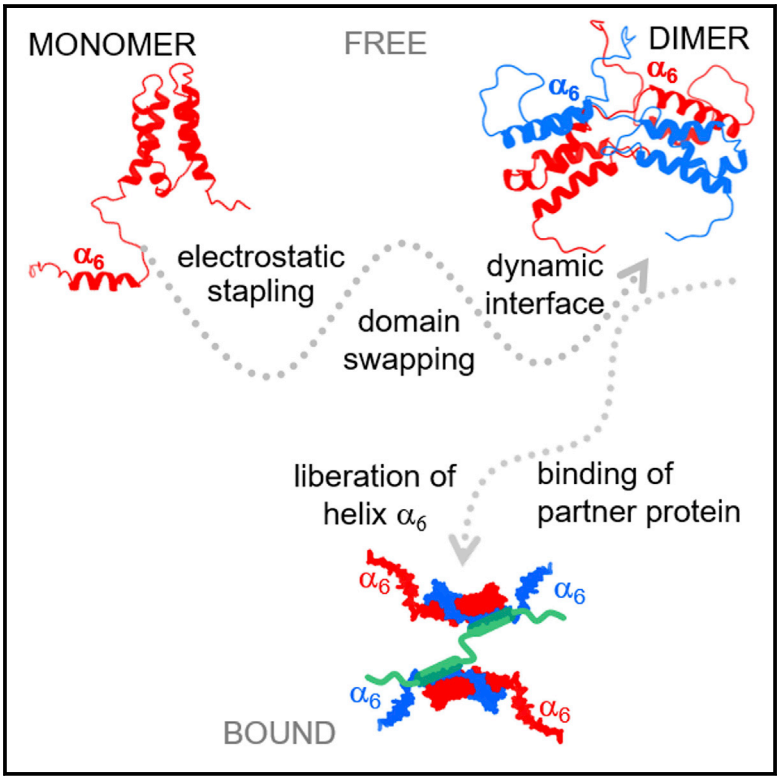
Structure

Article

Molecular Mechanism of LEDGF/p75 Dimerization

Graphical Abstract



Highlights

- Full-length LEDGF/p75 forms dimers with a low micromolar K_D
- LEDGF/p75 residues 345–467 is the minimal stable dimerization domain
- LEDGF dimer is stabilized by domain swapping and additional electrostatic stapling
- LEDGF dimerization does not impair binding of interaction partners

Authors

Vanda Lux, Tine Brouns, Kateřina Čermáková, ..., Frauke Christ, Zeger Debyser, Václav Veverka

Correspondence

zeger.debyser@kuleuven.be (Z.D.), vaclav.veverka@uochb.cas.cz (V.V.)

In Brief

Dimerization is an important process regulating eukaryotic transcription. Lux et al. use a combination of biophysical and biochemical techniques to investigate the dimerization mechanism of LEDGF/p75, the H3K36 methylation reader, and its effect on molecular interactions with other proteins.

Lux et al., 2020, Structure *28*, 1288–1299
December 1, 2020
https://doi.org/10.1016/j.str.2020.08.012

CellPress

CellPress

Structure

Article

Molecular Mechanism of LEDGF/p75 Dimerization

Vanda Lux,[1,7] Tine Brouns,[2,7] Kateřina Čermáková,[1,3] Pavel Srb,[1] Milan Fábry,[4] Marcela Mádlíková,[1] Magdalena Hořejší,[4] Zdeněk Kukačka,[5] Petr Novák,[5] Michael Kugler,[1] Jiří Brynda,[1,4] Jan DeRijck,[2] Frauke Christ,[2] Zeger Debyser,[2,*] and Václav Veverka[1,6,8,*]

[1]Structural Biology, Institute of Organic Chemistry and Biochemistry of the CAS, Prague 16000, Czech Republic
[2]Molecular Virology and Gene Therapy, KU Leuven, Molecular Virology and Gene Therapy, Leuven, 3000 Flanders, Belgium
[3]Department of Molecular & Cellular Biology, Center for Precision Environmental Health, and Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
[4]Structural Biology, Institute of Molecular Genetics of the CAS, Prague 14220, Czech Republic
[5]Protein Structure Characterization by Advanced Mass Spectrometry, Institute of Microbiology of the CAS, Prague 14220, Czech Republic
[6]Department of Cell Biology, Faculty of Science, Charles University, Prague 12800, Czech Republic
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: zeger.debyser@kuleuven.be (Z.D.), vaclav.veverka@uochb.cas.cz (V.V.)
https://doi.org/10.1016/j.str.2020.08.012

SUMMARY

Dimerization of many eukaryotic transcription regulatory factors is critical for their function. Regulatory role of an epigenetic reader lens epithelium-derived growth factor/p75 (LEDGF/p75) requires at least two copies of this protein to overcome the nucleosome-induced barrier to transcription elongation. Moreover, various LEDGF/p75 binding partners are enriched for dimeric features, further underscoring the functional regulatory role of LEDGF/p75 dimerization. Here, we dissected the minimal dimerization region in the C-terminal part of LEDGF/p75 and, using paramagnetic NMR spectroscopy, identified the key molecular contacts that helped to refine the solution structure of the dimer. The LEDGF/p75 dimeric assembly is stabilized by domain swapping within the integrase binding domain and additional electrostatic "stapling" of the negatively charged α helix formed in the intrinsically disordered C-terminal region. We validated the dimerization mechanism using structure-inspired dimerization defective LEDGF/p75 variants and chemical crosslinking coupled to mass spectrometry. We also show how dimerization might affect the LEDGF/p75 interactome.

INTRODUCTION

Lens epithelium-derived growth factor/75 (LEDGF/p75), a 530-residue long protein, was initially discovered as a transcriptional coactivator and component of the general RNA polymerase II transcription machinery (Ge et al., 1998). It was reported to act as a pro-survival protein, by binding heat shock- and stress-related elements in promoter regions of many stress-related genes (Shinohara et al., 2002), and was later indicated as a major cellular interaction partner of HIV-1 integrase (IN) (Cherepanov et al., 2003). Recently, LEDGF/p75 together with its ortholog HRP2 was identified as a key factor allowing RNA polymerase II to overcome the nucleosome-induced barrier to transcription elongation by assuming the "facilitates chromatin transcription complex" role in differentiated cells (LeRoy et al., 2019).

The protein domain responsible for its interaction with integrase was mapped to residues 347–429 and named the integrase binding domain (IBD) (Cherepanov et al., 2004). The N-terminal part of LEDGF/p75 contains a PWWP motif, a nuclear localization signal, and two AT-hook-like motifs, which are necessary for association of the protein with the chromatin (Llano et al., 2004; Turlure et al., 2006). The N-terminal portion of the protein (amino acid residues 1–326) is shared with the alternative splice variant LEDGF/p52. Unlike LEDGF/p52, LEDGF/p75 plays a crucial role in cancer through its interaction with mixed-lineage leukemia (MLL/MENIN) fusion proteins whereby LEDGF/p75 is involved in MLL-dependent transcription and leukemic transformation (Cermakova et al., 2016; Yokoyama and Cleary, 2008). LEDGF/p75 IBD also interacts with several other cellular proteins and protein complexes, including the phosphatidylinositol 3-kinase regulator JPO2 (CDCA7L) (Bartholomeeusen et al., 2007; Maertens et al., 2006), the zinc-finger protein POGZ (Bartholomeeusen et al., 2009), the cell-cycle regulator ASK (DBF4) (Hughes et al., 2010), the transcription elongation regulator IWS1 (Tesina et al., 2015), and a subunit of the Mediator complex MED1 (Sharma et al., 2018). Finally, LEDGF/p75 is also implicated in atopic and inflammatory disorders, and LEDGF/p75 autoantibodies have been identified (Wu et al., 2002).

To perform all these functions in parallel, the interactome of LEDGF/p75 is subjected to tight regulatory mechanisms. We recently demonstrated that all known cellular binding partners of LEDGF/p75 contain an intrinsically disordered IBD-binding

motif (IBM) (Sharma et al., 2018) that can be phosphorylated, which regulates the affinity of these proteins for LEDGF/p75 (Sharma et al., 2018). Also, post-translational sumoylation of LEDGF/p75 was observed at different sites in the N- and C-terminal parts of the protein and this modulates its transcriptional activity (Bueno et al., 2010).

Structural investigation of protein-DNA binding with scanning force microscopy (Vanderlinden et al., 2014) revealed the ability of LEDGF/p75 dimers to induce DNA bridging. Surprisingly, DNA synapse formation was not observed for the monomeric LEDGF/p75 molecule. This result was intriguing, since synapse formation is observed in all aspects of DNA metabolism, including transcription, replication, and recombination (Weickert and Adhya, 1993; Yanofsky, 1971). A putative role for dimerization is further supported by the fact that at least three known LEDGF/p75 eukaryotic binding partners—JPO2, ASK, and POGZ—bear two copies of IBMs that are relatively structurally confined and could benefit from the LEDGF/p75 dimeric organization (Sharma et al., 2018; Tesina et al., 2015). The N-terminal region of JPO2 implicated in LEDGF/p75 binding is intrinsically disordered and harbors two sequentially positioned IBMs (residues 9–33 and 65–91). Similarly, ASK contains two IBMs separated by a shorter linker sequence at the disordered C terminus (residues 614–637 and 655–674). While POGZ does not carry two distinct IBMs, it forms a dimeric assembly through the structured 23-kDa C-terminal DDE-1 domain followed by one copy of the IBM (residues 1,380–1,404). Each of these motifs is sufficient to bind the LEDGF/p75 IBD (Sharma et al., 2018; Tesina et al., 2015). Moreover, JPO2 can associate with two LEDGF/p75 IBD molecules simultaneously as determined by ITC (Sharma et al., 2018; Tesina et al., 2015). We decided to investigate whether dimerization plays a role in the context of LEDGF/p75 and its interactome.

RESULTS

Contributions of C-Terminal Interactions to LEDGF/p75 Dimerization

We first showed that the full-length LEDGF/p75 protein readily forms dimers in a concentration-dependent manner using the AlphaScreen assay (Figures 1A and 1B). While full-length glutathione-S-transferase (GST)-tagged LEDGF/p75 (GST-p75) and the IBD extended to the very C terminus (GST-Δ325 p75) formed assemblies with Flag-tagged full-length LEDGF/p75, the splice variant p52 (GST-p52) did not bind to full-length LEDGF/p75 in the absence of DNA (Figure 1B). We determined a dissociation constant of dimer formation for full-length LEDGF/p75 of 1 μM and for a C-terminal LEDGF/p75 construct (residues 325–530) of ~0.6 μM using microscale thermophoresis (Figure 1C). The values were further validated using analytical ultracentrifugation (Figure 1D), which confirmed the prevalence of a dimeric state at the 25 μM protein concentration. Next, we investigated which part of the protein is responsible for dimerization. The AlphaScreen data suggest that the minimal LEDGF/p75 dimerization region is limited to the IBD and C-terminal tail. During data collection for NMR structural characterization using the protein variant spanning residues 345–530, a frequent and relatively homogeneous C-terminal proteolysis product was detected. Subsequent analysis revealed that the C terminus was cleaved after the lysine at position 467, as signals from the following residues (468–530) were gradually disappearing from NMR spectra over time. Analytical ultracentrifugation and size-exclusion chromatography revealed that this proteolysis did not affect the capacity of $LEDGF_{345-467}$ to dimerize at the low micromolar concentration range (Figures 1D and 1E) and we therefore prepared a series of C-terminally extended constructs of the IBD up to residue K467 (Figure 1E). In addition, we introduced a deletion and charge-modifying point mutations in a region of the helix α_6 that was identified by NMR (Figure 2A). The purified LEDGF/p75 variants were analyzed using analytical size-exclusion chromatography (Figure 1E). The LEDGF/p75 IBD eluted from the size-exclusion chromatography column as a monomer and similar behavior was observed for a LEDGF/p75 construct spanning residues 345–441. The C-terminal extensions that included helix α_6 resulted in a shift toward the dimer, but the elution profile contained a shoulder, indicating the presence of a monomer-dimer equilibrium at the low μM concentration range (Figure 1E). Deletion of the entire helix α_6 or the point mutations within helix α_6 have effects similar to those of shortening the C-terminal tail, indicated by a shift in the equilibrium toward monomers. These results clearly suggest that helix α_6 plays a key role in stabilizing the LEDGF/p75 dimer. In the biochemical experiments, the variant spanning residues 345–467 was the most stable of the constructs tested. We postulated that it is the minimal stable region necessary for dimerization and used it in further studies.

