**Charles University in Prague**

**Faculty of Science**

**Department of Genetics and Microbiology**

**Mgr. Petr Těšina**

# Structural Studies of LEDGF/p75 Interactions

**PhD Thesis**

**Supervisor: Pavlína Maloy Řezáčová, Ph.D.**

**Department of Structural Biology**

**Institute of Organic Chemistry and Biochemistry, AS CR, v.v.i.**

**Prague, 2015**

## Statement of authorship:

I hereby declare, that I have written this thesis independently under the supervision of Pavlína Maloy Řezáčová, Ph.D. I have cited all the used information sources. This thesis serves only and exclusively for my PhD graduation at Faculty of Science of Charles University in Prague. Any part of this thesis was not used in support of an application for any other academical degree.

Prague ……. 2015 Mgr. Petr Těšina

# ACKNOWLEDGEMENTS

I would like to thank all my colleagues from the Department of Structural Biology. Many thanks to my supervisor Pavlína Řezáčová for all her help and support. Thank you for the opportunity to work on this challenging and intriguing project. Special thanks to Pavel Mader and Irena Sieglová, who introduced me to the biochemical methods and taught me the principles of good laboratory practice. None of my work would be possible without Milan Fábry, the cloning magician and spirit of the team. Thank you, Aleš Hnízda for all the fun, support and scientific debates. Thank you Petr Pachl, Jirka Brynda, Vlasta Král, Juraj Sedláček, Káťa Procházková, Magda Hořejší, Věra Mrkvičková and Jitka Kredbová. You guys created something extraordinary and precious and I was extremely lucky to be a part of it. Last but not least, thank you Václav Veverka, for bringing your NMR expertise into this project.

Many thanks to our colleagues from Leuven, Belgium. Thank you Káťa Čermáková and Jonas Demeulemeester for taking care of me in Leuven and for sharing all the fun and ideas. Thanks to Jan De Rijck and Zeger Debyser for the fruitful collaboration and inviting me over.

I would also like to thank some of my former colleagues, who trusted me long before I chose to strive for a PhD and set me on this path. Thank you, Braňo Fleischer for giving me the chance. Thank you, David Staněk for showing me what it means to be a great boss and scientist.

Very special thanks go to my family for not giving up and to Péťa Malá for all her love, patience and support.

This work was supported by the Grant Agency of the Charles University in Prague grant 200213; FP7 framework of the European Union (THINC, HEALTH-2007-2.3.2-1); the Ministry of Education of the Czech Republic LK11205 (programme 'NAVRAT') and 7E08066; the Academy of Sciences of the Czech Republic, projects RVO 61388963 and 68378050.

# ABSTRACT

LEDGF/p75 protein is a human transcriptional co-activator and epigenetic reader associated with transcriptionally active chromatin. It is crucial for HIV integration and MLL1 fusion-driven leukemia development. Interactions of LEDGF/p75 with HIV integrase (HIV IN) and MLL1–menin complex are considered an attractive therapeutic target for drug development. LEDGF/p75 interacts with both HIV IN and MLL1–menin complex through its integrase binding domain (IBD).

While the pathophysiological interactions of LEDGF/p75 IBD were intensively studied, little was known about the physiological ones. In addition to HIV IN and MLL1, the LEDGF/p75 IBD also interacts with JPO2, PogZ, ASK and MLL2. In search for specific inhibitors of LEDGF/p75 IBD interaction with HIV IN and MLL1, it is essential to obtain detailed information about its interactions with all binding partners.

The IBD–MLL1–menin complex has been structurally characterized, but only partially. Using NMR spectroscopy, we identified and mapped a novel part of the IBD–MLL1 interface. This additional interface is able to maintain the interaction between LEDGF/p75 and MLL1 even without the presence of menin, which was considered necessary. Moreover, colony forming assays of primary leukemic blasts revealed that this additional interface is essential for leukemic transformation. Interestingly, the newly defined interface on IBD overlaps with the binding site of the HIV IN.

Our analyses revealed structural details of LEDGF/p75 interactions with other physiological binding partners. We found that interactions with the LEDGF/p75 IBD are maintained by an intrinsically disordered IBD-binding motif (IBM) common to all known cellular partners. This interaction interface and its importance have been thoroughly validated by mutation analyses and two solution structures of IBD–IBM complexes have been solved. Utilizing the structural information, we explained how HIV IN out-competes the cellular proteins.

Based on the knowledge of the IBM, we identified and validated IWS1 as a novel LEDGF/p75 interaction partner. IWS1 is a human transcription factor which plays a key role in defining the composition of the RNA polymerase II elongation complex. It also interacts with the H3K36 (histone 3, lysine 36) methyltransferase Hypb/Set2. Trimethylated H3K36 represents a signature chromatin mark of active transcription which is recognized by PWWP domain of LEDGF/p75.

