communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-07895-2

Structural basis of thymidine-rich DNA recognition by *Drosophila* P75 PWWP domain

Zhaohui Jin[1,5], Zhe Meng[2,5], Yanchao Liu[3,5], Chongyang Li[2], Xuedi Zhang[2], Yue Yin[4], Guanjun Gao[2], Kun Dou[2] ✉ & Ying Huang[1] ✉

Drosophila P75 (dP75), a homolog of the human LEDGF/p75, is crucial for oogenesis by recruiting the histone kinase Jil-1 to euchromatin and impeding H3K9me2 spreading. Like LEDGF, dP75 binds transcriptionally active chromatin, but its precise mechanism remains unclear. Here we show that its PWWP domain prefers binding to thymidine-rich DNA over GC-rich sequences. Crystal structures both in apo and ssDNA-bound states, reveal a domain-swapped homodimer. The aromatic cage, known to recognize histone methyllysine, also engages thymine. Mutations in this cage mimic dP75 knockout phenotypes, including impaired chromatin binding, transposon upregulation, and female sterility. Although dP75 maintains chromatin-bound in H3K36A mutant flies, alterations in the aromatic cage disrupt this localization, underscoring its role in DNA binding. These findings reveal how dP75 targets euchromatin through a PWWP domain that integrates histone reading and nucleotide recognition, advancing our understanding of PWWP domains.

Lens epithelium-derived growth factor (LEDGF or p75), a chromatin-binding protein, is highly conserved from arthropods to mammals. The key role of LEDGF as a cofactor is highlighted in previous studies. It assists in integrating HIV-1 cDNA into human chromosomes by targeting the lentiviral preintegration complex (PIC) to active transcribing regions[1–5]. Additionally, it regulates HOX gene expression in cancers caused by Mixed-linear leukemia 1 (MLL1) fusions[6]. LEDGF, as an epigenetic regulator, comprises the N-terminal PWWP domain and the C-terminal integrase binding domain (IBD)[4,7]. The IBD domain, a helical bundle, exhibits the capacity to form complexes with a variety of proteins, including the HIV integrase, Menin, MLL1[1,7], and kinases that harbor an FxGF motif, such as Pogo transposable element with ZNF domain (POGZ), Cell division cycle-associated 7-like protein (JPO2), ASK (DBF4), and Interacts-with-Spt6 (IWS1)[7].

Drosophila P75 (dP75), the homolog of LEDGF in fly, is essential for female fertility[8]. Structurally, dP75 mirrors LEDGF; it also contains an N-terminal PWWP domain and a C-terminal IBD domain, the latter of which recognizes the FxGF motif of the histone H3S10 kinase Jil-1[8,9]. Therefore, dP75, also known as Jil-1 anchoring and stabilizing protein (JASPER), forms a complex with Jil-1 and stabilizes it in vivo, thereby safeguarding the transcription of genes, including the piRNA pathway components by inhibiting the spread of repressive transcriptional modification H3K9me2[8].

Both dP75 and LEDGF possess a PWWP domain at the N-terminus, which belongs to the 'Royal' family of classical histone readers[10,11]. The PWWP domain is critical for chromatin binding[12–20]. For LEDGF, deletion of the PWWP domain disrupts chromosomal binding during mitosis, impairs HIV-1 replication[4,12], and alters HIV cDNA integration sites in the genome[3,4,19]. The PWWP domain also exists in other epigenetic factors, such as DNA methyltransferase 3 A (Dnmt3a)[10,21], Brpf1[22], MSH6[23], NSD1/2[24,25], NDF[26], and PWWP2A[14,27], where it binds to H3K36me3. However, the affinity between PWWP domain and histone H3K36me3 peptides is relatively weak, within the millimolar range[17,28], questioning the efficacy of PWWP-domain containing proteins' recruitment to chromatin. Numerous studies have documented the structures of PWWP in complexes with dsDNA and nucleosomes, such as the PWWP domain of LEDGF in complex with H3K36me2/3-modified nucleosome complexes[18,29] and HRP3 PWWP with dsDNA oligo[16]. Consequently, the recorded high-

[1]Department of General Surgery, Shanghai Key Laboratory of Biliary Tract Disease Research, State Key Laboratory of Oncogenes and Related Genes, Xinhua Hospital, Shanghai Jiao Tong University, Shanghai, China. [2]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [3]CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China. [4]National Facility for Protein Science in Shanghai, Shanghai Advanced Research Institute, Chinese Academy of Science, Shanghai, China. [5]These authors contributed equally: Zhaohui Jin, Zhe Meng, Yanchao Liu.
✉e-mail: doukun@shanghaitech.edu.cn; huangy@shsmu.edu.cn

affinity interaction between the PWWP domain and DNA could possibly reinforce its hold on chromatin[10,13,16,30,31].

Our previous work showed that the dP75 PWWP domain exhibits a strong preference for DNA over histone H3K36me3[8]. The binding affinity for H3K36me3 exceeds 500 μM, as determined by isothermal titration calorimetry. Unlike other PWWP domains, dP75 PWWP binds with high affinity to thymidine (T)-rich single-stranded DNA (ssDNA) oligos while showing significantly less affinity for adenine (A), cytidine (C), and guanine (G)-rich ssDNAs. Yet, the molecular mechanism underlying the recognition of thymidine by the dP75 PWWP domain and its role in dP75's chromatin binding remains unclear.

In this study, we elucidated the structure and function of the dP75 PWWP domain using both in vitro crystallization and in vivo functional assays. We determined the crystal structures of the dP75 PWWP domain in its apo form, as well as in complex with ssDNA. Unlike previously reported PWWP structures, the dP75 PWWP domain forms a saddle-shaped, domain-swapped homodimer in both in vitro and in vivo contexts. We identified the aromatic cage in the PWWP domain to comfortably accommodate a thymidine nucleotide. Mutations in key residues involved in DNA binding validated the in vivo function of dP75. Notably, dP75 retains its chromatin-binding ability effectively in the absence of histone H3K36me3. Our findings provide valuable insights into the chromatin binding mechanism of dP75, demonstrating its capability to bind to T-rich DNA sequences independently of H3K36me3.

Results

Crystal structure of dP75 PWWP

To elucidate the molecular basis of dP75's chromatin localization, crystallization of the dP75 PWWP domain was performed. The dP75 PWWP domain shares a similar organizational structure with LEDGF, with the notable exception that dP75 lacks the two AT hooks present in LEDGF, which are located in the region adjacent to the C-terminal of the LEDGF PWWP domain (Fig. 1A). The crystal structure of the dP75 PWWP domain reveals a saddle-shaped, domain-swapped homodimer with an interface area of 1504.2 Å^2, indicating stable dimer formation. Each lobe of the homodimer resembles the classical PWWP fold, comprising a five-stranded β-barrel (β1-β5 and a long α-helix αC) (Fig. 1B, C). Additionally, the region that typically forms a loop in other PWWP structures (Fig. 1D), bridging α1 and α2 (residues 68-75), adopts a helical configuration in dP75. This folding extends α1 into α2, resulting in the elongated α-helix (αC) (Fig. 1B, E).

The homodimeric assembly of the dP75 PWWP domain can be attributed to interactions in two crucial regions between molecule A (mol A) and molecule B (mol B) (Fig. 1F). The first region involves intermolecular interactions within loop 1, which links residues 18–26, binding β1 and β2 in both molecules. This interaction includes a hydrogen bond between the carbonyl oxygen of V20 in mol A and the amide nitrogen of G22 in mol B, with a corresponding reciprocal interaction. Additionally, the carbonyl oxygen and amide nitrogen of Y23 in mol A form hydrogen bonds with the same amino acids in mol B. This is supplemented by a cation-π interaction between K74 in mol B and Y23 in mol A and a salt bridge between D56 in mol A and K21 in mol B. Furthermore, we identified a hydrogen bond between the amino group of K74 in mol B and the carbonyl oxygen of G22 in mol A, which strengthens the dimer interface. The second critical region involves the αC helices of both entities, where salt bridges and hydrogen bonds form between E86 in mol B and K64 in mol A, as well as between the hydroxyl group of Y60 in mol A and E86 in mol B. Furthermore, G45 in mol A engages with Q84 and K77 in mol B through hydrogen bonding between its amide nitrogen and carbonyl oxygen.