Analysis of Dimer-Stabilizing Contacts

To investigate the structural determinants of LEDGF/p75 dimer formation, we obtained data from NMR relaxation and paramagnetic relaxation enhancement measurements and complemented them with data from chemical crosslinking experiments. First, we obtained the sequence-specific backbone resonance assignments for $^{13}C/^{15}N$-labeled $LEDGF_{345-467}$. A detailed analysis of the backbone chemical shift values revealed the presence of an additional α helix within the C-terminal part of $LEDGF_{345-467}$ (helix α_6, Figure 2A) that was thought to be intrinsically disordered. Although well defined, helix α_6 is more dynamic than helices from the IBD, based on the order parameter derived from chemical shift values (S^2, Figure 2A). Detailed analysis of the steady-state heteronuclear Overhauser effect ($\{^1H\}$-^{15}N nuclear Overhauser enhancement [NOE]) data, supported by R_1 and R_2 relaxation rates obtained for the $LEDGF_{345-467}$ backbone resonances provided valuable insights into backbone dynamics. As expected, the IBD region (345–430) is highly rigid, whereas the segment between residues 430 and 460 within the C-terminal part of the protein appears to be more dynamic. The negative values of the $\{^1H\}$-^{15}N NOE observed for the C-terminal residues (460–467) of $LEDGF_{345-467}$ upstream of helix α_6 indicate that this segment is flexible. We prepared a $LEDGF_{345-467}$ construct with helix α_6 deleted ($LEDGF_{345-467}$ Δ443–453) and tested its ability to dimerize using size-exclusion chromatography (Figure 1E). A reduced capacity to migrate on the column at dimeric elution volumes suggested a central role for helix α_6 in LEDGF/p75 dimer stabilization.

Next, we introduced a single cysteine residue at specific positions (Figures 2B and S2) within $LEDGF_{345-467}$ that was subsequently covalently coupled to a paramagnetic or diamagnetic probe using thiol chemistry. This modification allowed to detect paramagnetic relaxation enhancement (PRE) in a

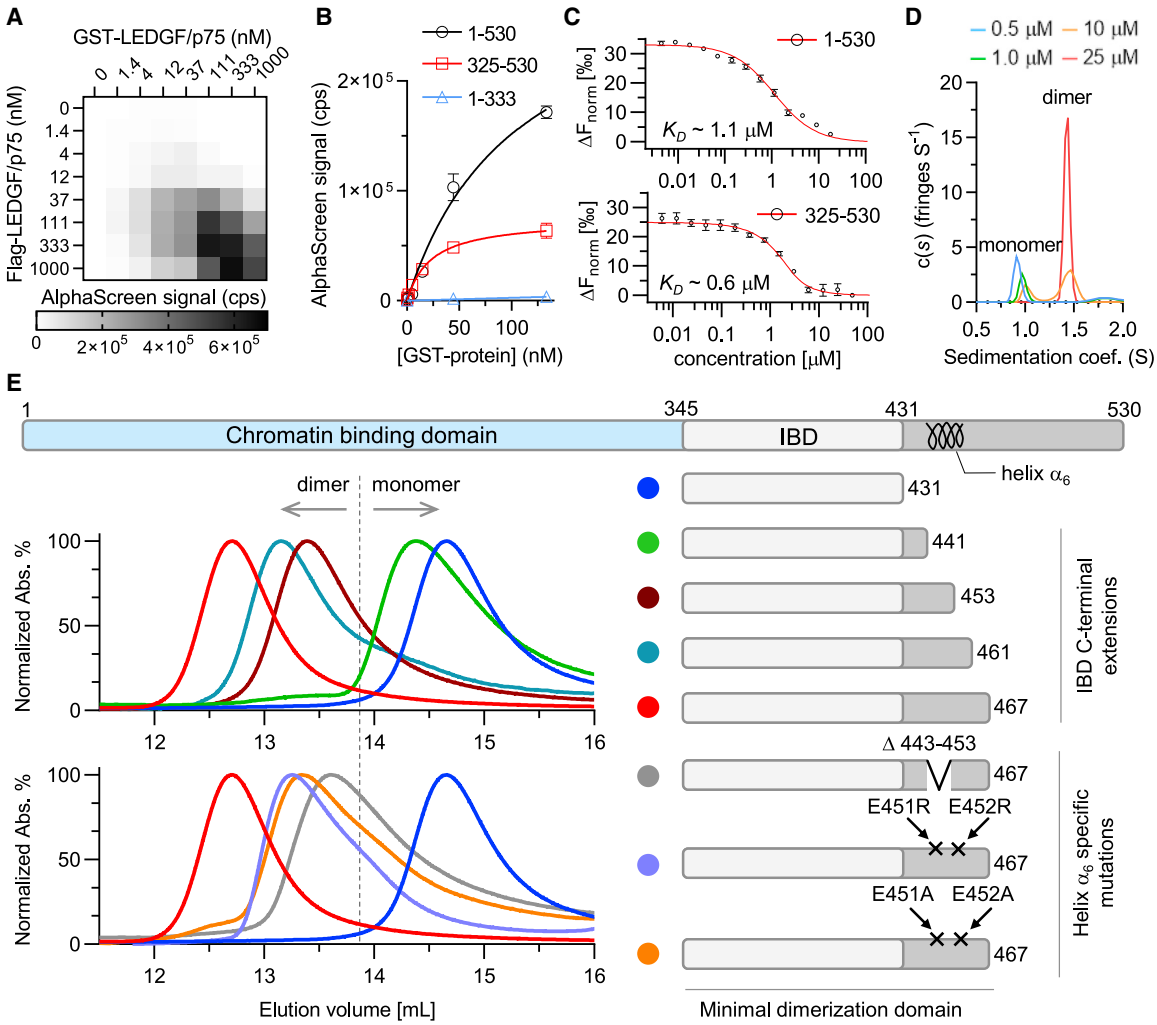


Figure 1. The C-terminal Interactions Contribute to LEDGF/p75 Dimerization

(A) AlphaScreen cross-titration between Flag-LEDGF/p75 and GST-LEDGF/p75 confirms dimerization.

(B) AlphaScreen binding assay of His_6-tagged LEDGF/p75 (200 nM) with GST-tagged variants (residues 1–530), p52 (1–333), and the C-terminal LEDGF construct (345–530), lacking the entire chromatin binding module, including the PWWP domain. The error bars represent the standard deviation of three measurements.

(C) MST titration yielded a K_D for dimerization of full-length LEDGF/p75 of 1.13 ± 0.05 μM and of 0.58 ± 0.04 μM for the C-terminal LEDGF construct (325–530), respectively. The error bars represent the standard deviation of three measurements.

(D) AUC with $LEDGF_{345-467}$ variable concentrations.

(E) The domain structure of LEDGF/p75 (top) and variants used in SEC (right). Color-coding matches the normalized size-exclusion chromatograms. The column calibration and the calculated apparent molecular weights of the eluting assemblies are shown in Figure S1.

sequence-specific manner across the entire protein. The central position of residue 409 enabled the monitoring of the positioning of the C terminus with respect to the IBD. The residues at positions 443 and 456 are flanking helix α_6, which was confirmed to play an important role in dimerization. The last modification was inserted at position 465 within the highly flexible C terminus as a negative control (Figure 2C). The naturally occurring cysteine 373 was mutated to alanine so that only one site per molecule is modified with the PRE probe. Residue 373 is hidden in the hydrophobic core of the IBD and this change did not affect the structure, as was confirmed by comparison of the wild-type and mutant 1D ^{1}H NMR spectra. We successfully prepared four variants containing S409C, S443C, T456C, and N465C mutations, respectively, in addition to C373A. All proteins were expressed and purified in expression media with a natural abundance or enriched for ^{15}N. The purification procedure was performed in the presence of a relatively high concentration of reducing agent (5 mM TCEP) to avoid unfavorable disulfide bridge formation. Interestingly, the S409C and T456C variants were more prone to forming covalent dimers than the other two cysteine variants, suggesting that these residues are close to one another in the dimer.

We formulated NMR samples for PRE measurements using two distinct approaches. First, we used ^{15}N $LEDGF_{345-467}$ cysteine variants and compared signal intensities in the 2D ^{15}N/^{1}H HSQC spectra of the paramagnetic and diamagnetic samples (Figure 2D). Reduced signal intensities indicate which intra- and intermolecular regions are close to the paramagnetic

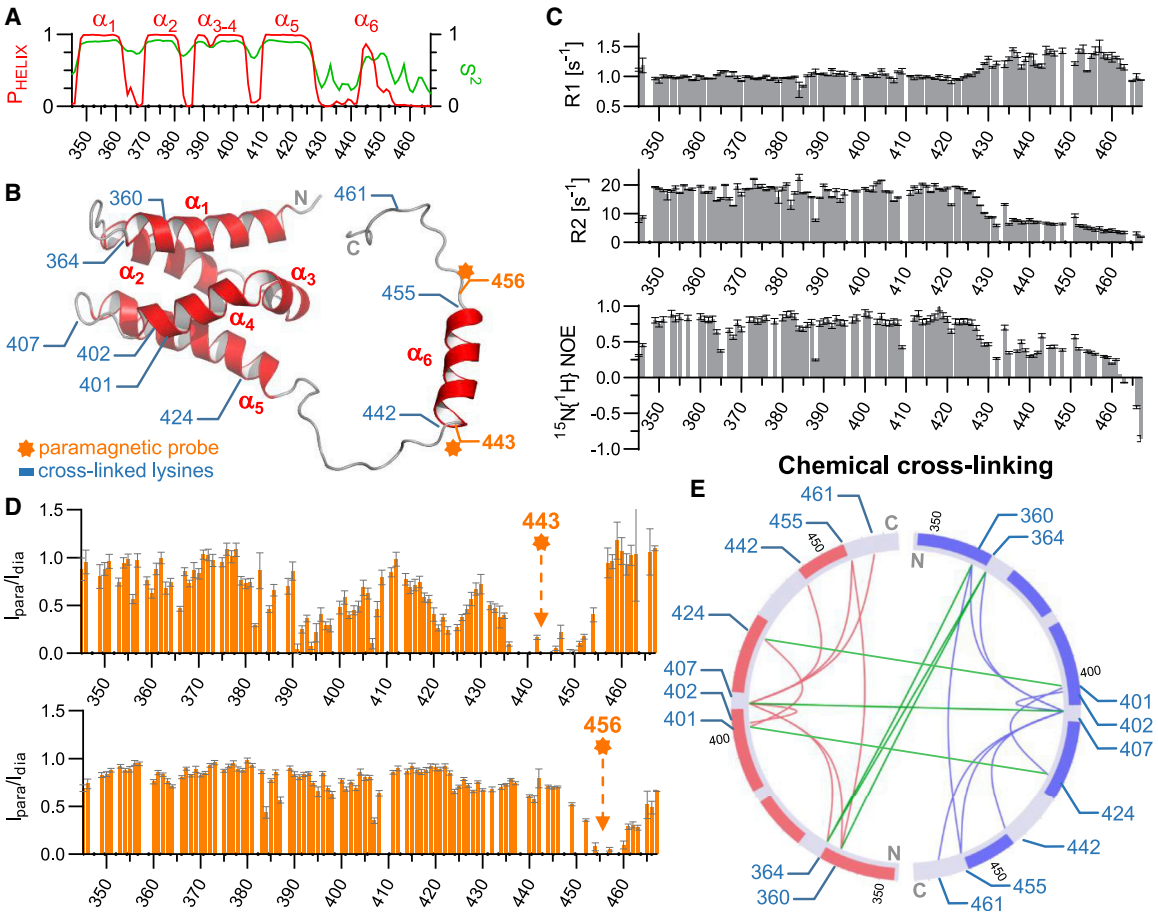


Figure 2. Dimer-Stabilizing Contacts

(A) Backbone NMR resonance assignment-derived α-helical propensity (P_{HELIX}; red) and order parameter (S^2; green) per residue (x axis).

(B) Model of $LEDGF_{345-467}$ highlighting the positions of α helices (red) within the monomer, residues involved in chemical crosslinking (blue), and residues to which the paramagnetic probe was coupled (orange asterisk and number).

(C) Backbone amide NMR relaxation rates (R1 and R2) and $\{^1H\}$–^{15}N steady-state NOE values obtained for $LEDGF_{345-467}$ residues (x axis).

(D) Residues affected by paramagnetic relaxation enhancement (PRE) in the $LEDGF_{345-467}$ dimer. The para- or diamagnetic probe was bound to a cysteine introduced at residues 443 or 456 (orange asterisk). R1 and R2 error values respresent the fitting error and NOE/PRE error bars represent error propagation from NMR signal intensities.

(E) Chemical crosslinking coupled to mass spectrometry to determine inter- (green lines) and intramolecular (blue or red lines) contacts within the $LEDGF_{345-467}$ dimer (each monomer is represented by a half-circle in red-gray or blue-gray).

nitroxide probe (12–29 Å). The paramagnetic probe at position 443 induced the strongest PRE effect (<0.5 intensity ratio) in four regions spanning residues 356–366, 380–410, 418–428, and 431–455 (Figure 2D). The probe at position 456 strongly affected neighboring residues in the primary sequence (445–467) and moderately affected several individual residues (365, 384, 387, 394, 395, 398, 407, and 408) and a stretch of residues at the boundary between the IBD and C-terminal tail (424–435) (Figure 2D). Introducing any probe at residue S409C caused a massive broadening of signals in 2D $^{15}N/^1H$ HSQC preventing a quantitative evaluation. The probe at residue 465 had no effect on neighboring residues.