Detailed characterization of physiological interaction interfaces on the IBD revealed a notable overlap with the region involved in interaction with HIV IN. Indeed, a large part of interaction interface on IBD is common for all the interaction partners and their interactions with LEDGF/p75 are mutually exclusive. The similar binding modes of LEDGF/p75 interaction partners represent a new challenge for the development of selective interaction inhibitors. However, in the case of MLL1, there is a unique additional menin-dependent interface engaged in its interaction with IBD. Such feature could allow for a specific inhibition.

# ABSTRAKT

Protein LEDGF/p75 je lidský transkripční koaktivátor, který je asociován s epigenetickými modifikacemi transkripčně aktivního chromatinu. Je také nezbytný v procesu integrace HIV do genomu hostitelské buňky a v leukemogenezi způsobené fúzními produkty genu kódujícího protein MLL1. Interakce proteinu LEDGF/p75 s HIV integrázou a komplexem MLL1–menin jsou považovány za potenciální cíl terapeutického zásahu. Interakce proteinu LEDGF/p75 s HIV integrázou i MLL1 jsou zprostředkovány tzv. integrázu vážící doménou (IBD – integrase binding domain).

Zatímco patofyziologické interakce IBD LEDGF/p75 byly intenzivně studovány, těm fyziologickým byla věnována malá pozornost. Mimo HIV integrázy a MLL1 interaguje LEDGF/p75 skrze IBD také s proteiny JPO2, PogZ, ASK a MLL2. Jelikož se interakce IBD s HIV integrázou stala cílem pro vývoj inhibitorů, bylo nutné získat informace o ostatních interakcích této domény. Jen díky detailní charakterizaci všech interakcí IBD je možné zajistit specifické cílení vyvíjených inhibitorů.

Proteinový komplex IBD–MLL1–menin byl již dříve strukturně charakterizován, ale pouze částečně. Pomocí NMR spektroskopie se nám podařilo zmapovat novou interakční oblast komplexu IBD–MLL1. Toto navazující interakční rozhraní je schopné zajistit interakci IBD s MLL1 i bez přítomnosti meninu, která byla považována za nezbytnou. Testování primárních leukemických buněk na schopnost tvořit kolonie odhalilo nezbytnost tohoto interakčního rozhraní v leukemické transformaci. Toto rozhraní na straně IBD se navíc překrývá s vazebnou oblastí HIV integrázy.

Naše analýzy odhalily strukturní detaily interakcí LEDGF/p75 s dalšími fyziologickými vazebnými partnery. Interakce s IBD jsou zprostředkovány vnitřně neuspořádaným vazebným motivem (IBM – IBD-binding motif), který je přítomen ve všech fyziologických vazebných partnerech. Důležitost tohoto motivu pro interakce s IBD byla ověřena pomocí mutačních analýz. Byly také vyřešeny dvě struktury komplexů IBD–IBM. Srovnáním dostupných strukturních informací jsme vysvětlili, proč má HIV integráza řádově vyšší afinitu vůči LEDGF/p75 než buněčné proteiny.

Díky znalosti IBM jsme zjistili, že protein IWS1 je dalším interakčním partnerem LEDGF/p75. IWS1 je lidský transkripční faktor, který hraje klíčovou roli ve složení elongačního komplexu RNA polymerázy II. IWS1 také interaguje s methyltransferázou Hypb/Set2, která methyluje lysin 36 histonu 3, čímž vytváří epigenetickou značku charakteristickou pro transkripčně aktivní chromatin. Tato epigenetická značka je pak rozpoznávána PWWP doménou proteinu LEDGF/p75.

Detailní charakterizace odhalila značný překryv interakčních rozhraní na IBD. Velká část těchto rozhraní je společná pro všechny interakční partnery IBD LEDGF/p75 včetně HIV integrázy a jednotlivé interakce se vzájemně vylučují. Tento překryv představuje komplikaci pro vývoj selektivních inhibitorů interakcí s IBD. Určitou výjimku v tomto směru představuje MLL1, který navíc využívá interakci s jinou částí IBD závislou na přítomnosti meninu, což by mohlo umožnit vývoj specifických inhibitorů.

## KEYWORDS

LEDGF/p75 / PogZ / JPO2 / PSIP1 / IWS1/ H3K36me3 / integrase / HIV / MLL / NMR / protein-protein interaction / leukemia

## KLÍČOVÁ SLOVA

LEDGF/p75 / PogZ / JPO2 / PSIP1 / IWS1/ H3K36me3 / integráza / HIV / MLL / NMR / protein-proteinová interakce / leukémie