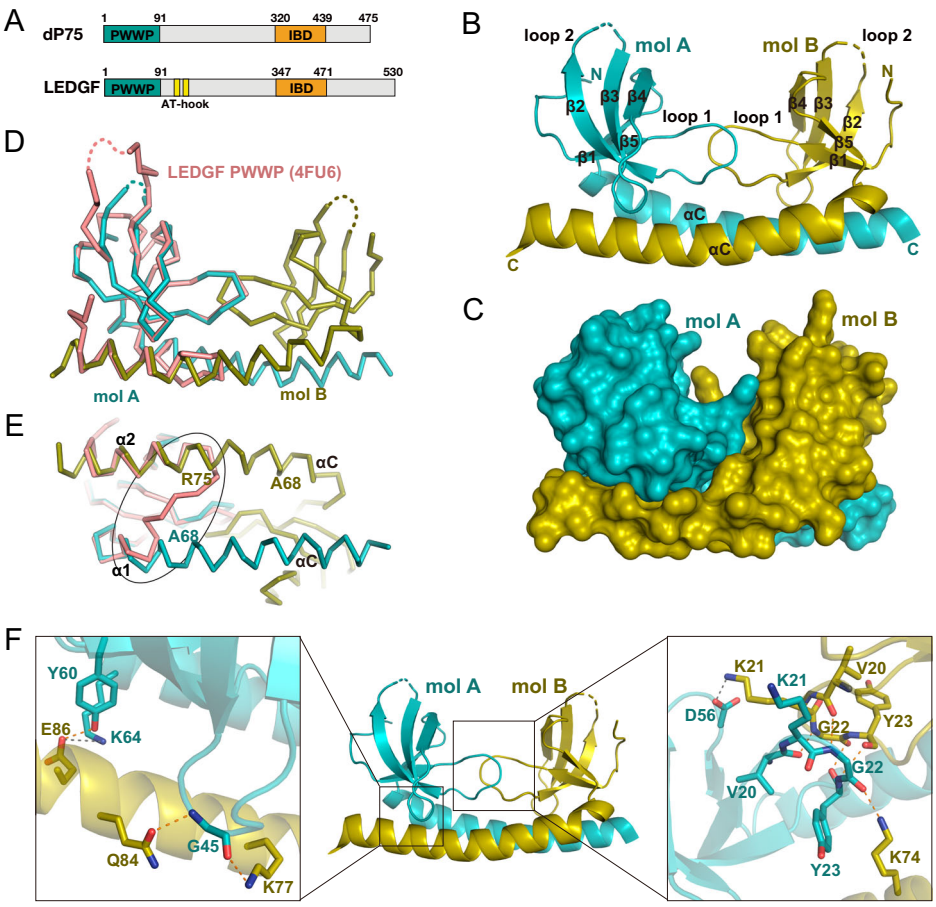


Fig. 1 | Crystal structures of dP75 PWWP domain. **A** Domain organizations of dP75 and LEDGF, respectively. **B** The overall structure of dP75 PWWP in its domain-swapped dimeric conformation. **C** Surface representation of dP75 PWWP dimer. **D, E** Superimposition of the PWWP domains of dP75 and LEDGF (4FU6). The loop 3 connecting α1 and α2 is circled (**E**). **F** Interactions between mol A and mol B. Key residues are shown in stick mode. Detailed interactions are zoomed up in the left and right boxes.

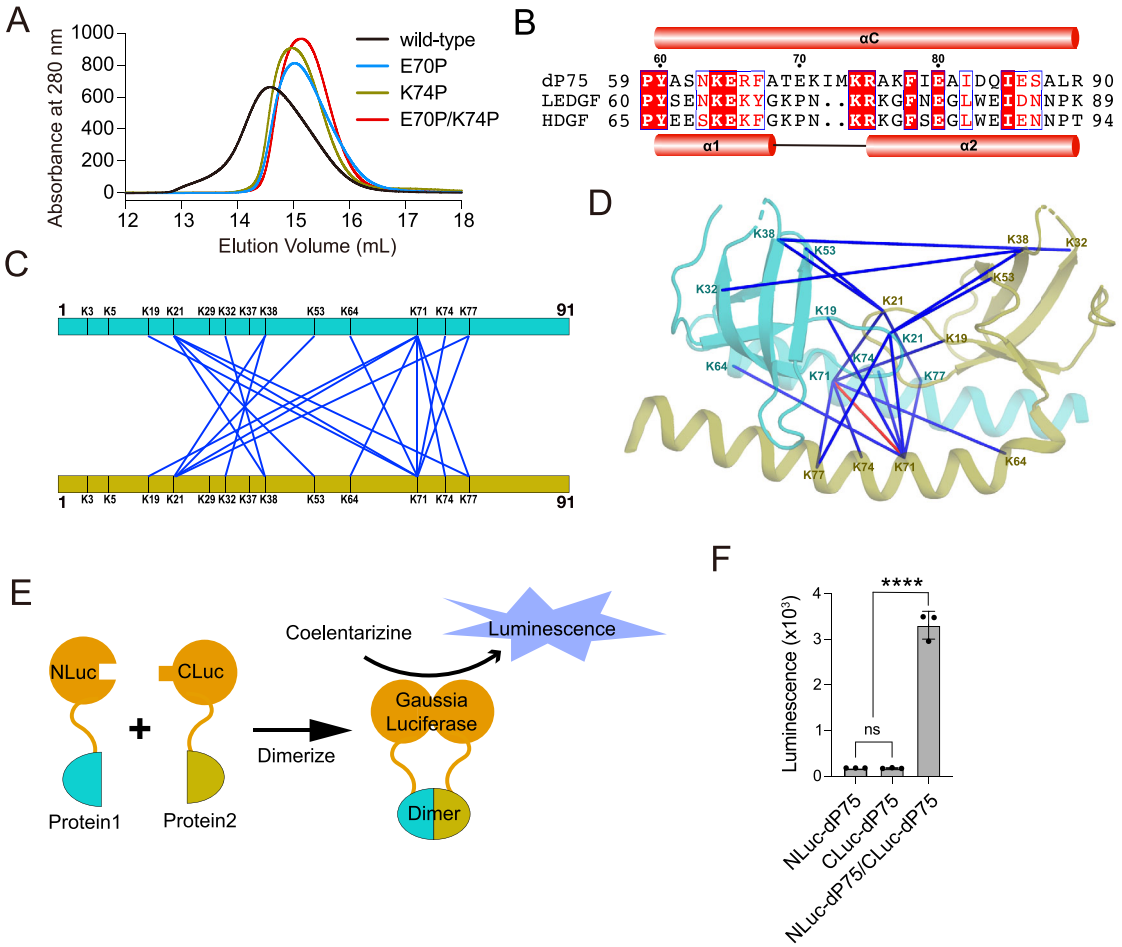


Fig. 2 | dP75 PWWP exists as a dimer. A SEC profiles of wild-type dP75 PWWP, two single mutants E70P and K74P, and the double mutant E70P/K74P. **B** Sequence alignment of dP75, LEDGF, and HDGF C-terminal helices of the PWWP domains. **C** The top 10 lysine pairs within the PWWP domain of dP75 were identified by crosslinking mass spectrometry. **D** Visualization of crosslinks in (**C**) using Xlink Analyzer in Chimera, based on the crystal structure. Blue solid lines indicate the distance between the Cα positions of the lysine residues, except that the distance between the two K71 Cα atoms is shown in red. **E** Schematic illustration for the split-luciferase assay. **F** Fly ovaries harboring transgenes of both NLuc-dP75 and CLuc-dP75 were homogenized and subjected to luminescence detection on the substrate. Transgenic flies with NLuc-dP75, and CLuc-dP75 only were used as controls. $n = 3$ biologically independent experiments. Error bars indicate the standard deviation of the 3 biological replicates. p value was calculated with one-way ANOVA. ****: $p < 0.0001$.

dP75 PWWP exists as a domain-swapped dimer both in solution and in vivo

To rule out the possibility that the formation of the domain-swapped dimer is an artifact of crystal packing, we first performed size-exclusion chromatography (SEC). The SEC elution profile revealed a single peak, indicative of a homogeneous conformation in solution (Fig. 2A). Given that proline residues can disrupt the extension of secondary structures, we generated two single mutants (E70P and K74P) and a double mutant E70P/K74P within the region that typically forms a loop connecting α1 and α2 in other PWWP structures (Fig. 2A, B). We hypothesized that these mutations would disrupt the extended α-helix, resulting in the formation of two short α-helices, as observed in other PWWP structures. The SEC profiles revealed that both single mutants exhibited retention volumes later than the wild-type dP75 PWWP, and the double mutant showed an even later retention volume than the single mutants, confirming a reversion to the monomeric state (Fig. 2A).

Next, we conducted crosslinking mass spectrometry on the full-length dP75 protein using disuccinimidyl suberate (DSS), which can identify two lysine residues close to each other in space. To rule out that dimerization only occurs in the short construct containing the PWWP domain and not in the full-length protein, the full-length dP75 protein was used in the assay (Supplementary Fig. 1 A-D). Top 10 crosslinks with the most spectrum counts were listed (Fig. 2C and Supplementary Table 1). Crosslinks within the DSS effective distance were shown, further supporting the dimerization structure (Fig. 2D). In this conformation, the Cα of the two lysine residues K71 from two molecules are 13.6 Å apart, corroborating the spatial proximity observed in the dimer.