In the second approach, only the variants prepared in media with a natural abundance of ^{15}N were modified with the para- or diamagnetic probe and then mixed in a 1:1 molar ratio with the ^{15}N-enriched $LEDGF_{345-467}$ C373A variant. Only signals from the ^{15}N-labeled protein, which cannot be modified with the probe, were visible in 2D $^{15}N/^1H$ HSQC. However, the PRE effect can, in the best-case scenario, be effective only by 50% of dimers, as ^{15}N molecules are evenly distributed between purely diamagnetic $^{15}N/^{15}N$ dimers and paramagnetic mixed $^{15}N/^{14}N$ dimers. Moreover, the insufficient exchange of monomers into $^{15}N/^{14}N$ dimers upon mixing the two $LEDGF_{345-467}$ species can lead to an even more reduced PRE response. This setup allowed us to visualize solely intermolecular dimeric interactions (Figure S2). Probes at positions 409, 443, and 456 are in the vicinity (12–29 Å) of residues 363–373, 398, and 405–410. Additional weak effects can be detected for residues 425–430. Again, the probe at residue 465 had only negligible effects.

Chemical crosslinking coupled with mass spectrometry was used as an independent method to determine inter- and intramolecular contacts within the $LEDGF_{345-467}$ dimer. To differentiate between the monomers in the dimer, ^{14}N- and ^{15}N-labeled $LEDGF_{345-467}$ molecules were mixed in a 1:1 molar ratio and

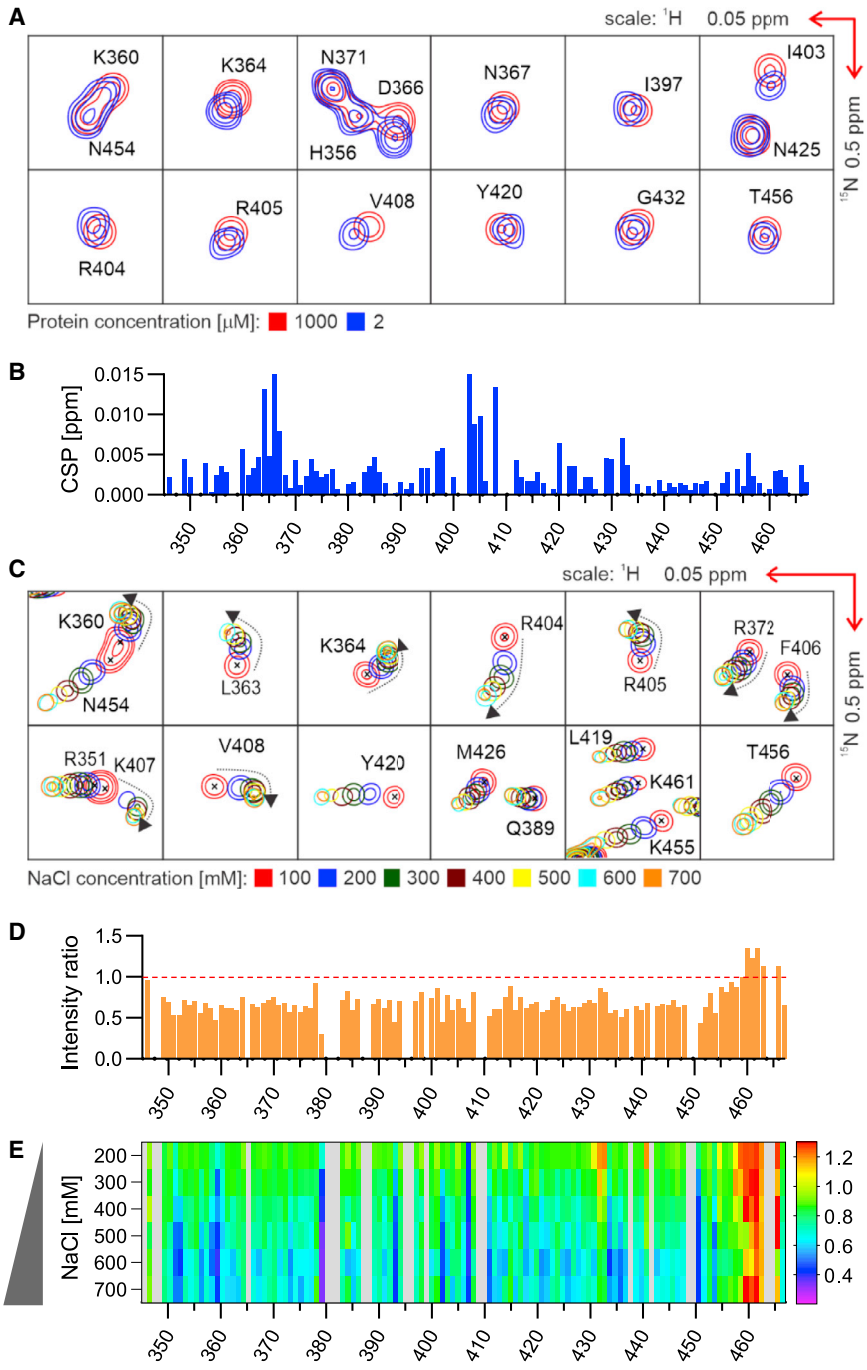


Figure 3. Protein Concentration and Ionic Strength Effects

(A) Comparison of the signals of selected residues in 2D $^{15}N/^{1}H$ HSQC spectra at 2 μM (blue) and 1,000 μM (red) $LEDGF_{345-467}$ concentrations.

(B) Chemical shift perturbations of $LEDGF_{345-467}$ residues (x axis) upon dilution from 1,000 to 2 μM.

(C) Examples of two sorts of behavior in 2D $^{15}N/^{1}H$ HSQC spectra upon increasing NaCl concentration (100–700 mM).

(D) Change in signal intensity in 2D $^{15}N/^{1}H$ HSQC spectra; comparison of 700 and 100 mM NaCl.

(E) Heatmap demonstrating signal intensity of each residue at all tested NaCl concentrations relative to signal intensity in 2D $^{15}N/^{1}H$ HSQC spectra in a buffer with 100 mM NaCl.

crosslinked with DSG (disuccinimidyl glutarate). In analogy to samples prepared for PRE measurements, the expected effective abundance of $^{15}N/^{14}N$ mixed dimers was reduced to 50%. Crosslinked peptides were digested with trypsin, analyzed by mass spectrometry, evaluated in LinX, and visualized in XiView (Graham et al., 2019) (Figures 2E and S3A–S3C). Dimeric contacts within the IBD were only detected between residues 360–364, 364–364, 364–407, 401–424, and 407–407. Intramolecular crosslinks were more abundant, in many cases within the IBD, but several linked K407 to 442, 455, and 461 within the C-terminal tail. These results are in a good agreement with the PRE analysis.

Dimer Formation Is Supported by Electrostatic Interactions

Next, we used NMR spectroscopy to follow changes induced by dissociation of dimers. First, we measured a series of 2D $^{15}N/^{1}H$ HSQC spectra at various protein concentrations (1,000–2 μM) and analyzed chemical shift perturbations (Figures 3A and 3B). As expected, the perturbations detected were modest and indicated that only a small fraction of the dimer is dissociated. A substantial enrichment of the monomeric fraction might be possible at a significantly lower concentration, which would be below detection level of NMR. The most affected regions consisted of residues 360–370 and 403–408. In addition, residues in the C-terminal region of the IBD (420–433) were moderately influenced by dilution. Interestingly, all three affected regions coincide with the binding site for LEDGF/p75 IBD interaction partners.

Considering that several positively charged residues at the C terminus of helix α_4 were among the residues most affected by dilution, we tested whether dimerization is supported by electrostatic interactions. In particular, we followed the effects of ionic strength on 2D $^{15}N/^{1}H$ HSQC spectra. During NaCl titrations, amide signals showed two distinct responses to increasing salt concentration (100–700 mM; Figures 3C and S3D). One subset of signals, from residues primarily located at the N terminus of $LEDGF_{345-467}$ or in the C-terminal region, followed a linear trajectory, while peaks in a second subset followed curved trajectories. Such a non-linear response could be a superposition of two linear trends—an effect of increasing the ionic strength, which is analogous to the movement observed for the first subset of signals; and an additive effect of dimer dissociation following the trend observed in the dilution experiments described above that is obvious in the last two to three titration

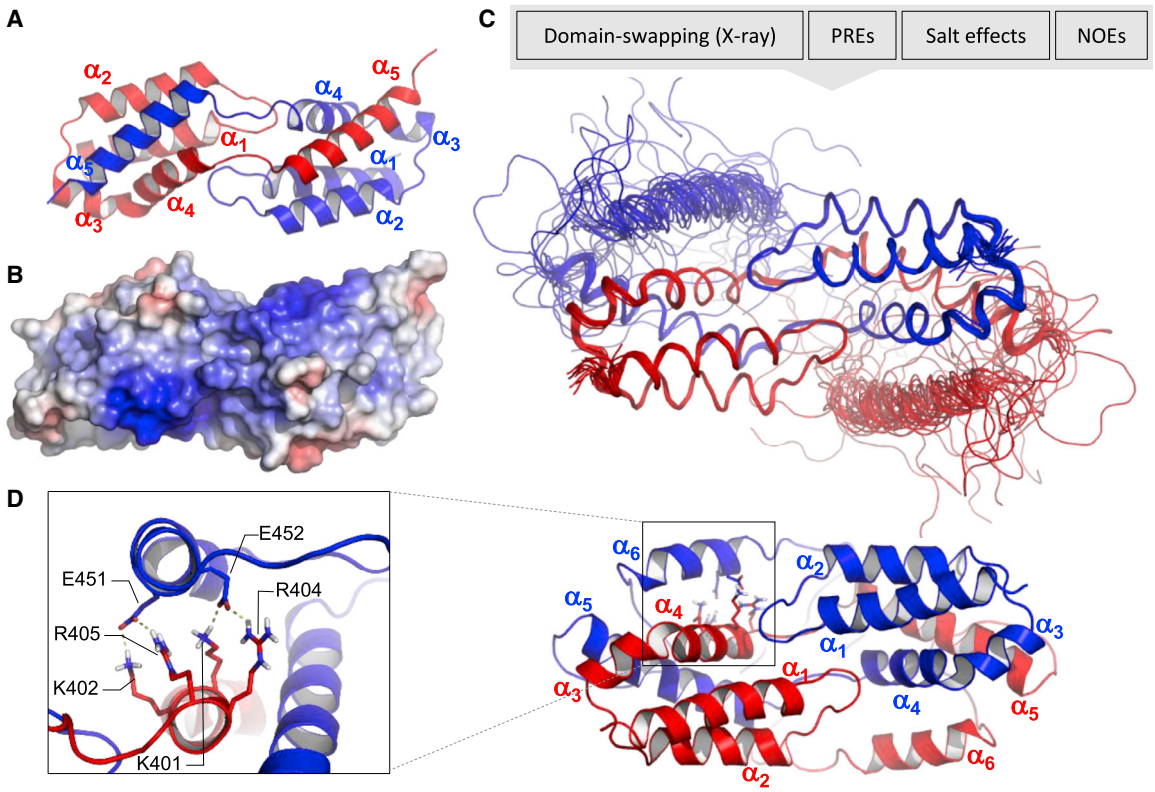


Figure 4. Structure of the LEDGF$_{345-467}$ Dimer

(A) Crystal structure of the LEDGF/p75 IBD domain-swapped dimer. The residues connecting helices α_4 and α_5 are fully extended instead of forming a loop. Such an orientation leads to a dimeric assembly stabilized by domain swapping, with helices α_5 exchanged between the monomers.

(B) Surface electrostatics representation of the LEDGF/p75 IBD dimer reveals two symmetrically positioned positively charged patches in the dimer surface. The positively charged areas are highlighted in blue, neutral regions in white, and negatively charged regions in red.

(C) A set of 20 converged structures obtained for the LEDGF$_{345-467}$ dimer using the crystallography-derived restraints for the domain-swapped dimer, NMR-based restraints (NOEs, dihedral restraints calculated form the backbone chemical shift assignments, PRE, and salt effects).

(D) The domain-swapped assembly is further stabilized by a network of electrostatic contacts between the negatively charged side chains of E451/E452 from helix α_6 and the positively charged side chains from K401, K402, R404, and R405 (helix α_4) (enlarged view).

steps with NaCl. This behavior is most apparent for K360, L363, K364, R372, R404, R405, F406, K407, and V408 (Figure 3C). The increasing NaCl concentration led to a gradual decrease in signal intensity for most of the residues except for 455–467 at the C-terminal end (Figures 3D, 3E, and S3D). Their intensity either increased or remained unchanged, which suggests that these residues are significantly more flexible in high salt concentration, perhaps due to dimer dissociation driven primarily by salt-dependent destabilization of electrostatic interactions.

Structure of the LEDGF$_{345-467}$ Dimer

The distribution of intra- and intermolecular PRE supported by dilution and salt titration experiments indicated that the LEDGF$_{345-467}$ dimer is stabilized by domain swapping facilitated by helix α_5 from the IBD and the more dynamic C-terminal helix α_6. We initially tried to structurally characterize the LEDGF$_{345-467}$ dimer using X-ray crystallography, but this effort was unsuccessful due to inherent dynamics of several regions of the construct. However, we obtained high-quality crystals for the IBD (residues 345–431) that yielded a domain-swapped dimeric structure highly similar to one published during the course of this study (Figure 4A; Table S2) (Hannon et al., 2018). The fact that we obtained reproducible results using a different protein construct indicates that swapping of helix α_5 is not a crystallographic artifact, as might be inferred from an estimated dimer dissociation constant in a millimolar range (Figures 1C and 1D) (Hannon et al., 2018). The swapping of helix α_5 between two monomers preserves all the key structural features of the monomer, except the straightening of the hinge region between helices α_4 and α_5.

