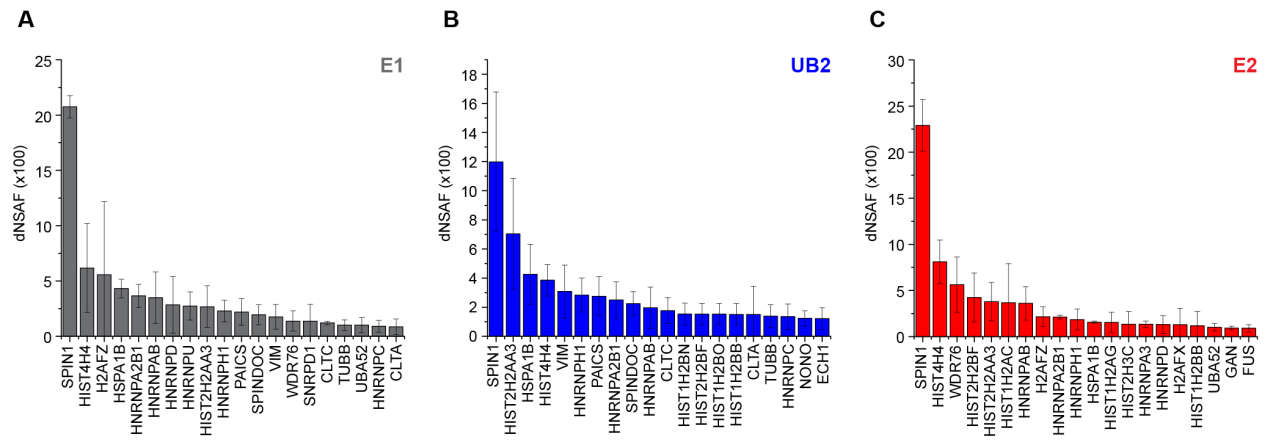
14

Fig. S3. Top 20 proteins ranked by abundance identified by MudPIT in E1 (A), UB2 (B), and E2 (C) fractions of SCAP using Halo-WDR76 and SNAP-SPIN1 as bait proteins. Each dNSAF value plotted is the average from three biological replicates; error bars represent standard deviations.
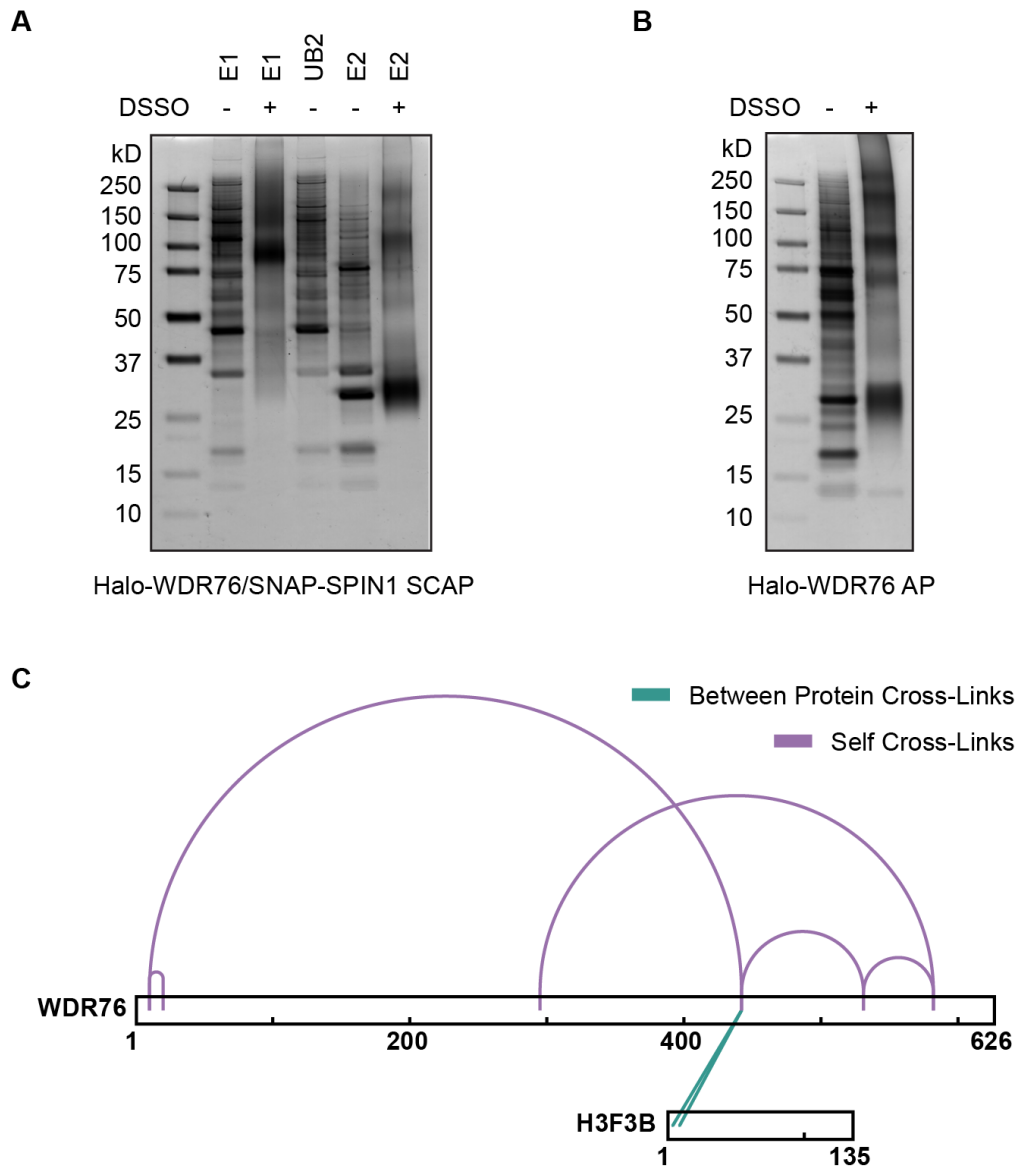
Fig. S4. Gel analysis of DSSO crosslinked samples and crosslinking mass spectrometry analysis of Halo-WDR76 single-bait purification. (**A**) Halo-WDR76/SNAP-SPIN1 SCAP purification and crosslinking. SNAP-SPIN1 was first purified followed by Halo-WDR76. E1 and E2 samples were crosslinked with DSSO. Samples were analyzed by SDS-PAGE gel and silver staining. 5% of crosslinked E2 sample used for mass spectrometry analysis were loaded. (**B**) Halo-WDR76 single-bait affinity purification and crosslinking. Co-purified proteins with or without DSSO crosslinking were analyzed by SDS-PAGE gel and silver staining. 5% of crosslinked sample used for mass spectrometry analysis were loaded. Same amount of protein standards was used in panel **A** and **B**. (**C**) DSSO crosslinked Halo-WDR76 purification samples analyzed by mass spectrometry. Linear view of protein-protein interactions that are linked by at least 2 crosslinks by xiView (11).
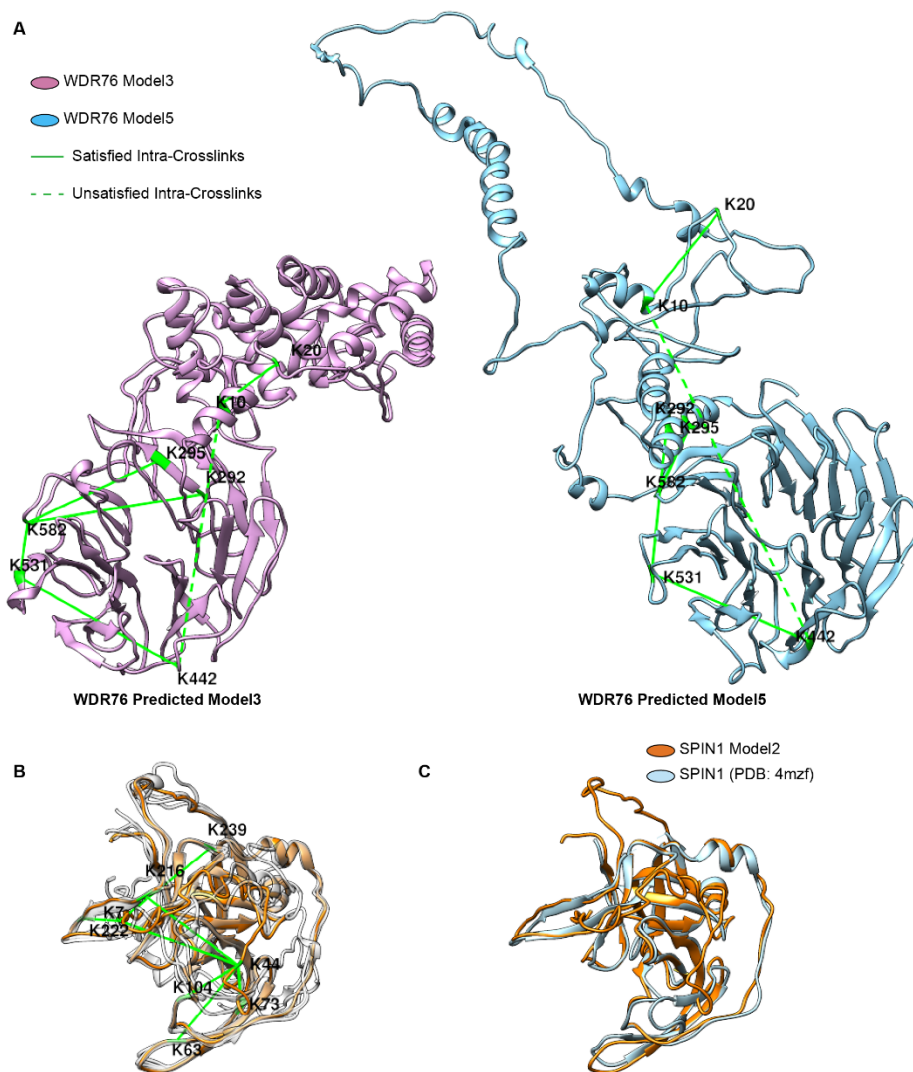
16