To validate the formation of the dP75 dimer in cells, we employed a developed dimer-reporter system derived from the *Escherichia coli* transcriptional machinery[32]. Specifically, the DNA-binding domain of AraC ($AraC_{DNA}$) cannot activate the transcription of the pBAD promoter unless it is fused with a dimerization domain. However, when fused to a dimerization domain, $AraC_{DNA}$ efficiently activates the pBAD promoter, leading to mTagBFP expression (Supplementary Fig. 2A). We fused $AraC_{DNA}$ with dP75 and examined mTagBFP expression under the pBAD promoter. dP75-$AraC_{DNA}$ constructs induced strong mTagBFP expression compared to control, suggesting dimer formation (Supplementary Fig. 2A, B). Given that the double mutation of E70P and K74P has been shown to disrupt dimer formation, we tested its effect on the dimer reporter. Compared to wild-type dP75, the BFP expression induced by the E70P and K74P mutant significantly decreased (Supplementary Fig. 2B, C), while the protein level of wild-type dP75 and mutant dP75 was comparable (Supplementary Fig. 2D). These results suggest that though E70P and K74P double mutation does not completely block dimer formation, it significantly disrupts dP75 dimer formation.

To further validate dP75 dimer formation in vivo, we carried out a canonical assay for protein-protein-interaction, the split-luciferase assay (Fig. 2E). The N- or C-terminus of luciferase was fused to dP75, and landed

to a specific locus of the genome to generate NLuc-dP75 and CLuc-dP75 transgenic flies, respectively. By mating these flies together, we obtained flies with both NLuc-dP75 and CLuc-dP75, and examined dP75 dimer formation by measuring the luciferase activity on the substrate. Compared to the control group, flies harboring both NLuc-dP75 and CLuc-dP75 exhibited significantly higher luminescence (Fig. 2F), suggesting that the dP75 indeed forms dimer in vivo.

dP75 PWWP prefers to bind thymidine-containing DNA

We superimposed the dimeric dP75 PWWP structure onto the cryo-EM structure of LEDGF PWWP in complex with an H3K36me3-modified nucleosome (PDB 8PC5). In the LEDGF PWWP and H3K36me3-modified nucleosome complex structure, loop 1 is inserted into the minor groove of dsDNA. However, this comparison revealed potential steric clashes that may occur when the dP75 PWWP dimer binds to nucleosome in the same way (Supplementary Fig. 3), suggesting that the dP75 PWWP domain may adopt an alternative binding mode to interact with DNA.

To investigate the DNA-binding characteristics of the dP75 PWWP domain, we performed fluorescence polarization (FP) assays with three double-stranded DNA (dsDNA) oligos: a 12 bp poly(A:T) oligo, a 12 bp poly(G:C) oligo, and a 20 bp random sequence oligo. Our results revealed that the dP75 PWWP domain exhibited strong binding to the 12 bp poly(A:T) sequence with a dissociation constant (K_D) of 18.2 ± 1.17 μM, whereas no binding was observed with the 20 bp random sequence oligo or for the 12 bp poly(G:C) oligo (Fig. 3A).

To further explore binding preferences, we tested the dP75 PWWP domain's affinity for single-stranded DNA (ssDNA) using four oligonucleotides with 5'-FAM labeling comprising 12 consecutive nucleotides of adenosine, thymidine, cytidine, and guanosine, respectively. The K_D values indicated a hierarchy of binding affinities: 24.2 ± 1.1 μM for the polythymidine sequence (FAM-polyT), suggesting the strongest interaction; 150.9 ± 38.9 μM for the poly-cytidine sequence (FAM-polyC), indicating moderate affinity; and negligible interaction with the poly-adenosine (FAM-polyA) and poly-guanosine sequences (FAM-polyG) (Fig. 3B). We also tested the thymidine-rich ssDNA binding ability of LEDGF PWWP. In contrast, LEDGF PWWP does not show any binding to the polyT sequence (Fig. 3B), indicating that the binding to thymidine-rich ssDNA is unique to dP75 PWWP.

The strong binding affinity to the 12 bp poly(A:T) dsDNA oligo, coupled with the high affinity for thymidine-rich ssDNA, suggests that the dP75 PWWP domain prefers thymidine-rich DNA sequences. Given that AT-rich dsDNA may partially unwind due to its low melting temperature[33], exposing thymidine nucleotides, these findings indicate that the dP75 PWWP domain is particularly suited for recognizing thymidine-rich regions within both dsDNA and ssDNA.

The overall structure of PWWP dimer in complex with ssDNA

To delineate the molecular basis for this preferential binding, we conducted co-crystallization experiments with a variety of oligonucleotides. Crystals of the dP75 PWWP domain in complex with a 5-mer ssDNA oligonucleotide bearing the sequence 5'-TCCCT-3' were obtained and diffracted to 2.4 Å (Fig. 4A, B and Supplementary Table 2). There are 16 dP75 PWWP molecules forming eight dimers in one asymmetric unit, and each molecule is bound to one ssDNA oligo.

The structure revealed that two positively charged grooves were sitting along the surface of the 'saddle-like' structure, potentially enhancing its DNA-binding capacity (Fig. 4C). The two oligos fit into the grooves. The formation of these grooves is due to the inclusion of many positively charged residues from both molecule A and molecule B (Fig. 4D). Unexpectedly, a thymidine is intricately nestled within the aromatic pocket, which traditionally serves as the canonical pocket for binding trimethyllysine in histone modification (Fig. 4C). In contrast, we could not find the similar positive charged groove on the surface of LEDGF PWWP (Fig. 4E, F), indicating the unique feature of dP75 PWWP domain-swapped dimer.

Recognition of the thymidine nucleotide by dP75 PWWP

The interactions between the dP75 PWWP domain and thymidine primarily involve hydrogen bonds and π-π stacking (Fig. 5A). Specifically, Y23 and F43 engage in π-π stacking to sandwich the thymine base. Additionally, T46 forms a hydrogen bond with the oxygen atom O^4 on the first thymidine (T-1), while the methyl group at the C^5 interacts with W26 hydrophobically. E48 forms another hydrogen bond with the nitrogen atom N^3. Furthermore, the main chain amide groups of N51 and K21 form hydrogen bonds with O^2 and the phosphate group of the second cytidine (C-2), respectively. K38, K53, and K21 from both molecules also interact with the backbone phosphate groups of C-2 and C-3 via electrostatic interactions.

Sequence alignment revealed that the residues involved in thymidine recognition are conserved (Supplementary Fig. 4). We assessed the function of these residues in DNA binding through site-directed mutagenesis. Substitution of Y23 and F43 with alanine completely abolished binding,

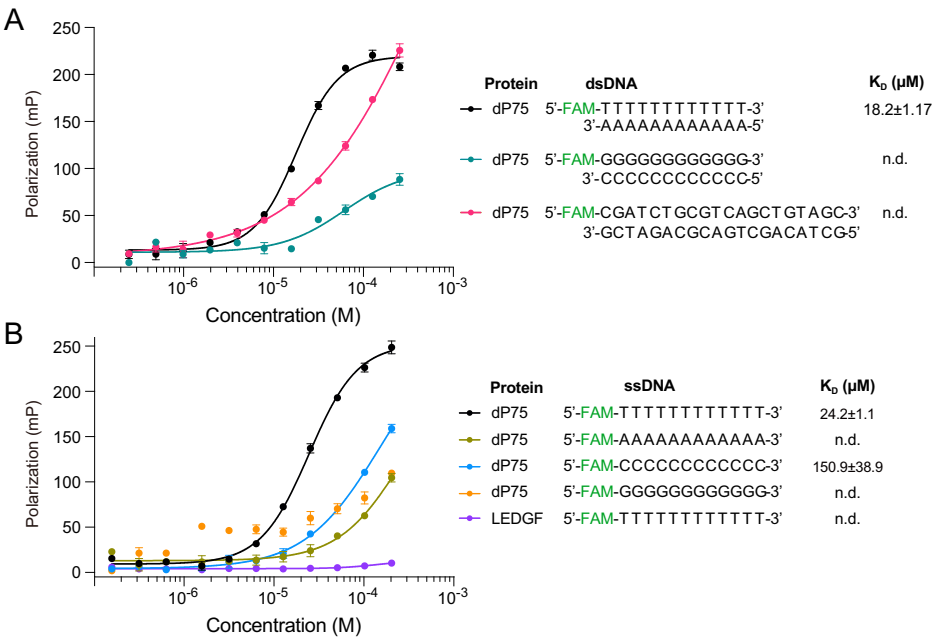


Fig. 3 | dP75 PWWP prefers to bind thymidine-rich DNA. A, **B** The binding affinities of dP75 PWWP to different dsDNA oligos (**A**) and ssDNA oligos (**B**) were measured by FP, respectively. Data are shown as mean ± s.d. ($n = 3$ biologically independent experiments).