The results obtained using PRE were converted into distance restraints that, in combination with NOESY-derived restraints, allowed for calculation of the LEDGF$_{345-467}$ dimer structure. Because of the differential local flexibility between the IBD and the C-terminal part of the molecule, we applied relatively less stringent 15 Å PRE-derived restraints only to residues showing more than 50% signal reduction in the presence of the paramagnetic probe. In addition, we confirmed that there were no distance restraint violations >0.5 Å observed for the IBD crystallographic data. Therefore, we generated a set of additional C^{α}-C^{α} long-distance restraints for residues from regions of regular secondary structure (helices α_1 to α_5) that were within 8 Å using the crystallographic data obtained for IBD to maintain the domain-swapped conformation with the IBD during NMR structural refinement.

To understand how this dimeric structure behaves in solution, we completed the side-chain NMR resonance assignments and calculated a preliminary set of structures for the $LEDGF_{345-467}$ dimer (Figure 4C). It revealed that the C-terminal helix α_6 downstream of the IBD is in a parallel orientation relative to the swapped helix α_5 from the IBD core, further stabilizing the dimeric assembly through intermolecular contacts with helix α_4 from the second monomer (Figure 4C). Interestingly, the IBD surface at the site of interaction with helix α_6 is formed by a cluster of positively charged residues (K401, K402, R404, and R405; large blue patches at top and bottom, Figure 4B) that are accessible for electrostatic interactions with two closely spaced negatively charged glutamate residues (E451 and E452) within helix α_6 (Figure 4D). Based on this observation, we designed two modified $LEDGF_{345-467}$ constructs with both glutamates swapped for either charge-reverting arginines or neutral alanines ($LEDGF_{345-467}$ E451R/E452R or E451A/E452A) and, in analogy to the helix α_6 deletion construct, investigated their dimerization properties using size-exclusion chromatography (Figure 1E, bottom part). Both constructs exhibited a significantly reduced capacity to elute from the column exclusively as dimers, highlighting the importance of these residues in the electrostatic stabilization of dimers. Therefore, in the final refinement we added distance restraints for two pairs of hydrogen bonds that stabilize this interaction (E451-K402/R405 and E452-K401/R404; Figure 4D) and calculated the set of water-refined structures for the $LEDGF_{345-467}$ dimer. Finally, we used data from chemical crosslinking for successful validation of the structural data (Figure S3C). Most C^{α}-C^{α} distances for crosslinked lysines were in the final set of structures within the 15 Å range, including all intermolecular and two intramolecular pairs (K360-K407 and K402-K407). The remaining intramolecular C^{α}-C^{α} distances, which included lysines from the more dynamic C-terminal region, were found distributed over broader ranges. In particular, K407 was separated from K442, K455, or K462 by 26.5–30.3, 14.0–18.5, or 12.4–31.1 Å, respectively. Similarly, the intramolecular distance between K360 and K455 was between 17.7 and 25.0 Å. The only crosslink that is not in good agreement with our structural data is an intramolecular link between K401 and K424 (~37 Å). Overall, the crosslinking data confirmed the domain-swapped arrangement of the dimer. Also, the overlay of the X-ray LEDGF/p75 IBD dimer and the representative NMR dimer illustrates the good agreement between the structures (Figure S4A). We further validated the agreement between the experimental PRE data and structures calculated for the $LEDGF_{345-467}$ dimer by a simulation of PRE effects induced by the paramagnetic probe attached to cysteines at positions 443 and 456, respectively (Figure S4B). As expected, the worst agreement was observed in the loop regions connecting the helices, perhaps due to an insufficient sampling of the conformational space during structure calculations.

$LEDGF_{345-467}$ Dimerization Does Not Affect the Interactome at the Molecular Level

In our previous work, we showed that a number of cellular binding partners of LEDGF/p75 interact with the IBD through an intrinsically disordered IBM (Cermakova et al., 2014; Sharma et al., 2018; Wu et al., 2002). Here, we investigated the effect of LEDGF/p75 dimerization on the interaction with three partners: MLL1, JPO2, and POGZ. The structural data for the $LEDGF_{345-467}$ dimer (Figure 4C) reveal that the dimer-stabilizing helix α_6 partially occludes the interaction surface utilized by helix 1, a common feature for all three binding partners (Sharma et al., 2018). Therefore, we expected that, in the case of a partner binding to the dimer, the C-terminal tail bearing helix α_6 must detach from the IBD core helices to make space for the partner. We titrated the ^{15}N-labeled $LEDGF_{345-467}$ dimer with unlabeled $MLL1_{123-160}$, $JPO2_{2-130}$, or $POGZ_{1117-1410}$, and followed the resulting changes in the 2D $^{15}N/^{1}H$ HSQC spectra (Figures 5A, 5C, 5D, S5, and S6). The gradual changes induced in the $LEDGF_{345-467}$ spectra by the MLL1-derived peptide indicate the "fast-exchange" regime (Figures 5A and S5) expected for the ~50 μM binding affinity reported for the monomeric IBD (Cermakova et al., 2014; Sharma et al., 2018; Wu et al., 2002). In the case of JPO2 and POGZ binding, most signals from the LEDGF/p75 interacting residues disappeared from the spectra at substochiometric concentrations and re-appeared at their "bound" positions in the presence of an excess of the partner protein (Figures 5C, 5D, and S8). Such "intermediate" to "slow" exchange regimes confirm the low micromolar binding affinity suggested by JPO2- and POGZ-derived peptides (Sharma et al., 2018; Wu et al., 2002).

The chemical shift perturbation graphs show that all partners interact with three stretches of residues (360–370, 400–410, and 420–430) and reveal that positions of signals from the C-terminal residues, including helix α_6, are not significantly affected by binding. However, there was an increase in signal intensities (2- to 4-fold) in this region in comparison with signals from the IBD residues, which suggests that the C terminus is liberated from the dimer upon partner binding. To confirm this hypothesis, we titrated the dimerization-deficient variant $LEDGF_{345-467}$ E451R/E452R with MLL1-derived peptide and quantified the effects on the NMR spectra (Figure 5B). While MLL1 binding affected the positions of similar residues, the C-terminal signal intensities did not increase as for the wild-type protein. Interestingly, the flexible negatively charged loops in all interaction partners contribute to complex formation by binding the same positively charged residues (K401, K402, R404, and R405; Figure 4B) (Sharma et al., 2018) that are involved in electrostatic dimer stabilization. The signal intensity ratio graphs also reveal that MLL1-derived peptide binding led to a moderate reduction in the IBD signal intensity, while the significantly larger JPO2 and POGZ protein domains induced an approximately 50% drop in 2D $^{15}N/^{1}H$ HSQC signal intensity in this region, reflecting the larger size of the LEDGF-JPO2 and LEDGF-POGZ complexes (Figure 5).

DISCUSSION

Tethering of various transcriptional regulators to the chromatin in a physiological as well as pathological context is one of the key biological roles attributed to LEDGF/p75 (for a review see Cermakova et al., 2016). Most known LEDGF/p75 interacting proteins are either constitutive dimers or contain two copies of IBMs (Sharma et al., 2018; Wu et al., 2002). In addition, nucleosomes can carry a pair of specific epigenetic marks, and related downstream processes might benefit from a "binary" organization of reader molecules. Several reports have already

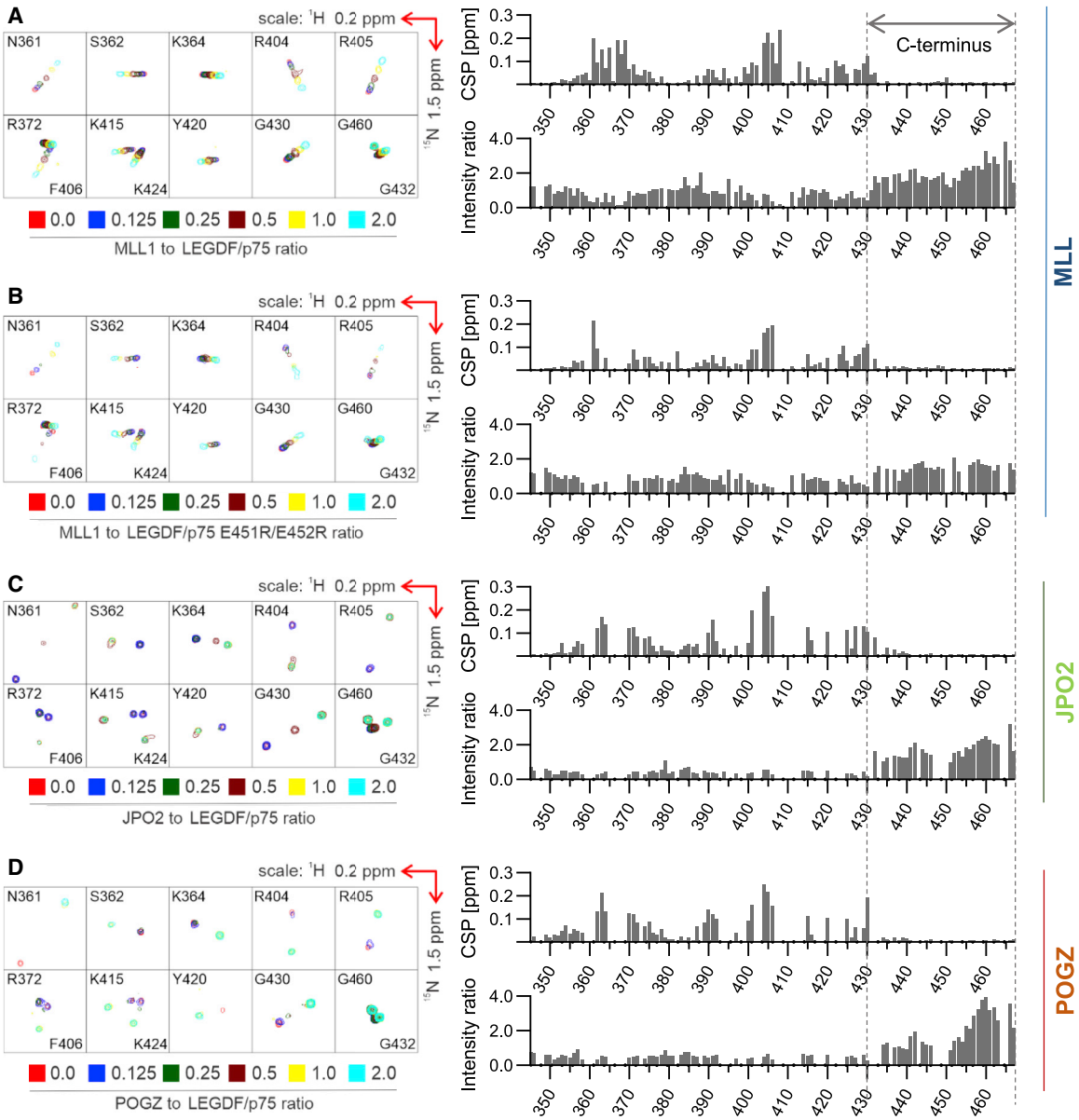


Figure 5. Interaction of LEDGF$_{345-467}$ with Binding Partners

(A) Representative regions from the 2D ^{15}N/^{1}H HSQC titration spectra illustrating effects of MLL1$_{123-160}$ additions on LEDGF$_{345-467}$ signals (left). The 50 μM ^{15}N-labeled LEDGF$_{345-467}$ protein construct was titrated with unlabeled peptide. Chemical shift perturbations of LEDGF$_{345-467}$ residues upon binding of MLL1$_{123-160}$ peptide at a 1:2 ratio (top right); signal intensity ratio of MLL1-bound versus free LEDGF$_{345-467}$ (bottom right).

(B) Data for MLL1$_{123-160}$ binding to the LEDGF$_{345-467}$ E451R/E452R variant.

(C and D) Data for JPO2$_{2-130}$ and POGZ$_{1117-1410}$ binding to LEDGF$_{345-467}$.

suggested that LEDGF/p75 forms a dimer in solution (Michel et al., 2009; Vanderlinden et al., 2014), which was not in agreement with published data (Cherepanov et al., 2005). To clarify this inconsistency, we addressed the following questions: (1) Is LEDGF/p75 capable of dimer formation? (2) What is the minimal domain responsible for dimerization? (3) What is its molecular mechanism?

We demonstrated that full-length LEDGF/p75 can form dimers with a low μM dissociation constant and identified the C-terminally extended IBD (residues 345–467) as the minimal dimerization region (Figure 1). Although we could also observe dimers for the IBD (residues 345–431) at mM concentrations (Figure S1C), the LEDGF$_{345-467}$ construct eluted from the size-exclusion column exclusively as a dimer over a broad concentration range. The LEDGF$_{345-467}$ NMR resonance assignment revealed an additional α helix (Figure 2A) downstream of the IBD that is more dynamic than the IBD helices, as suggested by analysis on NMR relaxation parameters. The C-terminal helix is essential for dimer integrity, as its deletion led to dimer destabilization (Figure 1E).