Fig. S5. Prediction of full length WDR76 and SPIN1 structures. (**A**) Ab initio prediction of full length WDR76 structure. Structural models were predicted by I-TASSER (12) from amino acid sequence. Identified self-crosslinks of WDR76 were mapped to each model and Cα-Cα distances between crosslinked lysine residues were measured. The best two models, model3 and model5, were displayed with crosslinked lysine labeled and colored in green. Crosslinks are visualized as green lines, while the unsatisfied crosslinks (with Cα-Cα distance between crosslinked sites larger than 35Å) are shown in dash. (**B**) Ab initio prediction of full length SPIN1 structure. Structural models were predicted by I-TASSER (12) from amino acid sequence. Total five models are aligned and displayed in overlays. Model2, which is the best model to satisfy all the identified self-crosslinks of SPIN1, is colored in orange. Crosslinked sites were labeled and colored in green. Crosslinks are visualized as green lines. (**C**) The best predicted full length SPIN1 structure model (colored in orange) is aligned and compared to published SPIN1 structure (chain B of PDB: 4mzf, colored in light blue).
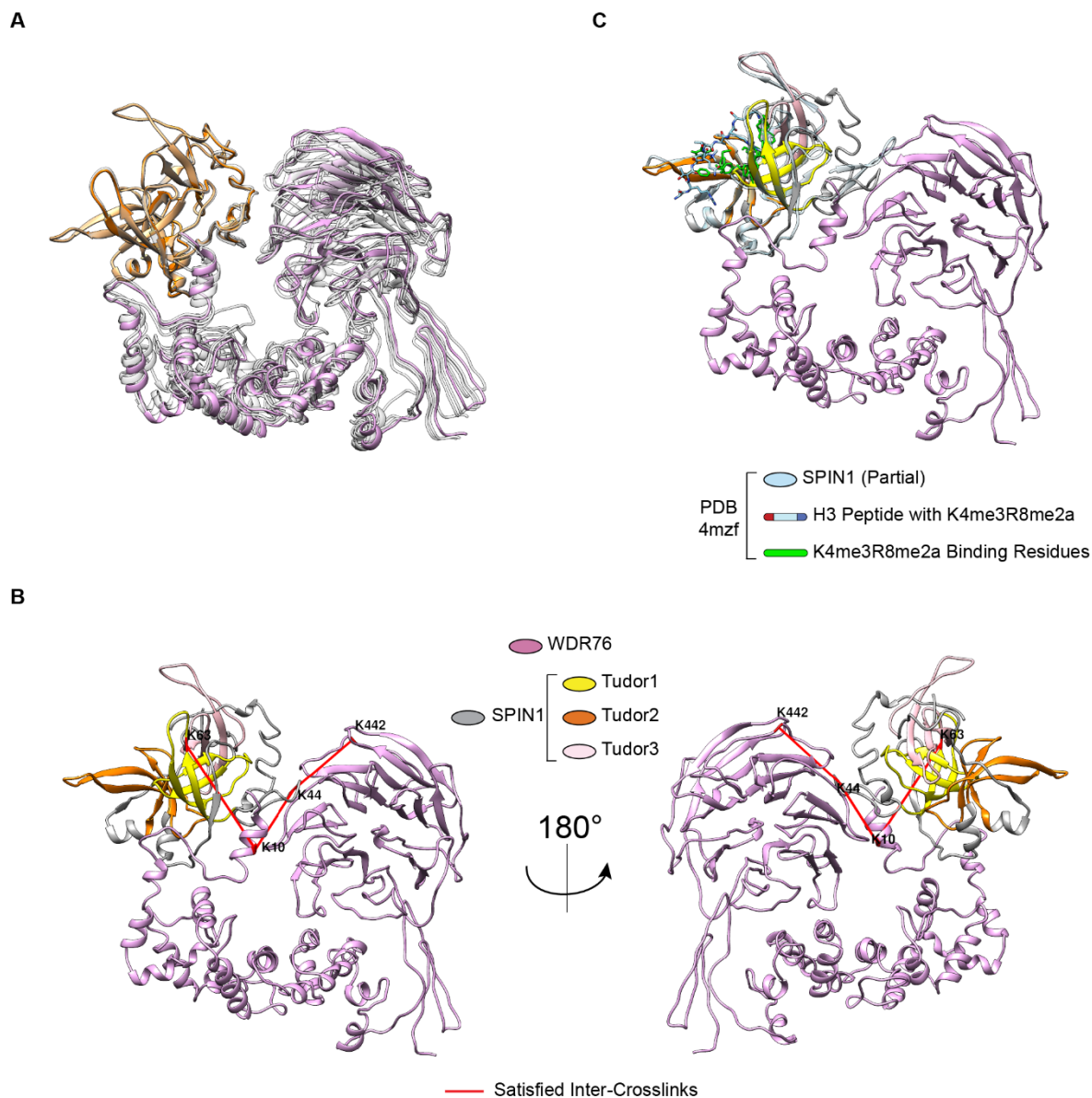
17