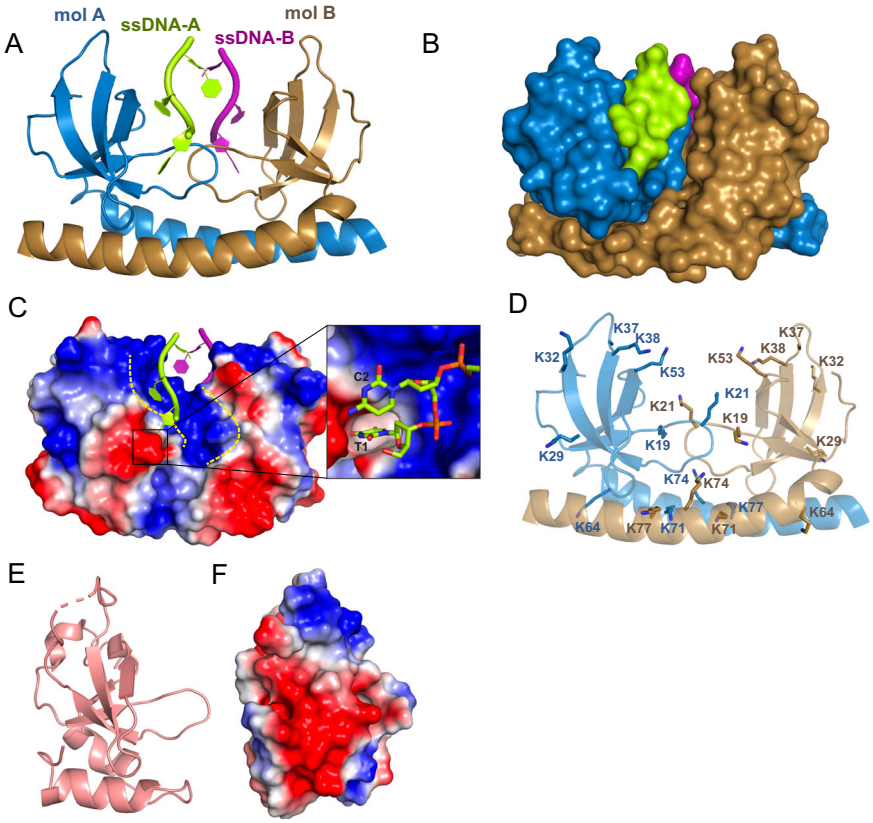


Fig. 4 | Overall structure of dP75 PWWP in complex with ssDNA. A The overall structure of dP75 PWWP in complex with ssDNA. Two ssDNA (green and purple) bind to each side of dP75 PWWP. **B** Surface representation of dP75 PWWP in complex with ssDNA. **C** The electrostatic potential surface of dP75 PWWP. Positively charged concave grooves are formed on each side of the dP75 PWWP. ssDNA binds to the groove. **D** Conformation of lysine residues in dP75 PWWP. (**E**) Cartoon representation of LEDGF PWWP. **F** The electrostatic potential surface of LEDGF PWWP.

highlighting their critical role. Similarly, replacing E48 with alanine significantly reduced the binding affinity ($K_D = 64.2 \pm 2.8$ μM) (Fig. 5B), underscoring the importance of the aromatic cage in recognizing T-containing ssDNA. In the crystal structure, ssDNA oligos were observed to engage within a positively charged groove. To evaluate the impact of electrostatic interactions, we substituted alanine for lysine residues at positions K21, K38, and K53, which interact with the ssDNA backbone. Mutating K21 resulted in a complete loss of binding, while mutations at K38 and K53 moderately decreased the binding strength, with dissociation constants of 60.9 ± 0.9 μM and 108.8 ± 2.9 μM, respectively (Fig. 5C).

Key residues in the PWWP domain are critical for dP75 functionality

To validate the key residues essential for the DNA-binding capacity of the dP75 PWWP domain, we generated transgenic flies with individual alanine substitutions and introduced these into a *dP75* knockout allele, using a wild-type *dP75* transgene as a control. dP75's nuclear localization and expression remain unaffected by point mutations in the PWWP domain (Supplementary Fig. 5). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) revealed that wild-type dP75 binds chromatin with a similar pattern to endogenous dP75 we characterized before[8]. However, mutants with alanine substitutions at Y23, W26, F43, or E48 showed dramatically decreased chromatin-binding capacity (Fig. 6A, B). Additionally, transgenes with these critical mutations failed to restore the sterility phenotype of the *dP75* knockout allele (Fig. 6C). Residues such as K21 and K53, which create the positively charged surfaces for binding the DNA backbone, are crucial for dP75's function in oogenesis, as indicated by the fertility assay observed in these mutants (Fig. 6C).

Next, we investigated whether dP75's chromatin-binding capacity is essential for restraining H3K9me2. In wild-type fly ovaries, H3K9me2 immunofluorescence showed limited bright foci in the nucleus, while the *dP75* knockout allele exhibited a broader deposition of H3K9me2[8]. Reintroducing wild-type *dP75* into the knockout mutant fully rescued the H3K9me2 spreading phenotype (Fig. 6D). In contrast, mutants with key residues Y23, W26, F43, and E48, though expressed at a comparable level as wild-type dP75 transgene, failed to restore this phenotype (Fig. 6D and Supplementary Fig. 6A-C).

To clarify the role of chromatin binding, we examined its involvement in transposon silencing. Previous studies, including ours, have shown that dP75 silences transposons in female gonads, with depletion in germ cells leading to transposon activation[8,34]. RNA fluorescence in situ hybridization (FISH) demonstrated elevated transposon mRNA levels in nurse cell nuclei and oocyte cytoplasm in the absence of dP75. Due to piRNA-mediated repression, transposon mRNAs were barely detectable in the control group (Fig. 6E). The wild-type *dP75* transgene completely rescued the transposon derepression phenotype of the null mutant, while point mutations impairing DNA-binding capacity failed to restore it (Fig. 6E). Our previous results showed that dP75 depletion leads to decreased expression of a piRNA component, BoYb, which may contribute to the transposon-activation phenotype. We thus examined *boYb* transcript level to determine whether the point mutants affect BoYb expression. In comparison to wild-type dP75, which restored the *boYb* transcript by more than 50%, the point mutants failed to restore *boYb* transcript (Supplementary Fig. 6D). These results are consistent with the transposon-elevation phenotype in the point-mutated flies, further confirming that the DNA-binding capacity is essential for dP75 to regulate its targets. Collectively, these findings identify that the key residues in the PWWP domain are crucial for chromatin binding, H3K9me2 restraining, transposon silencing, and female fertility.

The PWWP domain is critical for dP75's chromatin binding in the absence of H3K36me3

To further assess whether dP75 preferentially recognizes H3K36me3 or DNA in vivo, we utilized homozygous H3K36A *Drosophila* mutants, which lack H3K36 methylation[35]. This allows us to examine dP75's chromatin binding in the absence of H3K36me3. Validation of the H3K36A larvae genotype was conducted and documented (Supplementary Fig. 7A-C).

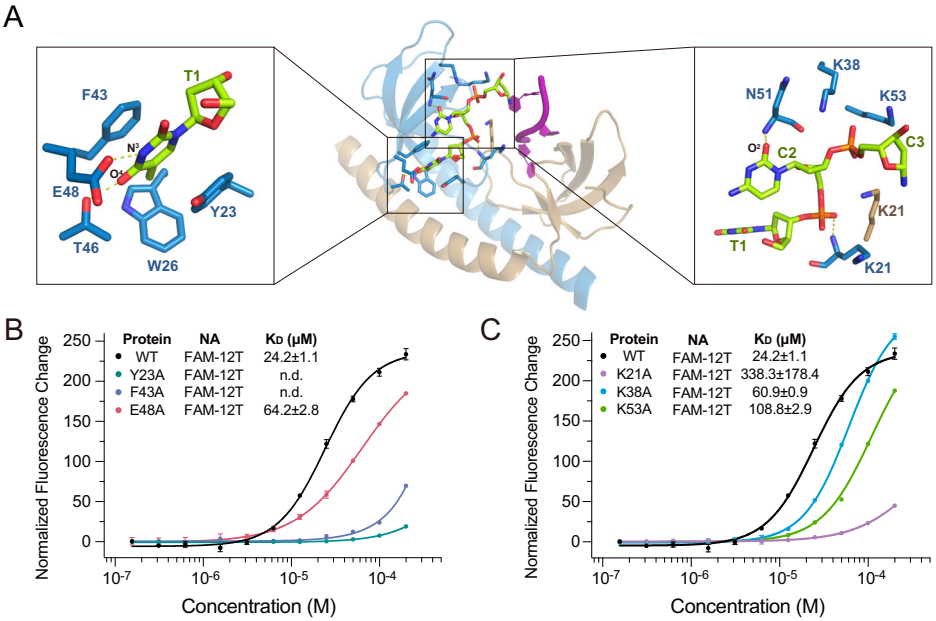


Fig. 5 | Recognition of the thymidine by the aromatic cage of dP75 PWWP. A The detailed interactions between dP75 PWWP domain and ssDNA. Hydrogen bonds are indicated by the dashed line. **B** The mutation of key residues in the aromatic cage affects the binding as measured by FP. **C** Mutation of key lysine residues affects the binding as measured by FP. Data are shown as mean ± s.d. ($n = 3$ biologically independent experiments).

Immunostaining with the H3K36me3 antibody confirmed the absence of K36 trimethylation in the H3K36A allele (Supplementary Fig. 7D). The homozygous H3K36A mutants are lethal at the pupal stage, hence allowing the examination of polytene chromosomes from third instar (L3) larvae[36]. Under wild-type conditions, dP75 localization was primarily observed at interband segments of DAPI staining (Fig. 7A). Absence of a chromatin signal in dP75-deficient mutants serves to affirm the specificity of the dP75 antibody used in this study (Supplementary Fig. 8). Notably, the dP75 localization pattern was retained in H3K36A specimens (Fig. 7A).