We observed a single set of resonances for every residue in NMR spectra collected for $LEDGF_{345-467}$, which suggested that the molecule forms a symmetric dimer. To get more insight into dimer organization, we utilized paramagnetic NMR spectroscopy. We generated four $LEDGF_{345-467}$ constructs carrying a single cysteine at different positions that we subsequently modified with either a diamagnetic or paramagnetic probe to assess PRE effects proportional to the distance from the paramagnetic center (Figures 2B and S2). We detected PREs either within the entirely ^{15}N-labeled dimer carrying either a paramagnetic or a diamagnetic probe (Figure 2D), or in a sample prepared by mixing the ^{15}N-labeled unmodified protein and unlabeled protein with a probe (Figure S2). Overall, the PRE effects in mixed samples, which provided information only about intermolecular contacts, were significantly weaker than the effects observed in ^{15}N-labeled proteins directly carrying the probe (Figures 2D and S2). This observation can be explained by a slow equilibrium of the monomer re-distribution in the context of the dimer that led to a relatively small population of assemblies consisting of the ^{14}N- and ^{15}N-labeled pairs, which could be theoretically formed by only 50% of the molecules present in the samples. The lower stability of cysteine-containing $LEDGF_{345-467}$ variants and probe instability did not allow for longer incubation times that would lead to higher abundance of mixed ^{14}N- and ^{15}N-labeled dimers. The distribution of PREs revealed that helix α_6 from the C terminus interacts with the IBD in the context of the dimer, with both intra- and intermolecular contacts (Figures 2D and S2). In addition, the IBD loop connecting helix α_4 and helix α_5 (residues 405–410) plays a crucial role in stabilization of intermolecular contacts.

LEDGF/p75 consists of a relatively high number of solvent-exposed lysine residues that are accessible for crosslinking agents. We identified crosslinked peptides obtained from samples of an equimolar mixture of ^{15}N-labeled and unlabeled $LEDGF_{345-467}$ constructs. The intramolecular crosslinks were more frequent than the intermolecular ones, because only half of the dimers can be formed by the mixture of ^{14}N- and ^{15}N-labeled molecules. However, the data fully supported the PRE experiments, with the ''back-folding'' of the C-terminal region and the IBD and unambiguous intermolecular contacts within the IBD loops between helices α_1 and α_2 as well as helices α_4 and α_5 (Figure 2E).

We investigated dimer disassembly using an NMR dilution experiment (Figures 3A and 3B). Considering the low μM K_D for this equilibrium (Figure 1C), we could only observe minor changes in the positions of signals between 1- and 2-μM samples, reflecting the presence of only a small fraction of free-state molecules in the diluted sample. Despite this limitation, dilution significantly affected the IBD loops connecting helices α_1 and α_2 as well as helices α_4 and α_5 (Figure 3B). Moderate effects were observed in helices α_5 and α_6, which supports earlier data obtained from chemical crosslinking and PRE experiments. Although we showed that helix α_6 contributes to enhanced dimerization, only the NMR salt titration experiments shed light on the nature of its role in dimer assembly. In general, increasing salt concentration affects the positions of backbone amide proton signals in NMR spectra due to changes in bulk magnetic susceptibility, electrostatic interactions, and van der Waals forces, and hydrogen bond formation between the solute and solvent molecules (Kukic et al., 2013). In the case of $LEDGF_{345-467}$, the gradual increase in salt led to both linear and non-linear changes in the positions of backbone amide proton signals (Figure 3C). Linear changes were attributed to expected bulk effects, but the non-linear behavior indicated an additional, ionic strength-dependent equilibrium in which dimer dissociation was caused by the weakening of electrostatic interactions. The strongest non-linear effects were again observed for residues within the IBD loops between helices α_1 and α_2 and helices α_4 and α_5. Interestingly, a pair of neighboring arginines (R404 and R405) was also affected, indicating their contribution to dimer-stabilizing electrostatic interactions. NMR sensitivity drops significantly with increasing ionic strength, especially when using cryoprobes. Indeed, this trend was observed for most of the $LEDGF_{345-467}$ signals (except for residues 455–46), perhaps due to increasing mobility that compensated for the loss in sensitivity (Figures 3D, 3E, and S3D). Our data revealed that the electrostatic interactions involving C-terminally located residues contribute significantly to $LEDGF_{345-467}$ dimer stability.

Domain swapping is relatively frequent in proteomes and was proposed to be one of the molecular mechanisms for evolution of oligomeric states of proteins (Bennett et al., 1994). The IBD domain-swapped conformational arrangement is in agreement with the biochemical (Figure 1E) and biophysical (Figures 2D and 2E) data obtained during this study. We combined restraints derived from the X-ray structure of the IBD swapped dimer, PREs, and NOESY spectra for determination of the 3D structure of the dimer. The resulting structure revealed that the domain swapping is maintained by helices α_5 and α_6 (Figure 4C). Moreover, a pair of negatively charged glutamates from helix α_6 was found to be well positioned for salt bridge formation with the positively charged patch on the IBD surface (Figure 4B). The importance of this electrostatic interaction was validated by introducing charge-modifying mutations that led to dimer destabilization (Figure 1E) due to electrostatic repulsion between the C terminus and positively charged residues in helix α_4.

The pocket formed by contacts between helices $\alpha_3/\alpha_4/\alpha_5$ and the positively charged region at the C terminus of helix α_5 within the IBD (Figures 4A and 4B) are the key interaction sites recognized by LEDGF/p75 cellular binding partners (Sharma et al., 2018). Interestingly, these features remain structurally unperturbed by domain swapping, although the positive side chains in helix α_4 contribute to the electrostatic stabilization of helix α_6 (Figure 4D). Using NMR, we tested binding of a short monomeric MLL1-derived peptide and two larger JPO2- and POGZ-derived protein constructs that represented dimeric arrangement of IBMs (Figure 5). Although these partners bound with different affinities and molar ratios, their interaction did not lead to dimer dissociation but only to detachment of the C-terminal region, including helix α_6, as the exposed positive patch on the IBD surface (K401, K402, R404, and R405) was required for electrostatic stabilization of newly established complexes. Surprisingly, the dimeric arrangement of $LEDGF_{345-467}$ does not rule out the possibility of LEDGF/p75 dimerization in the context of the MLL1/MENIN/LEDGF ternary complex (Huang et al., 2012), as there are no steric clashes between MLL1/MENIN assemblies binding to the LEDGF/p75 dimer (Figure 6A). Such an arrangement might lead to a stronger biological response upon ternary complex formation. We can envisage an analogous organization for the

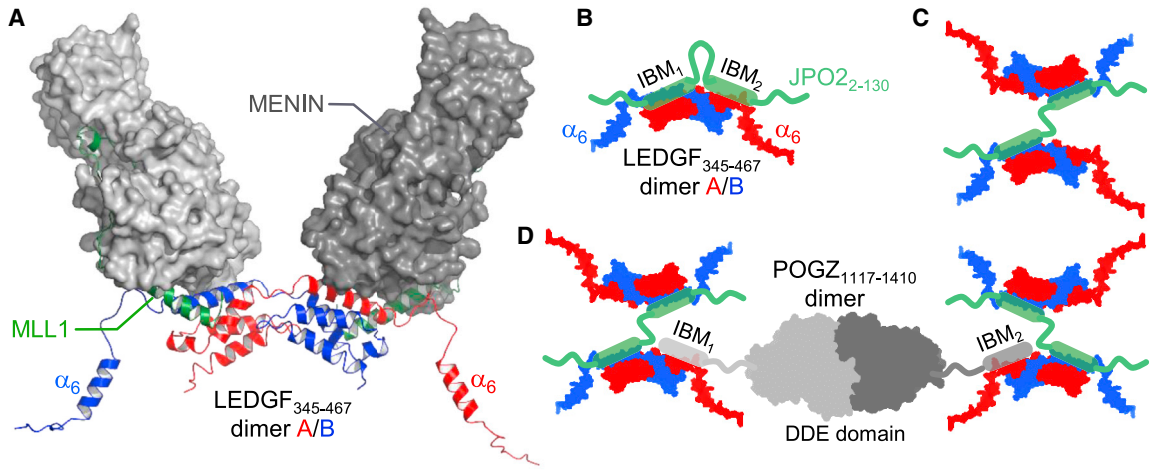


Figure 6. Accessible Binding Modes for LEDGF/p75 IBD Interacting Partners

(A) The dimeric organization of $LEDGF_{345-467}$ does not interfere with the MLL1/MENIN/LEDGF ternary complex assembly (Huang et al., 2012). Each half of the LEDGF/p75 dimer (red/blue) can accommodate one MLL1/MENIN binary complex (green/gray). Two IMBs from $JPO2_{2-130}$ (green) can bind to a single LEDGF/p75 dimer (B) or can "crosslink" two (C) or more LEDGF/p75 molecules by utilizing additional partners, such as $POGZ_{1117-1410}$ (gray) (D).

JPO2 and POGZ complexes in which a LEDGF/p75 dimer can bridge the "single-chain" two-motif JPO2 molecule or a constitutive POGZ dimer (Figure 6B). Alternatively, two motifs from JPO2 or POGZ could "crosslink" two or more LEDGF/p75 dimer molecules because of the relatively small size of their IBMs (Figures 6C and 6D). Indeed, this is highly speculative as there are no published *in vivo* experiments confirming the capacity of the LEDGF/p75 dimer to accommodate two different binding partners. This might be beneficial in the context of nucleosomes carrying a pair of H3K36me marks that can simultaneously accommodate two copies of LEDGF/p75, as revealed by recent cryoelectron microscopy data capturing the LEDGF/p75-nucleosome interaction (Wang et al., 2020). Unfortunately, most of LEDGF/p75 (residues 94–530) is not resolved, perhaps due to the highly dynamic properties of the modular LEDGF/p75 regions downstream of the PWWP domain. There is also a functional link with LEDGF/p75 dimerization, which indicates that at least two molecules of LEDGF/p75 or its ortholog HRP2 are required per nucleosome for efficient transcription *in vitro* (LeRoy et al., 2019). The LEDGF/p75 dimer could act as a molecular "glue" crosslinking the same or even different interaction partners and amplifying the related epigenetic signals.

Conclusions

Various reports point to LEDGF/p75 dimerization as an important molecular feature of this epigenetic reader. However, no direct evidence for LEDGF/p75 dimerization either *in vivo* or in a cellular context has been reported to date. Here, we have uncovered the molecular basis of dimer formation using NMR spectroscopy in combination with biochemical and biophysical techniques, including size-exclusion chromatography and mass spectrometry. We narrowed down the minimal dimerization domain to amino acid residues 345–467, which contain the structured protein binding platform IBD and a dynamic C-terminal region containing an additional helix responsible for dimer stabilization through electrostatic interactions. Furthermore, we successfully determined the structure of the $LEDGF_{345-467}$ dimer using experimentally derived NMR restraints, including PRE and NOESY data, and used chemical crosslinking and structure-inspired LEDGF variants for validation. $LEDGF_{345-467}$ dimerization does not affect the protein's capacity to interact with cellular partners *in vitro*. Moreover, it might be an important mechanism to locally increase concentration of interacting factors which enhance active transcription.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Material Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cloning
  - Protein Expression and Purification
  - AlphaScreen
  - Microscale Thermophoresis
  - Analytical Ultracentrifugation
  - Chemical Cross-Linking and Mass Spectrometry
  - NMR Spectroscopy
  - Para and Diamagnetic Spin-Labelling
  - Paramagnetic Relaxation Enhancement (PRE) Measurements
  - X-ray Crystallography
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2020.08.012.

ACKNOWLEDGMENTS

We thank Edward Curtis and Pavlína Řezáčová for valuable discussions. We would also like to acknowledge Dieter Braun and Evgeniia V. Edeleva for training with the Monolith NT.115 (Nanotemper). This work was supported by the GACR (grant no. 16-06357S to V.V. and 16-24309S to P.N.), the Flemish FWO (G.0595.13 to Z.D.), the Ministry of Education of the Czech Republic (LO1304), the European Regional Development Fund; OP RDE; the ChemBioDrug project (no. CZ.02.1.01/0.0/0.0/16_019/0000729), the European Commission H2020 (European Network of Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry Centers—project agreement no.731077) and KU Leuven through an IDO grant (to T.B.) and Herculesstichting 2016. We also acknowledge the Center of molecular structure Core Facility at BIOCEV, funded by RDE Funds (CZ.1.05/1.1.00/02.0109 BIOCEV) and supported by the Czech Infrastructure for Integrative Structural Biology (LM2015043 CIISB for CMS BIOCEV).

AUTHOR CONTRIBUTIONS

Conceptualization, V.L., T.B., K.Č., Z.D., and V.V.; Methodology, V.L., T.B., K.Č., Z.D., and V.V.; Investigation, V.L., T.B., K.Č., P.S., M.F., M.M., M.H., Z.K., P.N., M.K., J.B., J.D.R., and F.C.; Writing – Original Draft, V.L., T.B., K.Č., P.S., Z.D., and V.V.; Writing – Review & Editing, V.L., T.B., P.S., K.Č., Z.D., and V.V.; Funding Acquisition, V.V. and Z.D.; Supervision, V.V. and Z.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 30, 2020
Revised: July 19, 2020
Accepted: August 28, 2020
Published: September 17, 2020

REFERENCES

Bartholomeeusen, K., Christ, F., Hendrix, J., Rain, J.C., Emiliani, S., Benarous, R., Debyser, Z., Gijsbers, R., and De Rijck, J. (2009). Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ. J. Biol. Chem. *284*, 11467–11477.