Fig. S6. Structure modeling of WDR76:SPIN1 complex. (**A**) Docking of predicted model3 of WDR76 and model2 of SPIN1 with HADDOCK (13). Inter-crosslinks between WDR76 and SPIN1 were used in docking as restraints. An ensemble representation of the best four structures in the top cluster with the lowest HADDOCK score is displayed. The structure in solid color (WDR76 in plum color and SPIN1 in orange) has the best HADDOCK score in the cluster. (**B**) Visualization of crosslinks between WDR76 and SPIN1 in the heterodimer model with the best HADDOCK score. The three Tudor domains of SPIN1 are color coded. Crosslinked sites were labeled and colored in red. Crosslinks are visualized as red lines. (**C**) Visualization of the heterodimer model of WDR76:SPIN1 complex with the best HADDOCK score. A known structure of SPIN1 Tudor domains (light blue) containing a modified histone H3 N'-peptide (PDB: 4mzf) is aligned to the model. Key residues recognizing the modifications of H3 are colored in green.
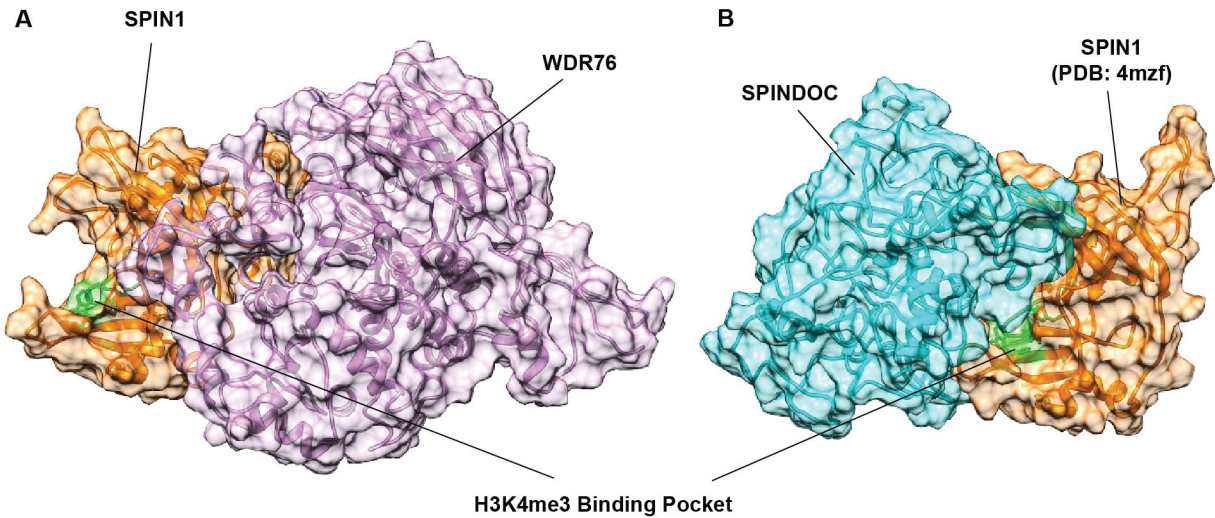
Fig. S7. Comparison of WDR76:SPIN1 and SPINDOC:SPIN1 Models. (A) To compare the structural models of two distinct SPIN1 containing complexes the heterodimer model of WDR76 and SPIN1 with the lowest HADDOCK score is displayed in surface view. WDR76 is shown in plum color and SPIN1 is shown in orange. (**B**) The best heterodimer model of SPIN1 and SPINDOC described by Liu *et al.