Analysis of polytene chromosome spreads from H3K36A larvae revealed no significant differences in the pattern and intensity of dP75 signals in 72% of the samples compared to controls. However, approximately 28% of the spreads from H3K36A larval salivary glands exhibited attenuated dP75 signals, though interband localization of dP75 remained detectable (Fig. 7A, B). This variance in immunofluorescence intensity may indicate intrinsic cellular heterogeneity within the salivary glands or suggest a potential reduction in dP75's affinity for chromatin in the absence of H3K36me3. Despite this variability, the consistent chromatin-association pattern of dP75 in H3K36A mutants supports the conclusion that H3K36me3 is not essential for dP75's chromatin-binding activity.

To investigate the in vivo role of the aromatic cage in dP75's DNA binding, we introduced Y23A and E48A mutations into the *Drosophila* H3K36A background. While native dP75 largely maintains chromatin binding in H3K36A mutants, the Y23A and E48A mutants showed minimal chromatin localization (Fig. 7C, D). This indicates that aromatic cage mutations influence dP75's chromatin association, even without H3K36me3, corroborating the crystal structure's demonstration of the aromatic cage's role in DNA binding.

Discussion

In this study, we elucidated the high-resolution crystal structures of the dP75 PWWP domain in both its apo form and in complex with DNA. We found that the dP75 PWWP domain forms a unique domain-swapped dimer and preferentially binds thymidine-rich DNA, including both double-stranded and single-stranded forms. Mutations of key residues highlighted the importance of the aromatic cage in thymidine recognition. Phenotypes such as transposon upregulation, H3K9me2 retention, impaired chromatin binding, and female infertility in mutant transgenic flies further underscored the significance of these residues. Importantly, we demonstrated that wild-type dP75 can effectively bind chromatin in H3K36A mutant flies, showing similar signal patterns and intensities in most samples compared to controls.

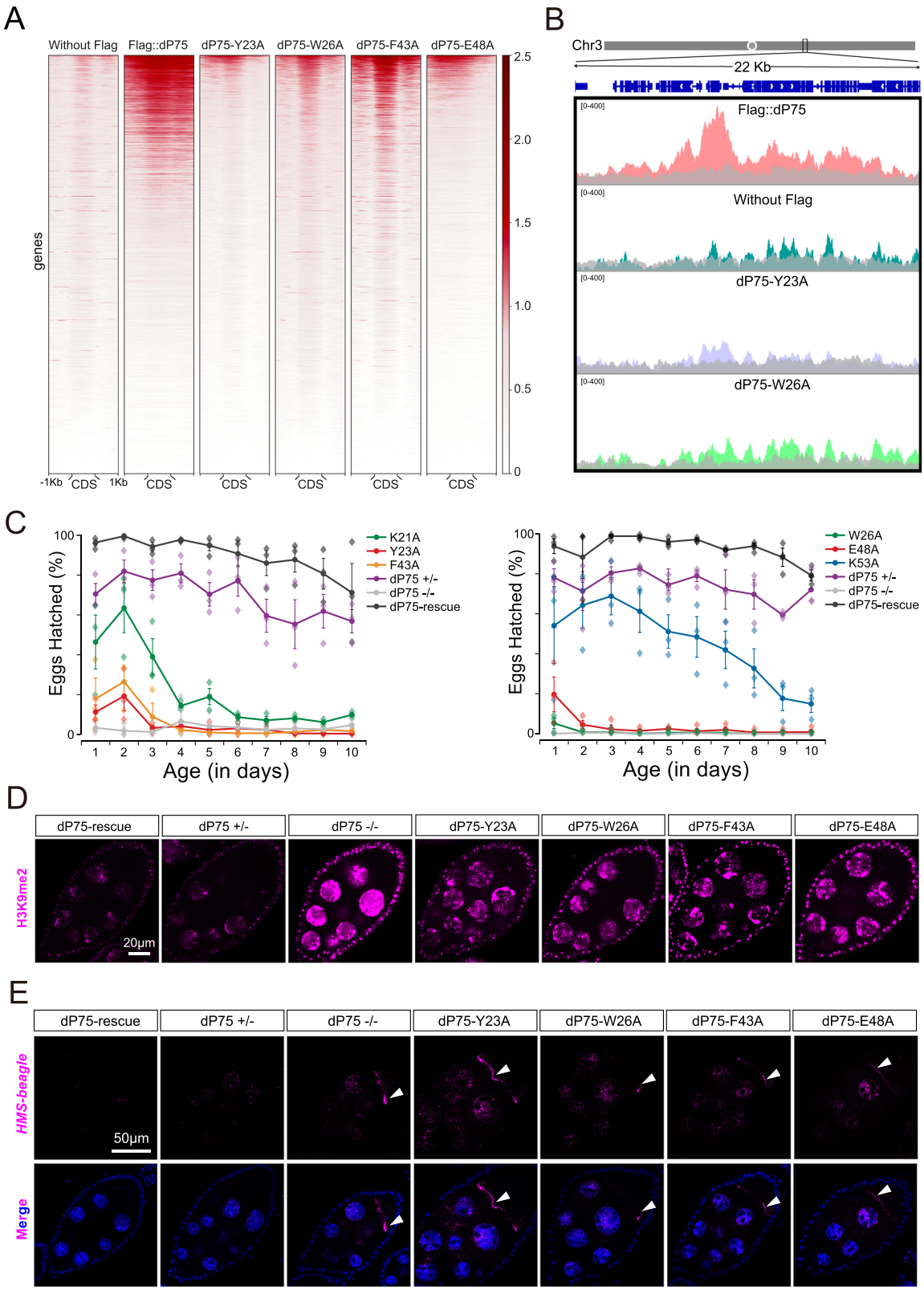
Structural analysis revealed that dimerization of the dP75 PWWP domain caused a conformational change in loop1, which connects β1 and β2, causing the Y23 residue to rotate approximately 56 degrees (Supplementary Fig. 9A, B). This rotation aligns the aromatic rings of Y23 and F43 parallel to each other, facilitating the π-π sandwich interactions with thymidine.

The structural constraints of the aromatic cage prevent it from accommodating larger bases such as adenine and guanine, resulting in weaker binding to these sequences (Supplementary Fig. 9C). The hydrophobic interaction between the thymine C^7 methyl group and W26 of dP75 enhances binding to thymine, while the absence of this methyl group in cytosine weakens cytosine binding by approximately 6.3 times (Fig. 3B). However, it is noteworthy that the aromatic cage of dP75 can only recognize individual thymine bases, with no evidence of recognition for adjacent bases.

This recognition mechanism bears some resemblance to the recognition of m^6A by the YTH domain. In comparison to the structure of YTHDC1 YTH domain in complex with an m^6A-modified RNA oligo (PDB 4RCJ), a similar π-π sandwich structure was observed (Supplementary Fig. 10). Three tryptophan residues formed the aromatic cage, with W470 and W411 positioned on the top and bottom of the m^6A base, respectively, and a hydrophobic interaction occurred between the methyl group of m6A and W465. Additionally, there is a hydrogen bond between D401 and N^1 of m6A. Thus, the PWWP domain of dP75 recognizes thymine using a DNA sequence-specific recognition mechanism analogous to how the YTH domain recognizes RNA m^6A.

The preference of dP75 PWWP for thymidine-rich sequences is particularly interesting given that its homolog, LEDGF, contains two AT-hook motifs that facilitate recognition of AT-rich dsDNA sequences—a feature absent in dP75. Therefore, dP75 PWWP likely changes the conformation of its aromatic cage through dimerization to bind AT-rich sequences. This conformational flexibility allows dP75 to bind both T-rich dsDNA and ssDNA, broadening its range of DNA substrates.

In addition to its established role in DNA interaction, the potential interaction between dP75 and the histone kinase Jil-1 is now considered. This interaction could offer deeper insights into dP75's functions within chromatin dynamics. Given Jil-1's involvement in histone modification and its implications in replication, there is a suggested potential relationship with dP75 in maintaining chromatin stability, especially during replication.



Further exploration of this relationship may elucidate how dP75 influences chromatin structure and stability, particularly within actively transcribed regions where replication and transcriptional activities converge.