Bartholomeeusen, K., De Rijck, J., Busschots, K., Desender, L., Gijsbers, R., Emiliani, S., Benarous, R., Debyser, Z., and Christ, F. (2007). Differential interaction of HIV-1 integrase and JPO2 with the C terminus of LEDGF/p75. J. Mol. Biol. *372*, 407–421.

Bax, A., Clore, G.M., and Gronenborn, A.M. (1990). H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13-enriched proteins. J. Magn. Reson. *88*, 425–431.

Bennett, M.J., Choe, S., and Eisenberg, D. (1994). Domain swapping: entangling alliances between proteins. Proc. Natl. Acad. Sci. U S A *91*, 3127–3131.

Brautigam, C.A. (2015). Calculations and publication-quality illustrations for analytical ultracentrifugation data. Methods Enzymol. *562*, 109–133.

Bueno, M.T., Garcia-Rivera, J.A., Kugelman, J.R., Morales, E., Rosas-Acosta, G., and Llano, M. (2010). SUMOylation of the lens epithelium-derived growth factor/p75 attenuates its transcriptional activity on the heat shock protein 27 promoter. J. Mol. Biol. *399*, 221–239.

Cermakova, K., Tesina, P., Demeulemeester, J., El Ashkar, S., Mereau, H., Schwaller, J., Rezacova, P., Veverka, V., and De Rijck, J. (2014). Validation and structural characterization of the LEDGF/p75-MLL interface as a new target for the treatment of MLL-dependent leukemia. Cancer Res. *74*, 5139–5151.

Cermakova, K., Weydert, C., Christ, F., De Rijck, J., and Debyser, Z. (2016). Lessons learned: HIV points the way towards precision treatment of mixed-lineage leukemia. Trends Pharmacol. Sci. *37*, 660–671.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Cherepanov, P., Ambrosio, A.L., Rahman, S., Ellenberger, T., and Engelman, A. (2005). Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc. Natl. Acad. Sci. U S A *102*, 17308–17313.

Cherepanov, P., Devroe, E., Silver, P.A., and Engelman, A. (2004). Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. J. Biol. Chem. *279*, 48883–48892.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. *278*, 372–381.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallographica. D Biol. Crystallogr. *60*, 2126–2132.

Ferrage, F. (2012). Protein dynamics by (1)(5)N nuclear magnetic relaxation. Methods Mol. Biol. *831*, 141–163.

Ferrage, F., Reichel, A., Battacharya, S., Cowburn, D., and Ghose, R. (2010). On the measurement of (1)(5)N-{(1)H} nuclear Overhauser effects. 2. Effects of the saturation scheme and water signal suppression. J. Magn. Reson. *207*, 294–303.

Ge, H., Si, Y., and Roeder, R.G. (1998). Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. EMBO J. *17*, 6723–6729.

Graham, M.J., Combe, C., Kolbowski, L., and Rappsilber, J. (2019). xiView: a common platform for the downstream analysis of crosslinking mass spectrometry data. bioRxiv. https://doi.org/10.1101/561829.

Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in C-13/N-15-enriched proteins by isotropic mixing of C-13 magnetization. J. Magn. Reson. Ser. B *101*, 114–119.

Grzesiek, S., and Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. J. Biomol. NMR *3*, 185–204.

Hannon, C., Cruz-Migoni, A., Platonova, O., Owen, R.L., Nettleship, J.E., Miller, A., Carr, S.B., Harris, G., Rabbitts, T.H., and Phillips, S.E.V. (2018). Cloning, purification and structure determination of the HIV integrase-binding domain of lens epithelium-derived growth factor. Acta Crystallogr. F Struct. Biol. Commun. *74*, 143–149.

Herrmann, T., Guntert, P., and Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. *319*, 209–227.

Huang, J., Gurung, B., Wan, B., Matkar, S., Veniaminova, N.A., Wan, K., Merchant, J.L., Hua, X., and Lei, M. (2012). The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. Nature *482*, 542–546.

Hughes, S., Jenkins, V., Dar, M.J., Engelman, A., and Cherepanov, P. (2010). Transcriptional co-activator LEDGF interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity. J. Biol. Chem. *285*, 541–554.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kukic, P., O'Meara, F., Hewage, C., and Nielsen, J.E. (2013). Coupled effect of salt and pH on proteins probed with NMR spectroscopy. Chem. Phys. Lett. *579*, 114–121.

LeRoy, G., Oksuz, O., Descostes, N., Aoi, Y., Ganai, R.A., Kara, H.O., Yu, J.R., Lee, C.H., Stafford, J., Shilatifard, A., et al. (2019). LEDGF and HDGF2 relieve the nucleosome-induced barrier to transcription in differentiated cells. Sci. Adv. *5*, eaay3068.

Liang, B., Bushweller, J.H., and Tamm, L.K. (2006). Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. J Am Chem Soc *128*, 4389–4397.

Llano, M., Delgado, S., Vanegas, M., and Poeschla, E.M. (2004). Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. J. Biol. Chem. *279*, 55570–55577.

Maertens, G.N., Cherepanov, P., and Engelman, A. (2006). Transcriptional coactivator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. J. Cell Sci. *119*, 2563–2571.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem *278*, 33528–33539.

Michel, F., Crucifix, C., Granger, F., Eiler, S., Mouscadet, J.F., Korolev, S., Agapkina, J., Ziganshin, R., Gottikh, M., Nazabal, A., et al. (2009). Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. EMBO J. *28*, 980–991.

Millan, C., Sammito, M., and Uson, I. (2015). Macromolecular ab initio phasing enforcing secondary and tertiary structure. IUCrJ *2*, 95–105.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Renshaw, P.S., Veverka, V., Kelly, G., Frenkiel, T.A., Williamson, R.A., Gordon, S.V., Hewinson, R.G., and Carr, M.D. (2004). Sequence-specific assignment and secondary structure determination of the 195-residue complex formed by the *Mycobacterium tuberculosis* proteins CFP-10 and ESAT-6. J. Biomol. NMR *30*, 225–226.

Sammito, M., Millan, C., Frieske, D., Rodriguez-Freire, E., Borges, R.J., and Uson, I. (2015). ARCIMBOLDO_LITE: single-workstation implementation and use. Acta Crystallogr. D Biol. Crystallogr. *71*, 1921–1930.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. *78*, 1606–1619.

Sharma, S., Cermakova, K., De Rijck, J., Demeulemeester, J., Fabry, M., El Ashkar, S., Van Belle, S., Lepsik, M., Tesina, P., Duchoslav, V., et al. (2018). Affinity switching of the LEDGF/p75 IBD interactome is governed by kinase-dependent phosphorylation. Proc. Natl. Acad. Sci. U S A *115*, E7053–E7062.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Shinohara, T., Singh, D.P., and Fatma, N. (2002). LEDGF, a survival factor, activates stress-related genes. Prog. Retin. Eye Res. *21*, 341–358.

Sjodt, M., and Clubb, R.T. (2017). Nitroxide labeling of proteins and the determination of paramagnetic relaxation derived distance restraints for NMR studies. Bio Protoc. *7*, e2207.

Tesina, P., Cermakova, K., Horejsi, M., Prochazkova, K., Fabry, M., Sharma, S., Christ, F., Demeulemeester, J., Debyser, Z., De Rijck, J., et al. (2015). Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. Nat. Commun. *6*, 7968.

Turlure, F., Maertens, G., Rahman, S., Cherepanov, P., and Engelman, A. (2006). 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.

Vanderlinden, W., Lipfert, J., Demeulemeester, J., Debyser, Z., and De Feyter, S. (2014). Structure, mechanics, and binding mode heterogeneity of LEDGF/p75-DNA nucleoprotein complexes revealed by scanning force microscopy. Nanoscale *6*, 4611–4619.

Veverka, V., Crabbe, T., Bird, I., Lennie, G., Muskett, F.W., Taylor, R.J., and Carr, M.D. (2008). Structural characterization of the interaction of mTOR with phosphatidic acid and a novel class of inhibitor: compelling evidence for a central role of the FRB domain in small molecule-mediated regulation of mTOR. Oncogene *27*, 585–595.

Veverka, V., Lennie, G., Crabbe, T., Bird, I., Taylor, R.J., and Carr, M.D. (2006). NMR assignment of the mTOR domain responsible for rapamycin binding. J. Biomol. NMR *36* (*Suppl 1*), 3.

Wang, H., Farnung, L., Dienemann, C., and Cramer, P. (2020). Structure of H3K36-methylated nucleosome–PWWP complex reveals multivalent cross-gyre binding. Nat. Struct. Mol. Biol. *27*, 8–13.

Weickert, M.J., and Adhya, S. (1993). The galactose regulon of *Escherichia coli*. Mol. Microbiol. *10*, 245–251.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallographica. D Biol. Crystallogr. *67*, 235–242.

Wu, X., Daniels, T., Molinaro, C., Lilly, M.B., and Casiano, C.A. (2002). Caspase cleavage of the nuclear autoantigen LEDGF/p75 abrogates its pro-survival function: implications for autoimmunity in atopic disorders. Cell Death Differ. *9*, 915–925.

Yanofsky, C. (1971). Tryptophan biosynthesis in *Escherichia coli*. Genetic determination of the proteins involved. JAMA *218*, 1026–1035.

Yokoyama, A., and Cleary, M.L. (2008). Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell *14*, 36–46.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
DH5α (Single shot - Max efficiency)	Invitrogen by Thermo Fisher Scientific	Cat#44-0097
BL21(DE3)	NEB	Cat#C2527I
BL21(DE3)pLysS (Oneshot)	Invitrogen by Thermo Fisher Scientific	Cat#44-0054
Chemicals, Peptides, and Recombinant Proteins		
Luria Bertani Broth	Sigma	Cat#L3022
LB Agar (Lennox)	Sigma	Cat#L2897
Ampicillin	Sigma	Cat#A9518
Ethyl β-D-thiogalactopyranoside	Carbosynth	Cat#ME02027
IPTG	Sigma	Cat#I6758
Glycerol	Penta	Cat#14550-11000
Deoxyribonuclease I (bovine pankreas)	Sigma	Cat#DN-25
DNase	Roche	Cat#4716728001
cOmplete^{TM} EDTA-free protease inhibitor	Roche	Cat#04693132001
Trizma base	Sigma	Cat#T6066
NaCl	Lachner	Cat#30093-AP0-G1000-1
NaCl	Sigma	Cat# S3014
EDTA	Sigma	Cat#E9884
β-mercaptoethanol	Sigma	Cat#M6250
TCEP	Sigma	Cat#C4706
DTT	GE Healthcare	Cat# GE17-1318-02
HEPES	Sigma	Cat#H3375
Imidazole	Sigma	Cat#56749
Tween 20	PanReac Applichem	Cat#A4974.0100
BSA	Sigma	Cat#A2153
MTSSL (CAS 81213-52-7)	Abcam	Cat#144308
1-Acetyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl Methanethiosulfonate (CAS 244641-23-4)	Toronto Research Chemicals	Cat#A188600
L-ascorbic acid	Sigma-Aldrich	Cat#765437
^{15}N ammonium sulphate	Cambridge isotope laboratories	Cat#NLM713-10
^{13}C glucose	Cambridge isotope laboratories	Cat#CLM1396-10
Deuterium oxide	Eurisotop	Cat#D214F
Simply Blue Safe Stain	Invitrogen	Cat#LC6060
Select Nickel Hisaffinity	Thermo fisher	Cat # 30210
Ni-NTA	Sigma	Cat#P6611-100mL
Superdex 75 10/300 GL	GE Healthcare	Cat#17-5174-01
Glutathione Sepharose-4 Fast Flow	GE Healthcare	Cat#17-5132-01
Amylose binding material	NEB	Cat#E8021S
HiTrap heparin HP column	GE Healthcare	Cat#GE17-0407-01
AlphaScreen Glutathione donor beads	Perkin Elmer	Cat#6765301
AlphaScreen Ni^{2+} chelate acceptor beads	Perkin Elmer	Custom service
Alphascreen anti-flag acceptor beads	Perkin Elmer	Custom service# CUSM81968000EA
Amicon Ultra-0.5 Centrifugal Filter unit 10kDa cutoff	Millipore	Cat#UCF501096