* (2) is displayed in surface view. SPINDOC is shown in cyan. SPIN1 chain in this model was adopted from PDB: 4mzf and is also colored in orange. In both complex models, the H3K4me3 binding of SPIN1 are marked in green.
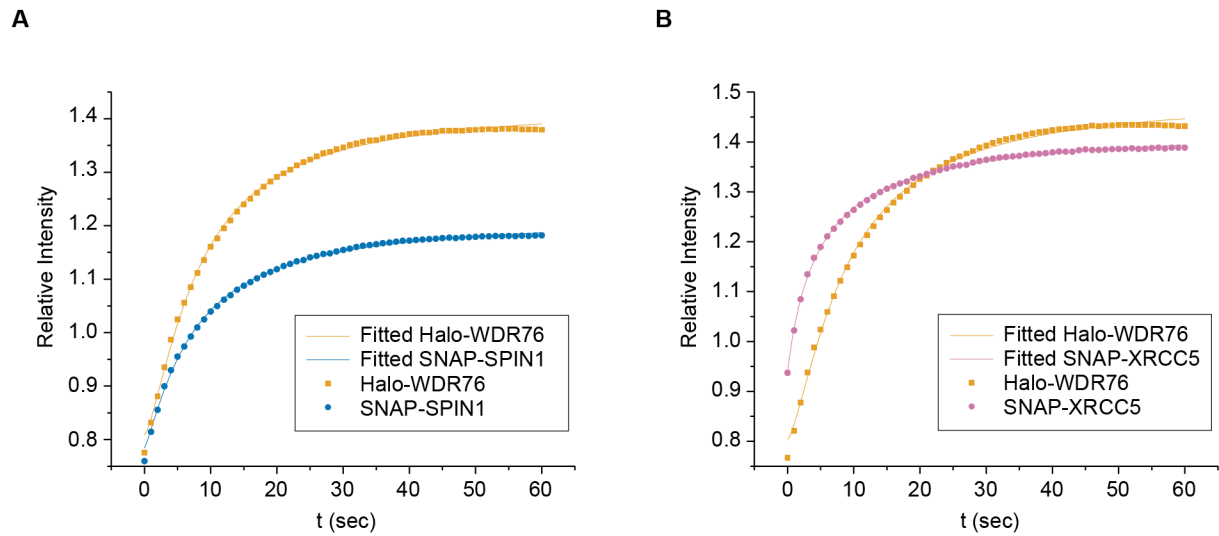
Fig. S8. The co-recruitment of WDR76 and SPIN1 to microirradiated region. The relative intensities of each protein at the laser-irradiated stripe were recorded for each second. The measurements from the moment of microirradiation up to 1 minute were used for nonlinear fitting in OriginPro (OriginLab, Northampton, MA). Each data point is an average from all the measured cells. (**A**) Halo-WDR76 and SNAP-SPIN1 were co-expressed and measured. (**B**) Halo-WDR76 and SNAP-XRCC5 were co-expressed and measured.
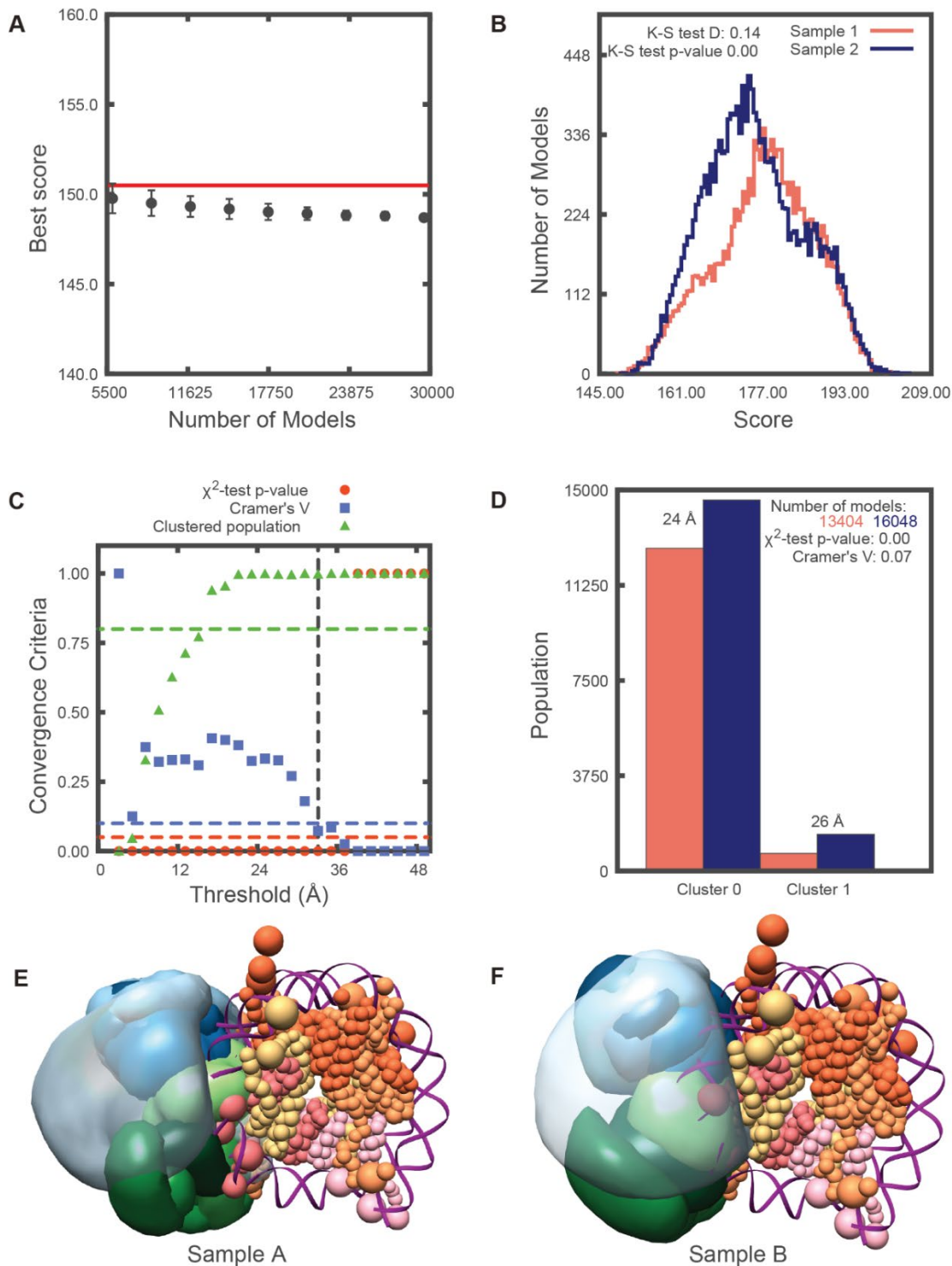
Figure S9. Sampling exhaustiveness protocol on WDR76:SPIN1:Nucleosome models (**A**) Results of test 1, convergence of the model score, for the 29,550 good-scoring models; the scores do not continue to improve as more models are computed essentially independently. The error bar represents the standard deviations of the best scores, estimated by repeating sampling of models 10 times. The red dotted line indicates a lower bound reference on the total score. (**B**) Results of test 2, testing similarity of model score distributions between samples 1 (red) and 2 (blue); the