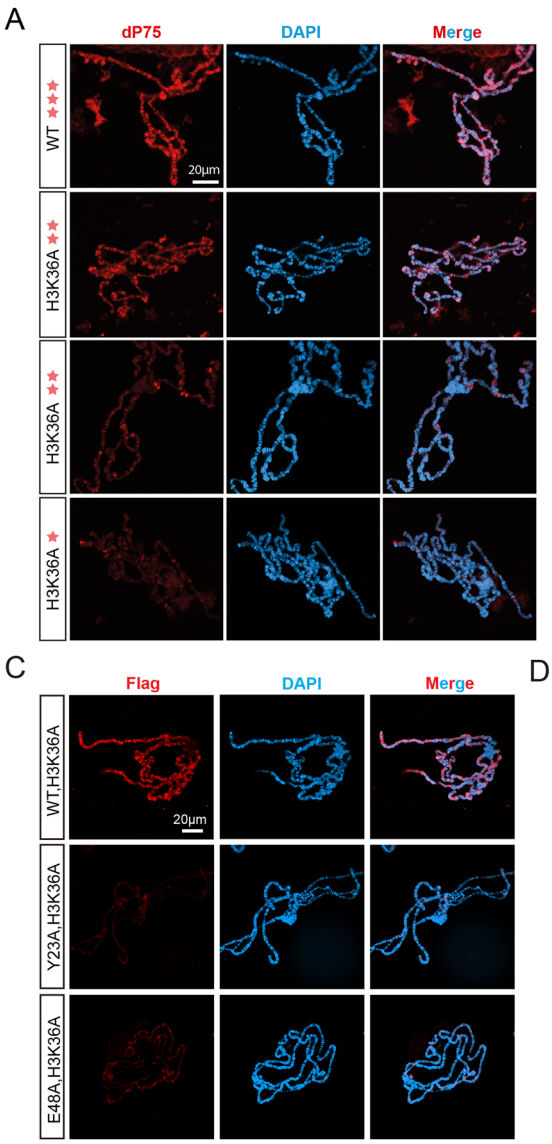
In summary, our study reveals a domain-swapped conformation of the dP75 PWWP domain and identifies its dual function as a histone and base reader. This study highlights the critical role of the dP75 binding pocket in ssDNA interactions, provides important insights into its in vivo functions, and underscores its relevance in chromatin biology. Our findings may elucidate a molecular mechanism for dP75 binding to chromatin.

Methods

The reagents used in this study are listed in Supplementary Table 3.

Fig. 6 | dP75 PWWP domain is critical for chromatin-binding, H3K9me2 restraining, transposon silencing, and female fertility. A Heatmap for Flag ChIP-seq signals of wild-type Flag::dP75 or Flag::dP75 with indicated point mutations. Each line from the heatmap represents the coding region of one gene plus its 1 kb upstream and downstream sequences. Two biological replicates were carried out and reached consistent results. **B** Genome-browser profile of the ChIP-seq coverage along a representative 22 kb window on chromosome 3. Flag ChIP performed in ovaries without the Flag tag serves as the negative control. **C** Fertility of females with indicated mutant transgenes in *dP75* knockout background. Y23, W26, F43, and E48 substitution to alanine individually exhibit sterility. $n = 3$ biologically independent experiments. Error bars indicate the standard deviation of the 3 biological replicates. **D** Representative images of H3K9me2 immunostaining in mutant and control ovaries. The full set of pictures with DAPI staining is shown in Supplementary Fig. 6A. **E** RNA FISH of transposon *HMS-beagle* in mutant and control ovaries. While wild-type *dP75* transgene fully restores *HMS-Beagle* activation in *dP75* knockout fly, transgenes with Y23, W26, F43, and E48 substitution to alanine individually show elevated transposon RNA signals compared to the control group.

Fig. 7 | H3K36me3 is not essential for dP75 binding to chromatin. A Immunofluorescence of dP75 in the salivary glands from wild type and H3K36A flies. dP75 was detected by immunostaining with the antibody against dP75 we generated before[8]. **B** Summarization of dP75 signal intensity from salivary glands of wild-type and H3K36A larvae. dP75 signal intensity was classified into three subgroups based on the brightness of the signals. N indicates the number of nucleus counted. Shown is the summarization of three biological replicates. **C** Immunofluorescence of Flag::dP75 and Flag::dP75 with substitution at the indicated amino acid in the salivary glands from H3K36A flies. **D** Summary of signal intensity of Flag::dP75 or point-mutated Flag::dP75 in salivary glands of H3K36A larvae. Flag::dP75 signal intensity was classified into three subgroups based on the brightness of the signals. N indicates the number of nucleus counted. Shown is the summarization of three biological replicates.



B

Genotype	Intensity of dP75 ★	★★	★★★	N
WT	-	40%	60%	80
H3K36A	28%	44%	28%	85

D

Genotype	Intensity of Flag::dP75 ★	★★	★★★	N
WT,H3K36A	-	4%	96%	160
Y23A,H3K36A	99%	1%	-	242
E48A,H3K36A	75%	24%	1%	128

Fly strains and husbandry conditions

The following fly strains were used: UASp-Flag-Myc-dP75 (attP40), UASp-Flag-Myc-dP75-K21A (attP40), UASp-Flag-Myc-dP75-Y23A (attP40), UASp-Flag-Myc-dP75-W26A (attP40), UASp-Flag-Myc-dP75-F43A (attP40), UASp-Flag-Myc-dP75-E48A (attP40), UASp-Flag-Myc-dP75-K53A (attP40), $dP75^{sg1}$/*TM6C*, *Sb*, Act5C-Gal4 (attP40), Ubi-NLuc-dP75 (attP2), Ubi-CLuc-dP75 (attP2), H3K36A, $dP75^{sg2}$/*TM6c*, *Sb*, and *Sco*/*CyO*, GFP; *Sb*/*TM6B*. The H3K36A strain was a generous gift from Dr. Guanjun Gao's lab[35], the *Sco*/*CyO*, GFP; *Sb*/*TM6B* strain was a gift from Dr. Jilong Liu lab, and all strains related to dP75 were generated by Z.M and K.D. All fly strains were maintained at room temperature on a standard agar-corn medium.

Cloning

The PCR-amplified fragments encoding the *Drosophila* melanogaster P75, the PWWP domain (spanning residues 9–90), and the human LEDGF/p75 PWWP domain (residues 4–89), were cloned into the pET28-SMT3 vector, which contains a 6×His tag at the N-terminus following a SUMO fusion protein. Mutants were generated using a Site-Directed Mutagenesis Kit (NEB, E0552S) according to the manufacturer's instructions. The plasmids

were transformed into *Escherichia coli* strain BL21 (DE3). The Mut Express II Fast Mutagenesis Kit (Vazyme, C214-01) was used to introduce the indicated point mutations to the UASp-Flag-Myc-dP75 vector. The positive colonies were validated by sequencing.

Dimerization assay in bacteria cells

The V5 tag, dP75, or dP75-E70P/K74P were cloned into the pWA21cBP-$AraC_{DNA}$-mTagBFP vector (modified from the pWA21cBP-EGFP-mTagBFP plasmid, vector ID: 0000447, from WeKwikGene, Westlake Laboratory), respectively. The V5 tag, dP75, or dP75-E70P/K74P were inserted in-frame with $AraC_{DNA}$ to generate fused proteins. The constructed plasmid was transformed into BW25113 competent bacteria. After Sanger sequencing, bacteria colonies with the right vector were incubated in an LB fluid medium supplemented with 100 μg/ml ampicillin, 0.02% arabinose, and 0.02% rhamnose. After incubation at 37 °C for 18–20 h, the bacteria cells were collected and imaged using Zeiss LSM 980 Airyscan (ZEISS, LSM980) to detect the BFP expression.

Split luciferase complementation assay

Gaussia luciferase was split into two parts, NLuc (17–109 aa) and CLuc (110–185 aa). dP75 was fused with NLuc and CLuc, respectively, and the corresponding transgenic fly lines were obtained. Ten pairs of fly ovaries were homogenized to extract proteins using cell lysis buffer in Renilla Luciferase Reporter Gene Assay Kit (Beyotime, RG016). The Protein level was determined using Detergent Compatible Bradford Protein Quantification Kit (Vazyme, E211-01). 200 μg proteins were used for luminescence detection on the substrate coelentarizine following the manufacturer's instructions of Renilla Luciferase Reporter Gene Assay Kit (Beyotime, RG016). The luminescence was detected by using multimode plate reader (Tecan, spark).

Protein expression and purification

For full-length dP75, its PWWP domain, and human LEDGF PWWP domain, the cells were grown at 37 °C to an OD_{600} of 0.6 and induced with 0.2 mM isopropylthiogalactoside overnight at 18 °C. The cells were collected by centrifugation at 4000 r.p.m. for 15 min and lysed using French Press (JNBIO). The proteins were purified using a HisTrap HP column (Cytiva), followed by the cleavage of the tag using Ulp1 protease. dP75 PWWP and LEDGF PWWP proteins were purified with a HisTrap S column (Cytiva) and Superdex G75 Hiload 16/60 column (Cytiva) in a buffer containing 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 1 mM dithiothreitol (DTT). The target proteins' fractions were pooled and concentrated to 10 mg/mL. For the full-length dP75, following the cleavage of His-SUMO tag, the protein was directly loaded on Superdex G75 Hiload 16/60 column and fractions containing full-length dP75 were pooled and concentrated to 5 mg/mL in a buffer containing 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 1 mM DTT.