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Amicon Ultra-4 Centrifugal Filter unit 10kDa cutoff	Millipore	Cat#UCF801096
Amicon Ultra-15 Centrifugal Filter unit 10kDa cutoff	Millipore	Cat#UCF901024
Millex-GS Syringe Filter Unit of 0.22 μm	Millipore	Cat#SLGSV255F
PD Minitrap G25	GE Healthcare	Cat#289180
384-well Optiwell microtiter plates white	Perkin Elmer	Cat#6007290
TEV protease	Lab source	N/A
Flag-LEDGF/p75	(Maertens et al., 2003)	N/A
GST-LEDGF/p75	(Cherepanov et al., 2004)	Gift from Mamuka Kvaratskhelia, University of Colorado, Denver, USA
$JPO2_{2-130}$	(Tesina et al.,2015)	N/A
$POGZ_{1117-1410}$	(Tesina et al., 2015)	N/A
$LEDGF_{345-431}$	This study	N/A
$LEDGF_{345-441}$	This study	N/A
$LEDGF_{345-453}$	This study	N/A
$LEDGF_{345-461}$	This study	N/A
$LEDGF_{345-467}$	This study	N/A
$LEDGF_{345-467\ \Delta443-453}$	This study	N/A
$LEDGF_{345-467\ E451A\ E452A}$	This study	N/A
$LEDGF_{345-467\ E451R\ E452R}$	This study	N/A
$LEDGF_{345-467\ C373A}$	This study	N/A
$LEDGF_{345-467\ C373A\ S409C}$	This study	N/A
$LEDGF_{345-467\ C373A\ S443C}$	This study	N/A
$LEDGF_{345-467\ C373A\ T456C}$	This study	N/A
$LEDGF_{345-467\ C373A\ N465C}$	This study	N/A
Critical Commercial Assays		
Morpheus complete	Molecular Dimensions	Cat#MD1-123
Protein labelling kit Red-maleimid	Nanotemper	Cat#L001
BCA kit	ThermoScientific Pierce BCA Protein Assay Kit	Cat#23225
Deposited Data		
$LEDGF_{345-467}$ dimer	This study	PDB: 6TVM
$LEDGF_{345-467}$ dimer	This study	BMRB: 34475
LEDGF/p75 IBD	This study	PDB: 6TRJ
LEDGF-MLL1-Menin	(Huang et al., 2012)	PDB: 3U88
Oligonucleotides		
Listed in Data S1		
Recombinant DNA		
pMCSG7	MCSG	http://bioinformatics.anl.gov/mcsg/technologies/vectors.html
$LEDGF_{345-431}$	This study	N/A
$LEDGF_{345-441}$	This study	N/A
$LEDGF_{345-453}$	This study	N/A
$LEDGF_{345-461}$	This study	N/A
$LEDGF_{345-467}$	This study	N/A
$LEDGF_{345-467\ \Delta443-453}$	This study	N/A
$LEDGF_{345-467\ E451A\ E452A}$	This study	N/A
$LEDGF_{345-467\ E451R\ E452R}$	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
$LEDGF_{345-467\ C373A}$	This study	N/A
$LEDGF_{345-467\ C373A\ S409C}$	This study	N/A
$LEDGF_{345-467\ C373A\ S443C}$	This study	N/A
$LEDGF_{345-467\ C373A\ T456C}$	This study	N/A
$LEDGF_{345-467\ C373A\ N465C}$	This study	N/A
Flag-LEDGF/p75	(Maertens et al., 2003)	N/A
GST-LEDGF/p75	(Cherepanov et al., 2004)	Gift from Mamuka Kvaratskhelia, University of Colorado, Denver, USA
$JPO2_{2-130}$	(Tesina et al.,2015)	N/A
$POGZ_{1117-1410}$	(Tesina et al., 2015)	N/A
Software and Algorithms		
ARCIMBOLDO_LITE	(Millan et al., 2015, Sammito et al., 2015	N/A
Coot	(Emsley and Cowtan, 2004)	http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/
Refmac5 program	(Murshudov et al., 1997)	http://www.ccp4.ac.uk
CCP4 suite	(Winn et al., 2011)	http://www.ccp4.ac.uk
MolProbity	(Chen et al., 2010).	N/A
XDS suite	(Kabsch, 2010)	N/A
TopSpin™	Bruker	N/A
Analysis 4.4 (Bruker Daltonics, Billerica, MA, USA)	(Bruker Daltonics, Billerica, MA, USA)	N/A
LinX software		(http://peterslab.org/downloads/SW/LinX/LinX.zip)
PyMOL	DeLano Scientific LLC	http://www.pymol.org/
GraphPad Prism	GraphPad Software Inc.	http://www.graphpad.com/scientific-software/prism/
Sednterp	Philo, J., Hayes, D. B. & Laue, T.	bitc.unh.edu
Sedfit	(Schuck, 2000)	http://www.analyticalultracentrifugation.com/sedfit.htm
Gussi	(Brautigam, 2015)	http://biophysics.swmed.edu/MBR/software.html
Cyana	Herrmann et al., 2002	N/A
Talos+	Shen et al., 2009	https://spin.niddk.nih.gov/bax/nmrserver/talos/
XiView	Graham et al., 2019	https://xiview.org/xiNET_website/index.php
Protein structure validation software		www.nesg.org

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for reagents and resources may be directed to, and will be fulfilled by the Lead Contact Václav Veverka (vaclav.veverka@uochb.cas.cz)

Material Availability

All the constructs generated in this study are available upon request.

Data and Code Availability

Atomic coordinates and structure factors for the crystal structure of $LEDGF_{345-431}$ were deposited in the Protein Data Bank (PDB) with the accession code 6TRJ. The structures, NMR restraints and resonance assignments obtained for $LEDGF_{345-467}$ were deposited in the PDB accession code: 6TVM and BioMagResBank with accession code: 34475.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Recombinant proteins were expressed using E.coli strain BL21(DE3) (NEB) and BL21(DE3)pLysS (Oneshot) (Invitrogen). E.coli were cultivated either in LB media or minimal media supplemented with $^{15}N(NH_4)_2SO_4$ and ^{13}Cglucose as the nitrogen or carbon source. Further specific information are specified in Method Details

METHOD DETAILS

Cloning

The expression plasmids coding for all LEDGF variants, $JPO2_{2\text{-}130}$ and $POGZ_{1117\text{-}1410}$ were derived from the pMCSG7 vector (T7 driven, ampicillin resistance). The protein coding sequence is preceded by an N-terminal His_6 affinity tag and a tobacco etch virus (TEV) protease recognition site. As a result of TEV protease cleavage, five additional residues (SNAAS) were retained at the N-terminus of the natural protein sequence. Point mutations and deletions were introduced by site-directed mutagenesis and checked by sequencing.

Protein Expression and Purification

All proteins used in AlphaScreen were overexpressed in *E. coli* BL21 (DE3)pLysS cells using Lysogeny broth (LB) medium supplemented with ampicillin. Bacterial cultures were grown at 37 °C, except for His-tagged proteins grown at 30°C. Protein expression was induced with 0.5-1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6. Cultures were harvested after overnight expression at 18°C, but GST- and MBP-tagged proteins were expressed for 4h at 37°C. Bacterial pellets were stored at −20°C. Cells were resuspended in a lysis buffer (50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM dithiothreitol (DTT), 0.1 μg/mL DNAse (Thermo Scientific) and protease inhibitor (cOmplete, EDTA-free, Roche) and sonicated (MSE 150 Watt Ultrasonic Disintegrator). The supernatant obtained by centrifugation of the lysate was first filtered on a Millex-GS Syringe Filter Unit of 0.22 μm (Millipore). Flag-tagged proteins were purified on a 5 mL HiTrap heparin and Superdex 75 10/300 GL column (GE Healthcare Life Sciences). GST-LEDGF/p75 was purified by affinity chromatography on Glutathione Sepharose-4 Fast Flow (GE Healthcare Life Sciences). All His-tagged proteins were affinity purified on Ni-NTA-resin (Thermo Fisher Scientific, Waltham, MA, USA) and additionally, over a HiTrap Heparin HP resin (GE Healthcare Life Sciences). MBP-tagged proteins were applied to amylose resin (NEB, Ipswich, MA, USA). All proteins were stored in Tris–saline buffer supplemented with 10% (v/v) glycerol at –80°C.

For all other experiments (NMR, SEC, chemical cross-linking) proteins were prepared using the following procedure. Expression plasmids were transformed into competent BL21(DE3) *E.coli* cells. The transformed colonies were grown in LB or minimal medium supplemented either with $(^{15}NH_4)_2SO_4$ or also ^{13}C glucose for double labeled proteins at 37°C to an OD_{600} of 0.8. Cultures were then transferred to 18°C and protein production was induced by 0.25 mM of ETG. After overnight protein expression for 15h at 18°C, bacterial cultures were pelleted by centrifugation (5000×g, 4°C, 20 minutes) and stored at -20°C. Each gram of bacterial pellet was lysed using 10 mL of lysis buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl, 2 mM β-mercaptoethanol, 10 μM EDTA) supplemented with 1× protease inhibitor cocktail tablet-EDTA free (Roche), DNase I (DN25 Sigma-Aldrich,2 units/mL) and sonicated (Branson Sonifier 250) until a homogeneous lysate was obtained. The lysates were centrifuged at 30,000×g for 30 minutes and the supernatant was loaded on Ni-chelate agarose resin. Bound fractions were eluted with an elution buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol, 10 μM EDTA). Fractions containing the protein of interest were pooled, TEV protease was added and cleavage of His_6-tag was performed along with dialysis to 50-100x excess of lysis buffer overnight at 4°C. The cleaved His_6-tag was removed on Ni-chelate agarose and the protein of interest retained in the flow-through fraction, which was then concentrated with an Amicon Ultra centrifugal filtration unit. The final purification by size exclusion chromatography was performed on Superdex 75 or 200 10/300 GL (GE Healthcare Life Sciences) in 20 mM HEPES pH 7.0, 100 mM NaCl, 1 mM TCEP. The same setup was also used for analytical size-exclusion experiments with Superdex 75 10/300 GL (GE Healthcare Life Sciences), in which the column was calibrated with protein standards of known molecular weight (Low molecular weight calibration kit 28403841, GE Healthcare Life Sciences). The protein concentration of the purified proteins was measured with a NanoDrop spectrophotometer and confirmed by amino acid analysis. The purity was determined by SDS-PAGE followed by Coomassie Brilliant Blue and silver staining. The proteins were shock-frozen in liquid nitrogen and stored at -80°C.

AlphaScreen

AlphaScreen is a bead-based, non-radioactive Amplified Luminescent Proximity Homogeneous Assay Screen (AlphaScreen, Perkin Elmer) to study protein-protein interactions. All reactions were performed in a final volume of 25 μL in white 384-well Optiwell microtiter plates (PerkinElmer). Components were diluted in assay buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM DTT, 0.1% (w/v) BSA and 0.1% (v/v) Tween 20). Protein concentrations and dilutions are always indicated in the figures; concentrations were determined in cross-titration experiments. Firstly, all proteins were applied and incubated on the plate for 1 hour at 4°C. Subsequently, the corresponding acceptor and donor beads (PerkinElmer) were added, bringing all proteins to the indicated final concentrations. After 1 hour of incubation at 28°C, the plate was analyzed in an EnVision Multi-label Reader in AlphaScreen mode (PerkinElmer). Data were evaluated using GraphPad Prism 7.0. After background subtraction data were fitted with a nonlinear regression - one site specific binding ($Y = Bmax * x(K_D + x)$; Bmax is the maximum specific binding).

Microscale Thermophoresis

LEDGF/p75 was labeled using the NanoTemper Protein Labeling Kit RED- MALEIMIDE (Cat. No. L001), according to the manufacturer's protocol. After labeling, the protein concentration was adjusted to 20 nM as it gave a reasonable fluorescence intensity. Unlabeled LEDGF/p75 was diluted in buffer containing 25 mM Tris pH 7.4, 150 mM NaCl, 1 mM DTT, 0.05% Tween 20 (v/v) and 0.1% BSA (w/v) to a final concentration of 35 μM. Both the labeled and unlabeled LEDGF/p75 were first centrifuged for at least 30 minutes before titrating and mixing both proteins. Both proteins were mixed in a 1:1 ratio and incubated for 15 min on ice before loading the samples using premium capillaries into the Monolith instrument. Measurements of LEDGF/p75 dimerization were performed on a NanoTemper Monolith NT.115 instrument run at 30°C. An LED Power of 60% and a Laser Power of 5% were used.

Analytical Ultracentrifugation

The oligomeric state of the $LEDGF_{345-467}$ at the 0.5-25 μM concentration range was analyzed in aBeckman Optima analytical ultracentrifuge (AUC). The sedimentation-velocity experiment was performed at 55,000 rev.min^{-1} using an An-60 Ti rotor for 21 h recording interference scans every 5 min at 20°C in 20 mM HEPES pH 7.0, 100 mM NaCl and 1 mM TCEP. The density and viscosity of the buffer (1.00391 g.cm^{-3}, 1.0264 mPa.s) was calculated using Sednterp (http://bitcwiki.sr.unh.edu) and the partial specific volume of the protein was 0.73. Results were processed using SEDFIT (Schuck, 2000), fitting to the c(s) model at confidence level ratio 0.95 and the graph was prepared using GUSSI (Brautigam, 2015) and GraphPad Prism.