difference in the distribution of scores is significant (Kolmogorov-Smirnov two-sample test p-value less than 0.05) but the magnitude of the difference is small (the Kolmogorov-Smirnov two-sample test statistic D is 0.14); thus, the two score distributions are effectively equal. (**C**) Results of test 3, three criteria for determining the sampling precision (Y-axis), evaluated as a function of the RMSD clustering threshold (X-axis). First, the p-value is computed using the $\chi 2$ -test for homogeneity of proportions (red dots). Second, the effect size for the $\chi 2$-test is quantified by the Cramer's V value (blue squares). Third, the population of models in sufficiently large clusters (containing at least 10 models from each sample) is shown as green triangles. The vertical dotted grey line indicates the RMSD clustering threshold at which three conditions are satisfied (p-value > 0.05 [dotted red line], Cramer's V < 0.10 [dotted blue line], and the population of clustered models > 0.80 [dotted green line]), thus defining the sampling precision of 33 Å. (**D**) Populations of sample 1 and 2 models in the clusters obtained by threshold-based clustering using the RMSD threshold of 33 Å. Cluster precision is shown for each cluster. (**E**) and (**F**) Results of test 4: comparison of localization probability densities of models from sample A and sample B for the major cluster (92% population). The cross-correlation of the density maps of the two samples is greater than 0.97.
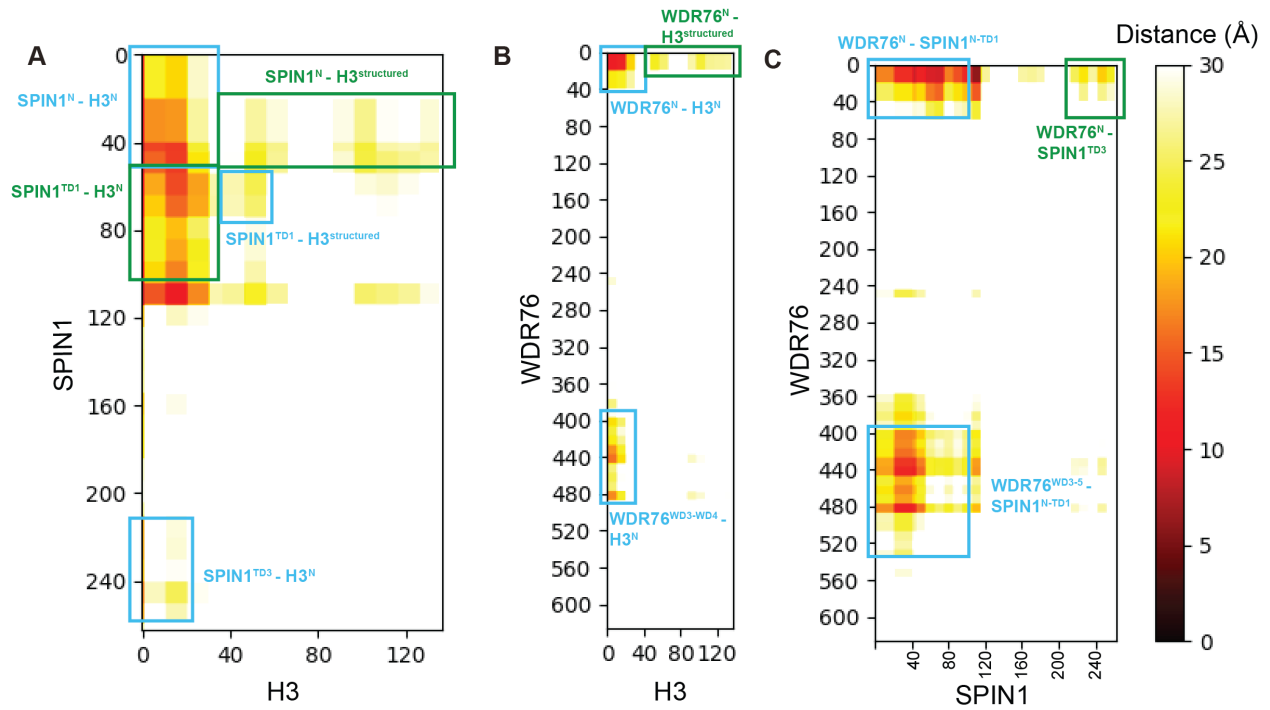
Figure S10. Distance maps. The distance maps show the average pairwise residue distances in the ensemble of WDR76:SPIN1:Nucleosome models for the (**A**) SPIN1-H3, (**B**) WDR76-H3, and (**C**) WDR76-SPIN1 protein pairs. For a pair of residues, the map indicates the distance between the surfaces of the corresponding beads averaged over the ensemble. Prominent interactions are highlighted with rectangles. All the distances above 30 Å are represented with white patches.
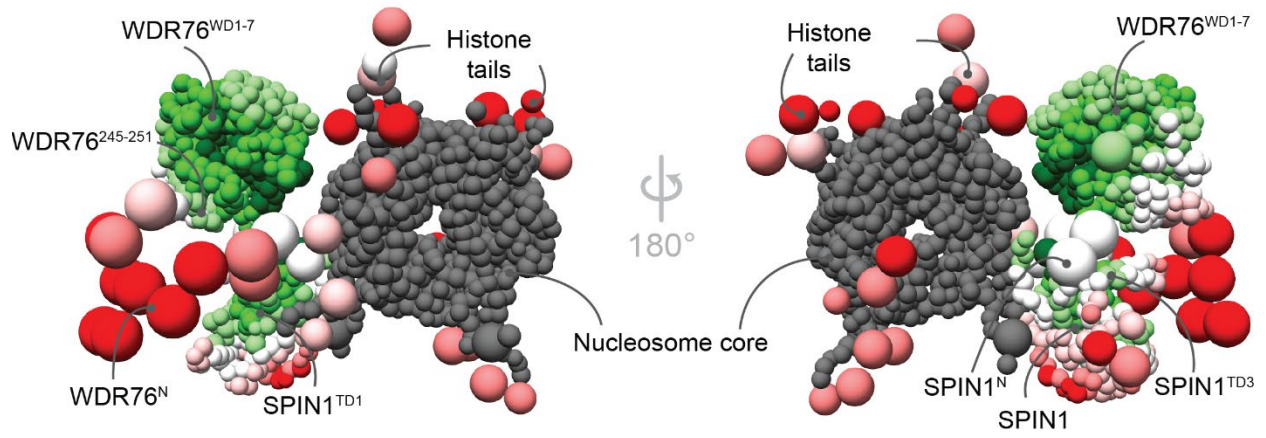
Figure S11. Precision analysis of WDR76:SPIN1:Nucleosome integrative structure Regions of high precision are shown in shades of green while regions of low precision are shown in shades of red.