Crystallization, data collection, and structure determination

Crystals of dP75 PWWP apo-form domain-swapped dimer were grown at 0.1 M Bicine, pH 8.5, 30% PEG1500. Crystals of dP75 PWWP in complex with ssDNA-2 (5'-TCCCT-3') were grown at 0.1 M Tris-HCl, pH 8.5, and 25% PEG3350. Crystals were grown at 17 °C using the hanging drop vapor diffusion method. For data collection, crystals were flash-frozen in liquid nitrogen in the above reservoir solution supplemented with glycerol. Diffraction data were collected at beamlines BL18U1 and BL19U1 of Shanghai Synchrotron Radiation Facility (SSRF). The data sets were processed using HKL3000[37]. The structures were solved by molecular replacement with the LEDGF (PDB 4FU6) structure as a search model. Model building and structure refinement were carried out using COOT[38] and PHENIX[39]. The structural graphics were generated using PyMol (The PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC). The statistics of data collection and refinement are summarized in Supplementary Table 2.

Crosslinking coupled mass spectrometry

Purified dP75 protein was incubated with 1 mM disuccinimidyl suberate (DSS; Pierce) in a reaction buffer containing 50 mM Hepes, pH 7.5, and 500 mM NaCl at room temperature for 2 h. To terminate the reaction, the quenching buffer (0.5 M ammonium bicarbonate) was added to a final concentration of 20 mM. The reaction was further incubated at room temperature for 20 min. Samples were precipitated by acetone and digested for 16 h at 37 °C by trypsin at an enzyme-to-substrate ratio of 1:50 (w/w). The tryptic-digested peptides were desalted and loaded onto an in-house packed capillary reverse-phase C18 column (40 cm length, 100 μm ID x 360 μm OD, 1.9 μm particle size, 120 Å pore diameter) connected to an Easy LC 1200 system. The samples were analyzed with a 120 min-HPLC gradient from 6% to 35% of buffer B (buffer A: 0.1% formic acid in Water; buffer B: 0.1% formic acid in 80% acetonitrile) at 300 nL/min. The eluted peptides were ionized and directly introduced into a Q-Exactive mass spectrometer using a nano-spray source. Survey full-scan MS spectra (from m/z 300–1800) were acquired in the Orbitrap analyzer with resolution r = 70,000 at m/z 400. Crosslinked peptides were identified and evaluated using pLink2[40] software and then visualized using Xlinker Analyzer[41] in Chimera[42].

Fluorescence polarization assays

The oligos were synthesized with fluorescein amidite (FAM) labeled at 5′-end (FAM-12T: 5′-FAM-TTTTTTTTTTTT-3′; FAM-12A: 5′-FAM-AAAAAAAAAAAA-3′; FAM-12C: 5′;-FAM-CCCCCCCCCCCC-3′; FAM-12G: 5′-FAM-GGGGGGGGGGGG-3′; 12T-FAM: 5′-TTTTTTTTTTTT T-FAM-3′; FAM-random: 5′-FAM-CGATCTGCGTCAGCTGTAGC-3′) (Bioneer). Fluorescence Polarization (FP) experiments were performed in 96 well plates (Corning, 3650) with the fluorescence reader Synergy™ NEO (BioTek). Serial dilutions of purified wild-type and mutant dP75 PWWP proteins were prepared in FP assay buffer (100 mM NaCl, 10 mM Tris-HCl, pH 8.0) with final concentrations ranging from 0.244 to 250 μM. FAM-labeled oligos were then added to a final concentration of 4 nM for a final assay volume of 100 μL. The reaction mixture was incubated for 30 min at room temperature. Polarization was measured at an excitation wavelength of 485 nm and an emission wavelength of 528 nm. Each plate was read three times, and the values were averaged before analysis. All dissociation constant (K_D) values were determined by fitting the titration curve with GraphPad Prism 9.0 software using the Dose-Response model.

Hatch ratio test

Five virgin flies of the indicated genotypes and five wild-type male flies of 4–6-day-old were put together in an empty bottle. An apple juice agar plate with a pat of wet yeast paste was served daily to the flies and collected after 24 h. The collected plates were maintained at room temperature for another 24–30 h to allow eggs to hatch. Hatching ratios were calculated by dividing the number of hatched eggs by the number of total eggs laid. Three independent replicates were carried out for each experiment.

Immunostaining

Four to five pairs of ovaries from 4-6-day-old flies were dissected in PBS and fixed in PBS with 4% paraformaldehyde for 8 min at room temperature. Then, the ovaries were washed in PBST (PBS solution with 0.2% Triton X-100) 3 times for a total of 30 min. Ovaries were incubated in 10% normal goat serum overnight at 4 °C. Primary antibodies were added in blocking buffer, and ovaries were incubated overnight at 4 °C. Ovaries were washed three times for a total of 30 min and then incubated in a blocking buffer with a secondary antibody overnight at 4 °C. Ovaries were washed 3 times for a total of 30 min, followed by incubation in PBST with DAPI for 5 min. Then, the ovaries were washed twice in PBST and mounted with 50% glycerol in PBS. Primary antibodies used in this study include anti-Flag (1:300), anti-dP75 (1:50), anti-H3K36me3 (1:200), and anti-H3K9me2 (1:300). Secondary antibodies used in this study include Alexa Fluor 594 Goat Anti-Rabbit IgG(H + L) (1:300) and Alexa Fluor 594 Goat Anti-Mouse IgG(H + L) (1:300). For the polytene

squash and immunostaining experiments, the protocol described in Johansen et al. 2009 was exactly followed[36]. Images were obtained using Zeiss LSM 980 Airyscan2.

Western blotting

Five pairs of ovaries were dissected on ice-cold PBS, and proteins were extracted using RIPA lysis buffer (Beyotime, P0013B). Then 5×loading buffer was added to the samples and boiled at 95 °C for 15 min. Then the protein was loaded in SDS-PAGE gel and transferred into NC membrane. After incubation in blocking buffer, primary antibody, and secondary antibody, the chemiluminescent signal was detected using Amersham ImageQuant 800 (Cytiva, AI800). The following antibodies were used: anti-dP75 (1:500), anti-FLAG antibody (1:3000), anti-β-Tubulin antibody (1:3000), Goat Anti-Rabbit IgG, HRP Conjugated (1:3000) and Goat Anti-Mouse IgG, HRP Conjugated (1:3000). Details of the antibodies are in Supplementary Table 3.

ChIP-Seq and ChIP-Seq data analysis

ChIP experiments were performed according to previous method[8]. Briefly, 50–60 pairs of ovaries from 4 to 6-day-old flies were dissected in ice-cold PBS for each sample, and fixed in 1.8% paraformaldehyde. The reaction was quenched by adding 250 mM glycine. The samples were washed in PBS twice and in sonication buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% sodium dodecyl sulfate, and 1×Protease inhibitor) once at 4 °C. Samples were sonicated using a bioruptor sonicator with the setting of 30 s, on/ 30 s, off, total 30 cycles. Then, the samples were spun at 4 °C to remove the pellet. 5% supernatant was saved as input, and the rest supernatant was used to IP. For IP, each sample were diluted 10 times in dilution buffer (16.7 mM Tris-HCl, pH8.0, 1.1 mM EDTA, 1.1% Triton X-100, 167 mM NaCl and 1×Protease inhibitor), mixed with 50 µL Flag agarose beads (GenScript, L00432-5), and incubated overnight at 4 °C. The beads were then washed in Wash buffer A (20 mM Tris-HCl, pH8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 150 mM NaCl), Wash buffer B (20 mM Tris-HCl, pH8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 500 mM NaCl), Wash buffer C (10 mM Tris-HCl, pH8.0, 1 mM EDTA, 1% NP-40, 1% Sodium deoxycolate, 250 mM LiCl), and TE buffer (10 mM Tris-HCl, pH8.0, 1 mM EDTA). Next, the samples were eluted in elution buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM EDTA, and 0.5% sodium dodecyl sulfate) at 65 °C for 6 h. Then, the eluted DNA was extracted using VAHTS DNA Clean Beads (Vazyme, N411-01).

VAHTS DNA Clean Beads were used according to the manufacturer's instructions. The DNA libraries were prepared using VAHTS Universal DNA Library Prep Kit for Illumina® V3. The libraries were sequenced by Illumina NextSeq with pair-end 150nt runs. The reads were mapped via BWA[43] and analyzed using deepTools[44]. The enrichment of the IP group over input was calculated for each genotype. The same number of mapped reads was used in control and experimental groups for generating the heatmaps, and data from two independent biological replicates were analyzed and reached consistent results.

RNA FISH

The RNA FISH experiments were carried out according to previous method[8]. Briefly, 3–4 pairs of ovaries were dissected in ice-cold PBS and fixed in 4% formaldehyde for 20 min. Samples were washed in PBST (0.1% Triton X-100) once and PBS twice at room temperature. Samples were soaked in 70% ethanol at 4 °C overnight. And then samples were washed in Wash buffer A^+ (Wash buffer A with 10% formamide) once and incubated in Hybridization buffer with Stellaris RNA FISH probe (1:100) overnight at 29 °C. The samples were washed once in Wash buffer A^+, stained in DAPI, washed in buffer B, and mounted with 50% glycerol in PBS. The RNA FISH kit (LGC Biosearch Tech, wash buffer A: Cat# SMF-WA1-60, Hybridization Buffer: Cat# SMF- HB1-10, Wash Buffer B: Cat# SMF- WB1-20) was used.