Chemical Cross-Linking and Mass Spectrometry

$LEDGF_{345-467}$ ^{15}N enriched and with natural abundance of ^{15}N was mixed 1:1 molar ratio and applied on Superdex 75 10/300 GL column (GE Healthcare Life Sciences). Peak fractions were cross-linked with 20 molar excess of DSG (disuccinimidyl glutarate) and digested by trypsin. One μg of peptide mixture was injected onto an analytical reversed phase column (Zorbax 300SB-C18 3.5 μm, 0.3 × 150 mm from Agilent Technologies, Santa Clara, CA, USA) with a trap pre-column (Zorbax 300SB-C18 5 μm, 0.3 × 5 mm) and separated using a 5%–40% acetonitrile gradient over 35 min. Throughout chromatography, 0.1% formic acid was used as an ion-pairing agent. The HPLC system was directly coupled with the ESI source of an FT-ICR mass spectrometer solariX XR equipped with 15T superconducting magnet Bruker Daltonics (Billerica, MA, USA). The mass spectrometer was operated in positive broadband mode over a 250–2500 m/z range with 1M data point acquisition. Data processing was performed using Data Analysis 4.4 (Bruker Daltonics, Billerica, MA, USA), and cross-links were assigned within 1 ppm mass accuracy using in-house-developed LinX software (http://peterslab.org/downloads/SW/LinX/LinX.zip).

NMR Spectroscopy

NMR spectra were acquired at 25 °C on an 850 MHz Bruker Avance spectrometer, which is equipped with a triple-resonance ($^{15}N/^{13}C/^{1}H$) cryoprobe. The sample volume was either 0.16 or 0.35 mL, in buffer (20 mM HEPES pH 7.0, 100 mM NaCl, 1 mM TCEP), 5 -10% D_2O/90-95% H_2O. A series of double- and triple-resonance spectra (Renshaw et al., 2004; Veverka et al., 2006) were recorded to obtain sequence-specific backbone resonance assignment for the studied LEDGF/p75 constructs. For full structural determination of $LEDGF_{345-467}$ we extended resonance assignment to aliphatic side-chains using a combination of standard 3D CC(CO)NH (Grzesiek et al., 1993), HBHA(CO)NH (Grzesiek and Bax, 1993) and (H)CCH-TOCSY experiments (Bax et al., 1990). Aromatic side-chain resonance assignments and ^{1}H-^{1}H distance restraints were derived from 3D $^{15}N/^{1}H$ NOESY-HSQC and $^{13}C/^{1}H$ NOESY-HMQC, which were acquired using a NOE mixing time of 100 ms.

The structural calculation was carried out in Cyana (Herrmann et al., 2002) using NOESY data in combination with backbone torsion angle restraints, generated from assigned chemical shifts using the program TALOS+ (Shen et al., 2009) and paramagnetic relaxation enhancement derived distance restraints (described below). First, the combined automated NOE assignment and structure determination protocol (CANDID) was used for automatic NOE cross-peak assignment. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle restraints were used to calculate a set of converged structures with no significant restraint violations (distance and van der Waals violations <0.5Å and dihedral angle constraint violations <5°). The 30 structures with the least restraint violations were selected for further analysis using the Protein Structure Validation Software suite (www.nesg.org). The statistics for the resulting structures are summarized in Table S1. The structures, NMR restraints and resonance assignments were deposited in the Protein Data Bank (PDB, accession code: 6TVM) and BMRB (accession code: 34475).

To follow changes in the chemical shifts of a protein upon partner protein binding or change in conditions (protein or salt concentration), we calculated chemical shift perturbations (CSPs). The CSP of each assigned resonance in the 2D $^{15}N/^{1}H$ HSQC spectra of the protein in the free state is calculated as the geometrical distance in ppm to the peak in the 2D $^{15}N/^{1}H$ HSQC spectra acquired under different conditions using the formula: $\Delta\delta = \sqrt{\Delta\delta_H^2 + (\Delta\delta_N \cdot \alpha)^2}$, where α is a weighing factor of 0.2 used to account for differences in the proton and nitrogen spectral widths (Veverka et al., 2008).

The relaxation data were obtained using 300 μM $LEDGF_{345-467}$ ^{15}N labeled in 20 mM HEPES pH 7.0, 100 mM NaCl, 1 mM TCEP. The relaxation delays (in seconds) were 0.0448, 0.0672, 0.112, 0.179, 0.246, 0.381, 0.784, 1.23 for R_1 and 0, 0.0310, 0.0628, 0.0931, 0.124, 0.186, 0.279 for R_2 measurements. The recovery delay was set to 1.2 s with 4 or 8 scans per FID. The {^{1}H}–^{15}N steady state nuclear Overhauser enhancement (NOE) was measured in an interleaved manner using a 5 s long train of selective 180° pulses for

proton irradiation, separated by 22 ms delays (Ferrage et al., 2010), and centered at 8.2 ppm in order to avoid water saturation. The offset of these pulses was shifted by 50 kHz in the reference experiment to ensure constant sample heating (Ferrage, 2012; Ferrage et al., 2010).

Para and Diamagnetic Spin-Labelling

Variants of $LEDGF_{345-467}$ were modified as follows: the endogenous cysteine residue 373 was mutated to alanine and cysteine residues were introduced additionally at positions 409 (S409C), 443 (S443C), 456 (T456C) and 465 (N465C). All proteins were expressed either in LB media or in ^{15}N enriched minimal media and purified as described above with the exception that the size exclusion chromatography was performed in buffer with 5 mM TCEP to prevent disulfide bond formation of introduced cysteine residues. The paramagnetic MTSSL probe (CAS 81213-52-7; 1-λ1-Oxidanyl-2,2,5,5-tetramethyl-3-(methylsulfonylsulfanylmethyl) pyrrole) was used for spin labeling and the diamagnetic variant (CAS 244641-23-4, 1-Acetyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl Methanethiosulfonate) was used as a control as it should have no effect on signal intensities. Probes were dissolved at 100 mM concentration in DMSO.

Purified $LEDGF_{345-467}$ variants at 300-500 μM concentration were incubated on ice for at least 1 h with 30 mM TCEP to reduce all disulfide bonds. Protein (0.5 mL) was then applied to a PD-10 desalting column, equilibrated with 20 mM HEPES pH 7.0, 100 mM NaCl, and it was eluted with 1 mL of 20 mM HEPES pH 7.0, 100 mM NaCl directly into a tube with 0.5 mL of 3 mM paramagnetic or diamagnetic probe in 20 mM HEPES pH 7.0, 100 mM NaCl (1.5 mL of total volume after elution from PD-10 column and 1 mM para- or diamagnetic probe). After mixing, additional para- or diamagnetic probe solution was added to a final probe concentration of 2 mM the mixture was incubated with the probe overnight at 4°C, in the dark. In general, a 15-20 molar excess of probe was used compared to protein. Labeled protein was concentrated in an Amicon Ultra centrifugal filtration unit and unbound probe and DMSO were removed by several rounds of washing with 20 mM HEPES pH 7.0, 100 mM NaCl buffer in an Amicon Ultra centrifugal filtration unit. Labeling efficiency was evaluated by MALDI-TOF. The molecular mass of the protein modified by the paramagnetic probe or diamagnetic probe was increased by 184 or 211 Da, respectively.

Note: The commonly used approach, where the paramagnetic variant of the protein is prepared and measured first and the diamagnetic variant is subsequently prepared by reducing the sample with ascorbic acid was tested, but the paramagnetic probe was insufficiently reduced by ascorbic acid, especially for the variant with labelled residue 409. It seemed that this residue is not accessible for ascorbic acid even at a high excess (up to 20- fold) and extended incubation period. A similar observation was reported previously (Liang et al., 2006). In addition, extended NMR data acquisition (repetition delay extended to 4s for a quantitative analysis) at room temperature resulted in a gradual precipitation of the measured protein, which resulted in reduced overall signal intensity for the diamagnetic variant. Therefore, we considered it more quantitative to split the protein sample into two halves and label one part with the paramagnetic and the other with the diamagnetic probe. During this procedure, both para- and diamagnetic samples were handled identically, including a similar data acquisition time. The labelling efficiency was over 80% as confirmed by mass spectrometry.

Paramagnetic Relaxation Enhancement (PRE) Measurements

The protein sample (160 μL) was substituted with 10% D_2O and 2D $^{15}N/^{1}H$ HSQC were measured in a 3 mm NMR tube with a repetition delay of 4 s. To detect inter- and intramolecular interactions, ^{15}N enriched LEDGF C373A S409C labeled with para- or diamagnetic probe for (PRE) were measured directly using NMR. LEDGF cysteine variants C373A in combination with S443C, T456C or N465C were labeled only with paramagnetic probe and 2D $^{15}N/^{1}H$ HSQC were measured. Subsequently, sample was reduced into diamagnetic version with 10 fold molar excess of ascorbic acid (freshly prepared 500 mM ascorbic acid with pH adjusted to 5.6, so that upon dilution it has pH 7.0) and 2D $^{15}N/^{1}H$ HSQC measurement with same settings was repeated.

To distinguish solely intermolecular interactions, LEDGF cysteine variants (C373A in combination with S409C, S443C, T456C and N465C) expressed in LB media (naturally abundant nitrogen isotopes) were PRE labeled (para or diamagnetic probe, PRE source) and they were mixed in 1:1 molar ratio with ^{15}N labeled $LEDGF_{345-467}$ C373A (PRE affected and NMR visible). Only signals from the ^{15}N labeled $LEDGF_{345-467}$ C373A were detected in the 2D $^{15}N/^{1}H$ HSQC spectra. Signals with decreased intensity indicate residues of $LEDGF_{345-467}$ C373A in proximity to the paramagnetic probe located on the LEDGF cysteine variant (C373A in combination with S409C, S443C, T456C and N465C) modified with the PRE probe (Sjodt and Clubb, 2017) (Figure S2). In case of a fast exchange of protomers between the oligomers, the distribution of dimer would be 1 (homodimer of 15N protein): 2 (mixed dimer = one monomer 15N and one monomer 14N) :1 (homodimer of 14N protein). Homodimers of 14N protein cannot be detected in the NMR under the used experimental conditions and the 15N homodimer should by default be no effect of a paramagnetic probe. Therefore, in the best-case scenario there would be a 14N/15N mixed dimer present in 50% of all dimers. As a prerequisite, the dimers already formed within each variant must dissociate upon mixing to establish a new equilibrium. The relatively slower homodimer dissociation rate prevented the enrichment of a mixed dimer population which leads to lower PRE effect than expected. To achieve higher populations of mixed dimers we would have to incubate the sample over an extended period of time which was not feasible due to the stability issues of mutated variants.

We introduced PRE-derived 15Å distance restraints into structural calculation between the Cα atoms from a residue modified by the paramagnetic probe (S443 and T456) and residues significantly affected by PRE in the uniformly ^{15}N labeled dimers ($I_{para}/I_{dia} < 0.5$; Figure 2D). The PREs from the $^{15}N/^{14}N$ mixed samples were not used due to the ineffective protomer exchange between the unmodified ^{15}N labeled and modified ^{14}N labeled dimers. The spin label was not present in structural calculations that were carried

out using the wild-type sequence. Back-calculation of I_{para}/I_{dia} ratios was performed using PyMOL (calculation of distances) and in-house R script (evaluation of formula (1) and (2) in (Sjodt and Clubb, 2017)). A correlation time of 9 ns (estimated from R_2/R_1 ratios obtained for rigid residues in the IBD) and amide proton R_2 equal to 24 s^{-1} was used.

Note: There are some residual intensities as a consequence of the differential S/N ratios due to the overlaps of the affected systems. Therefore, we did not use peaks from the heavily overlapped regions for further analysis.

X-ray Crystallography

Crystals of LEDGF/p75 IBD were prepared using sitting drop vapor diffusion with a protein-precipitant ratio of 1:1. The protein solution (25 mg/mL in 25 mM Tris pH 7.5, 200 mM NaCl, 2 mM β-mercaptoethanol) was screened against Morpheus® screen (Molecular Dimensions). Several conditions yielded harvestable crystals. The most promising crystals formed within 3 days in condition H11 (0.1 M DL-Glutamic acid monohydrate; 0.1 M DL-Alanine; 0.1 M Glycine; 0.1 M DL-Lysine monohydrochloride; 0.1 M DL-Serine; 0.1 M Tris (base); BICINE pH 8.5; 20% v/v Glycerol; 10% w/v PEG 4000). Crystals were flash frozen in liquid nitrogen without any additional cryoprotectant. X-ray diffraction data at 100 K were collected at an in-house diffractometer (MicroMax-007 HF microfocus equipped with PILATUS 300K detector, Rigaku). Diffraction data were processed using the XDS suite of programs (Kabsch, 2010). Crystal parameters and data collection statistics are summarized in Table S1.

The crystal structure was solved using ARCIMBOLDO_LITE (Millan et al., 2015; Sammito et al., 2015) implemented in the CCP4 suite (Winn et al., 2011). The refinement was carried out with theRefmac5 program (Murshudov et al., 1997), with 5% of the reflections reserved for cross-validation. The Coot program was used for model rebuilding and addition of water molecules (Emsley and Cowtan, 2004). The quality of the crystallographic model was assessed with MolProbity (Chen et al., 2010). The final refinement statistic is summarized in Table S1. Atomic coordinates and structure factors for the crystal structure of $LEDGF_{345-431}$ were deposited in the PDB with the accession code 6TRJ.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistics for the resulting NMR structures are summarized in Table S1 and statistics of X-ray diffraction data collection and structure refinement are summarized in Table S2.