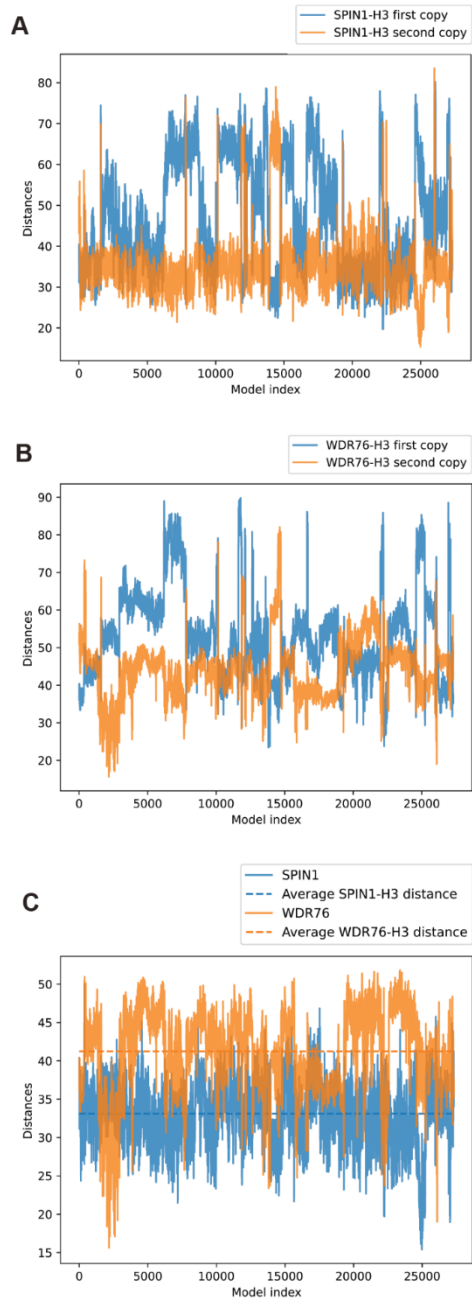**Figure S12. Distances of SPIN1 and WDR76 to H3.**
The following distances are computed for each model in the major cluster. Distance between each molecule of H3 and SPIN1 (**A**) and WDR76 (**B**). (**C**) Distance between SPIN1/WDR76 and the closest H3 molecule to each. All distances were between the center of mass of H3 and the center of mass of the SPIN1/WDR76 domain closest to the corresponding H3 molecule.