Stellaris RNA FISH probes were designed and purchased from LGC Bioresearch Technology.

Statistics and Reproducibility

One-way ANOVA was used to evaluate the statistical significance of the luminescence results and the BFP results, respectively, between control and experimental groups. *p*-values were calculated using GraphPad Prism 10 software. Standard deviation was calculated using GraphPad Prism 10 software for the luminescence results, BFP results, fertility assays, the immunofluorescence signal intensities, and the RT-qPCR results, with mean ± s.d. presented. For FP, bind curves were fitted in a Dose Response model which incorporates a Hill equation with GraphPad Prism 10 software. Data were normalized and are presented as the mean ± s.d. of the interpolated K_D from three independent experiments.

For all the data presented in this manuscript, at least two biological replicates were performed, and the replicates reached consistent conclusion. The exact n number for each statistical analysis was indicated in the corresponding figure legend and method.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic coordinates and associated structure factors have been deposited in the Protein Data Bank under accession codes 6KD6 and 6KCO. The deep sequencing data have been deposited to NCBI, (PRJNA1195629). The crosslinking mass spectrometry data have been deposited to ProteomeXchange with the accession codes PXD058799. The data generated or analyzed during the current study are included in the Figures and Supplementary information files. The uncropped and unedited blots and gel images are included in Supplementary Fig. 11. Source data are provided in Supplementary Data 1. Fly strains generated in this study are available from K.D. upon request.

Received: 7 August 2024; Accepted: 6 March 2025;
Published online: 15 March 2025

References

1. Rijck, J. D., Bartholomeeusen, K., Ceulemans, H., Debyser, Z. & Gijsbers, R. High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region. *Nucleic Acids Res.* **38**, 6135–6147 (2010).
2. Christ, F. & Debyser, Z. The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy. *Virology* **435**, 102–109 (2013).
3. Gijsbers, R. et al. LEDGF Hybrids Efficiently Retarget Lentiviral Integration Into Heterochromatin. *Mol. Ther.* **18**, 552–560 (2010).
4. Ferris, A. L. et al. Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc. Natl. Acad. Sci. USA***107**, 3135–3140 (2010).
5. Cherepanov, P. et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* **278**, 372–381 (2003).
6. Yokoyama, A. & Cleary, M. L. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* **14**, 36–46 (2008).
7. Tesina, P. et al. Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. *Nat. Commun.* **6**, 7968 (2015).
8. Dou, K. et al. Drosophila P75 safeguards oogenesis by preventing H3K9me2 spreading. *J. Genet. Genomics* **47**, 187–199 (2020).
9. Albig, C. et al. JASPer controls interphase histone H3S10 phosphorylation by chromosomal kinase JIL-1 in Drosophila. *Nat. Commun.* **10**, 5343 (2019).
10. Rondelet, G., Maso, T. D., Willems, L. & Wouters, J. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *J. Struct. Biol.* **194**, 357–367 (2016).

11. Maurer-Stroh, S. et al. The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **28**, 69–74 (2003).
12. Turlure, F., Maertens, G., Rahman, S., Cherepanov, P. & Engelman, A. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. *Nucleic Acids Res.* **34**, 1653–1665 (2006).
13. Chen, L.-Y. et al. Domain swapping and SMYD1 interactions with the PWWP domain of human hepatoma-derived growth factor. *Sci. Rep. UK* **8**, 287 (2018).
14. Pünzeler, S. et al. Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. *Embo J.* **36**, 2263–2279 (2017).
15. Qiu, C., Sawada, K., Zhang, X. & Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.* **9**, 217–224 (2002).
16. Tian, W. et al. The HRP3 PWWP domain recognizes the minor groove of double-stranded DNA and recruits HRP3 to chromatin. *Nucleic Acids Res.* **47**, 5436–5448 (2019).
17. Eidahl, J. O. et al. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.* **41**, 3924–3936 (2013).
18. Wang, H., Farnung, L., Dienemann, C. & Cramer, P. Structure of H3K36-methylated nucleosome–PWWP complex reveals multivalent cross-gyre binding. *Nat. Struct. Mol. Biol.* **27**, 8–13 (2020).
19. Llano, M. et al. Identification and Characterization of the Chromatin-binding Domains of the HIV-1 Integrase Interactor LEDGF/p75. *J. Mol. Biol.* **360**, 760–773 (2006).
20. Shun, M.-C. et al. Identification and Characterization of PWWP Domain Residues Critical for LEDGF/p75 Chromatin Binding and Human Immunodeficiency Virus Type 1 Infectivity. *J. Virol.* **82**, 11555–11567 (2008).
21. Bröhm, A. et al. Methylation of recombinant mononucleosomes by DNMT3A demonstrates efficient linker DNA methylation and a role of H3K36me3. *Commun. Biol.* **5**, 192 (2022).
22. Vezzoli, A. et al. Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat. Struct. Mol. Biol.* **17**, 617–619 (2010).
23. Laguri, C. et al. Human Mismatch Repair Protein MSH6 Contains a PWWP Domain That Targets Double Stranded DNA. *Biochem.-us* **47**, 6199–6207 (2008).
24. Zhang, M. et al. Histone and DNA binding ability studies of the NSD subfamily of PWWP domains. *Biochem. Bioph. Res. Co.* **569**, 199–206 (2021).
25. Sankaran, S. M., Wilkinson, A. W., Elias, J. E. & Gozani, O. A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin. *J. Biol. Chem.* **291**, 8465–8474 (2016).
26. Fei, J. et al. NDF, a nucleosome destabilizing factor that facilitates transcription through nucleosomes. *Gene Dev.* **32**, 682–694 (2018).
27. Link, S. et al. PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. *Nat. Commun.* **9**, 4300 (2018).
28. Rona, G. B., Eleutherio, E. C. A. & Pinheiro, A. S. PWWP domains and their modes of sensing DNA and histone methylated lysines. *Biophys. Rev.* **8**, 63–74 (2016).
29. Koutná, E. et al. Multivalency of nucleosome recognition by LEDGF. *Nucleic Acids Res.* **51**, 10011–10025 (2023).
30. Zhang, M. et al. Crystal structure of the BRPF2 PWWP domain in complex with DNA reveals a different binding mode than the HDGF family of PWWP domains. *Biochimica Et. Biophysica Acta BBA Gene Regul. Mech.* **1864**, 194688 (2021).
31. Li, J. et al. H3K36 methylation and DNA-binding both promote loc4 recruitment and Isw1b remodeler function. *Nucleic Acids Res.* **50**, 2549–2565 (2022).
32. Zhang, H. et al. Bright and stable monomeric green fluorescent protein derived from StayGold. *Nat. Methods* **21**, 657–665 (2024).
33. Wegrzyn, K. & Konieczny, I. Archaeal Orc1 protein interacts with T-rich single-stranded DNA. *BMC Res. Notes* **14**, 275 (2021).
34. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. *Mol. Cell* **50**, 749–761 (2013).
35. Zhang, W. et al. Probing the Function of Metazoan Histones with a Systematic Library of H3 and H4 Mutants. *Dev. Cell* **48**, 406–419.e5 (2019).
36. Johansen, K. M. et al. Polytene chromosome squash methods for studying transcription and epigenetic chromatin modification in Drosophila using antibodies. *Methods* **48**, 387–397 (2009).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
39. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 213–221 (2010).
40. Chen, Z.-L. et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404 (2019).
41. Kosinski, J. et al. Xlink Analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015).
42. Pettersen, E. F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
43. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
44. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

Acknowledgements

We thank the Kiryl Piatkevich lab for generously sharing the pWA21cBP-EGFP-mTagBFP plasmid and the Yang Yu lab from Guangzhou Medical University for supplying the pENTR4-NLlu and pENTR4-CLlu plasmids. We also acknowledge the staff at beamlines 18U1 and 19U1 at Shanghai Synchrotron Radiation Facility (SSRF) for assistance with data collection, as well as the core facilities of Molecular and Cellular Biology and Molecular Imaging of the School of Life Science and the HPC Platform at ShanghaiTech University. We thank the Fly Core Facility at the CAS Center for Excellence in Molecular Cell Science for help with Drosophila embryo injections. This work was supported by the National Natural Science Foundation of China (grant numbers 32171186 and 91940302 to Y.H.; 32100469 to K.D.) and the Science and Technology Foundation of Shanghai (grant number 20PJ1410600 and 20ZR1435900 to K.D.).

Author contributions

Conceptualization, Y.H. and K.D.; Methodology (in vitro assays), Z.J., Y.L.; Methodology (in vivo assays), Z.M.; Software and ChIP-seq analysis, C.L.; Mass spectrometry assays, Y.Y.; Reagent, X.Z., and G.G.; Validation, Z.J., Y.L., and Z.M.; Writing, Y.H., K.D., and Z.M.; Funding acquisition, Y.H. and K.D.

Competing interests

The authors declare no competing of interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07895-2.

Correspondence and requests for materials should be addressed to Kun Dou or Ying Huang.

Peer review information *Communications Biology* thanks Jinrong Min and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Jun Wei Pek and Mengtan Xing.