Data S1. Information of mass spectrometry datasets. (**A**) Full mass spectrometry data accessibility. (**B**) Spectra and peptide FDR information of MudPIT experiments. (**C**) FDR reports of crosslinking mass spectrometry datasets by Proteome Discoverer 2.4 (Thermo Scientific, San Jose, CA). (Provided as an Excel Worksheet.)

Data S2. Additional tables of MudPIT data. (**A**) Full contrast table of Halo-WDR76 and Halo-SPIN1 AP-MS data. (**B**) Full contrast table of Halo-WDR76/SNAP-SPIN1 SCAP-MS data. (**C**) Full contrast table of Halo-WDR76 AP-MS data generated using cells without endogenous WDR76. The data of E2 fraction in Data S2B was also compared in the same table. In all the contrast tables, peptide counts, spectral counts, sequence coverages, dNSAF values and QSPEC (6) results are included. (**D**) Comparing different SPIN1 containing complexes from SCAP-MS data. Data searched using 2016 database to be compared to published SPIN1:SPINDOC data (2). (**E**) Spectral counts of total and modified peptides of histone H3 containing K9. The percentage values were calculated from spectral count or intensity. (**F**) List of proteins enriched in SCAP E2 fraction over other samples. NCBI RefSeq protein accessions are used here to distinguish protein identities. (Provided as an Excel Worksheet.)

Data S3. Table of EpiProfile Analysis of Histone PTMs from Halo-SPIN1 and Halo-WDR76. (**A**) EpiProfile output table for histone ratios from Halo-WDR76 and Halo-SPIN1 cell total histones and histones associated with affinity purified complexes. (**B**) EpiProfile output table for histone ratios of single posttranslational modifications from Halo-WDR76 and Halo-SPIN1 cell total histones and histones associated with affinity purified complexes. (**C**) Statistical analysis of histone single posttranslational modification ratios from Halo-WDR76 and Halo-SPIN1 cell total histones and histones associated with affinity purified complexes. (Provided as an Excel Worksheet.)

Data S4. Additional tables of XL-MS uploaded to xiView (11) **for visualization and full Halo-WDR76 and SCAP WDR76:SPIN1 dataset.** (**A**) Full table of Halo-WDR76/SNAP-SPIN1 SCAP-XL data. (**B**) Annotations of WDR76 and SPIN1. (**C**) Metadata for of Halo-WDR76/SNAP-SPIN1 SCAP-XL. (**D**) Full table of Halo-WDR76 AP-XLMS data. (**E**) Metadata for Halo-WDR76 AP-XLMS. The protein names provided in metadata are gene names used by Uniprot database. (**F**) Complete XL-MS dataset from Halo-WDR76. (**G**) Complete XL-MS dataset from SCAP WDR76:SPIN1. (Provided as an Excel Worksheet.)

Data S5. Additional data of structural modeling. (**A**) Cα-Cα distances between crosslinked sites measured from predicted WDR76 and SPIN1 models. (**B**) Results of DisVis (14, 15) analysis. (**C**) Parameters, statistic results and cluster RMSD values of HADDOCK analysis. (**C2**) HADDOCK statistics and RMSD values for docking of WDR76 model5 and SPIN1. (**D**) Cα-Cα distances between inter-protein crosslinked sites measured from the top WDR76:SPIN1 heterodimer models. (Provided as an Excel Worksheet.)

Data S6. Additional data of fluorescence imaging experiments. (**A**) Measurements and statistics of apFRET experiments. (**B**) Measurements and fitting of FCCS experiments. (**C**) Measurements of UV laser-induced microirradiation assay. (**D**) Maximum recruitments and

statistics of microirradiation assay. (**D**) Non-linear fitting results of recruitments in the microirradiation assay. (Provided as an Excel Worksheet.)

Table S1: Inputs for Bayesian Integrative Modeling the Integrative Modeling Platform

Protein name	Domain name	Residue range	Atomic structure (PDB or Alphafold ID)	Multi-scale coarse-grained bead representation (number of residues per bead)	Number of copies modeled
SPIN1	N	1-44	-	20	1
		45-53	4h75	1, 10	
	Tudor1 (TD1)	54-103	4h75	1, 10	
	Between Tudor1 and Tudor2	104-132	-	20	
	Tudor2 (TD2)	133-182	4h75	1, 10	
	Between Tudor2 and Tudor3	183-213	-	20	
	Tudor3 (TD3)	214-259	4h75	1, 10	
	C	260-262	-	20	
WDR76	N	1-244	-	20	1
		245-276	AF- Q9H967-F1-model_v4	1, 10	
		277-287	-	20	
	WD40 (WD1)	286-327	AF- Q9H967-F1-model_v4	1, 10	
	WD40 (WD2)	328-392		1, 10	
	WD40 (WD3)	393-438		1, 10	
	WD40 (WD4)	439-485		1, 10	
	WD40 (WD5)	486-532		1, 10	
	WD40 (WD6)	533-581		1, 10	
	WD40 (WD7)	582-621		1, 10	
	C	622-626	-	20	
H2A	N	1-13	-	10	2
	structured	14-119	5gt0	1, 10	
	C	120-130	-	10	
H2B	N	1-27	-	10	2
	structured	28-126	5gt0	1, 10	
H3	N	1-36	-	10	2
	structured	37-136	5gt0	1, 10	
H4	N	1-23	-	10	2
	structured	24-103	5gt0	1, 10	

Table S1. Representation table. Representation of WDR76, SPIN1, and histones used in the integrative structural model of the WDR76:SPIN1:Nucleosome complex. Regions with known structures were modeled as rigid bodies with 1- and 10- residues per bead based on their atomic structure (X-ray or high-confidence AlphaFold prediction), while regions with unknown structures were modeled as flexible strings of beads with 20 (10) residues per bead for WDR76:SPIN1 (histones). The table specifies the nomenclature for protein domains, their residue ranges, the PDB ID or AlphaFold ID of the structure if known, and the number of copies of each protein in the model. (See Table Above.)

Supplemental References

1.   J. M. Gilmore *et al.*, WDR76 Co-Localizes with Heterochromatin Related Proteins and Rapidly Responds to DNA Damage. *PloS one* **11**, e0155492 (2016).
2.   X. Liu *et al.*, Driving integrative structural modeling with serial capture affinity purification. *Proceedings of the National Academy of Sciences* 10.1073/pnas.2007931117, 202007931 (2020).
3.   T. Xu *et al.*, ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *Journal of proteomics* **129**, 16-24 (2015).
4.   D. L. Tabb, W. H. McDonald, J. R. Yates, 3rd, DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of proteome research* **1**, 21-26 (2002).
5.   Y. Zhang, Z. Wen, M. P. Washburn, L. Florens, Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Analytical chemistry* **82**, 2272-2281 (2010).
6.   H. Choi, D. Fermin, A. I. Nesvizhskii, Significance analysis of spectral count data in label-free shotgun proteomics. *Molecular & cellular proteomics : MCP* **7**, 2373-2385 (2008).
7.   S. Sidoli, N. V. Bhanu, K. R. Karch, X. Wang, B. A. Garcia, Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J Vis Exp* 10.3791/54112 (2016).
8.   Z. F. Yuan *et al.*, EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *Journal of proteome research* **17**, 2533-2541 (2018).
9.   C. Dominguez, R. Boelens, A. M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* **125**, 1731-1737 (2003).
10.   J. P. Rodrigues *et al.*, Clustering biomolecular complexes by residue contacts similarity. *Proteins* **80**, 1810-1817 (2012).
11.   M. Graham, C. Combe, L. Kolbowski, J. Rappsilber, xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data. *bioRxiv* 10.1101/561829, 561829 (2019).
12.   J. Yang *et al.*, The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12**, 7-8 (2015).
13.   G. C. P. van Zundert *et al.*, The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725 (2016).
14.   G. C. van Zundert, A. M. Bonvin, DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics* **31**, 3222-3224 (2015).
15.   G. C. van Zundert *et al.*, The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **429**, 399-407 (2017).