# Table of contents

# ABBREVIATIONS

| | |
|---|---|
| aa | amino acid |
| ASK | Activator of S-phase Kinase |
| BME | β-mercaptoethanol |
| BSA | solvent-accessible surface area |
| CCD | Catalytic Core Domain |
| CD | Circular Dichroism |
| CDCA7L | Cell Division Cycle-Associated 7-Like protein |
| DSF | Differential Scanning Fluorimetry |
| DTT | Dithiothreitol |
| FPLC | Fast Protein Liquid Chromatography |
| GST | Glutathione S-transferase |
| H3K36me3 | tri-methylated lysine 36 of histone 3 |
| $His_6$ | Hexahistidine tag |
| HIV | Human Immunodeficiency Virus |
| HIV IN | Human Immunodeficiency Virus Integrase |
| HP1α | Heterochromatin Protein 1 alpha |
| HRP-2 | Hepatoma-derived growth factor Related Protein 2 |
| HSQC | Heteronuclear Single Quantum Correlation |
| IBD | Integrase Binding Domain of LEDGF/p75 |
| IBM | IBD-interaction motif |
| IDP | Intrinsically Disordered Protein |
| IPTG | isopropyl-β-D-thiogalactopyranoside |
| ITC | Isothermal Titration Calorimetry |
| IWS1 | Interacts With Spt6 1 |
| $K_d$ | Dissociation Constant |
| KIX | Kinase-Inducible domain interacting |
| LB | Lysogeny Broth |
| LC-MS | Liquid Chromatography – Mass Spectrometry |
| LEDGF | Lens Epithelium Derived Growth Factor |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionization – Time Of Flight |
| MBP | Maltose-Binding Protein |

| | |
|---|---|
| MLL | Mixed-Lineage Leukemia |
| NLS | Nuclear Localization Signal |
| NMR | Nuclear Magnetic Resonance |
| NOE | Nuclear Overhauser Effect |
| NOESY | Nuclear Overhauser Effect SpectroscopY |
| NTD | N-Terminal Domain |
| PDB | Protein Data Bank |
| PogZ | Pogo transposable element with Zinc finger domain |
| PSIP1 | PC4 and SFRS1-Interacting Protein |
| SEC | Size-Exclusion Chromatography |
| SID | Specific Interaction Domain |
| Sp1 | Simian virus 40 Promoter transcription factor 1 |
| SRD | Supercoil-Recognition Domain |
| TEV | Tobacco Etch Virus |
| TOCSY | TOtal Correlation SpectroscopY |
| $V_r$ | Retention volume |
| WT | Wild Type |

# 1. INTRODUCTION

Lens epithelium-derived growth factor (LEDGF/p75) is a human transcriptional co-activator that plays an essential role in HIV integration and mixed lineage leukaemia 1 (MLL1) fusion-driven leukaemia. LEDGF/p75 interacts with HIV integrase (HIV IN) and MLL1 via its integrase binding domain (IBD). Through the IBD–HIV IN interaction, LEDGF/p75 mediates several key steps in the process of integration of HIV DNA into the host cell genome. Moreover, it is also indispensable for MLL-dependent transcription and leukemogenesis. Therefore, interactions of LEDGF/p75 IBD with the HIV IN and MLL1–menin complex are considered an attractive therapeutic target for drug development.

While the pathophysiological interactions of LEDGF/p75 IBD were studied extensively, little was known about the physiological ones. In addition to MLL1, the LEDGF/p75 IBD also interacts with JPO2, PogZ, Cdc7–ASK and MLL2. In this regard, the IBD of LEDGF/p75 serves as a protein interaction hub for several cellular proteins and the HIV IN. As an epigenetic reader associated with transcriptionally active chromatin, LEDGF/p75 is considered to be a chromatin-tethering factor for its binding partners. Whether there is a yet undiscovered defining physiological function of LEDGF/p75 is still not clear. Nonetheless, its interactions with MLL1 and BMI1 are crucial for *HOX* genes regulation and their disruption leads to perinatal mortality and severe developmental defects in mice.

Interfering with such critical processes regulated by LEDGF/p75 would be undesirable. Therefore, a truly selective inhibitor that would target one of its many protein-protein interactions while not interfering with the others is required. Comparison of structural data would provide information necessary for development of such specific inhibitors. Moreover, gaining deeper insight into the interactions of IBD with cellular binding partners may contribute to the understanding of complex biological processes and facilitate future efforts in the development of efficient chemotherapies.

In this work, we set out to characterize interactions of LEDGF/p75 IBD with three binding partners: transcriptional coactivator MLL1, transcriptional repressor JPO2 and domesticated transposase PogZ. We have thoroughly characterized these proteins and their interactions with LEDGF/p75 IBD. Using NMR spectroscopy and comprehensive mutation analyses, we have mapped their interacting regions on amino acid level and obtained detailed structural information. Utilizing this information we have also identified and validated a novel LEDGF/p75 IBD interaction partner. Moreover, we have revealed how HIV IN out-competes

the cellular proteins and uncovered new challenges and opportunities for the development of selective IBD interaction inhibitors.

# 2. AIMS OF THE THESIS

This thesis is focused on physiological interactions of LEDGF/p75 IBD. The general idea is to characterize and structurally describe the physiological interactions and compare them with available data regarding the IBD–HIV IN complex. Such information could be critical for the development of selective IBD interaction inhibitors. The main goals are specifically:

- Design and clone expression vectors encoding interacting regions of LEDGF/p75 and its binding partners.
- Produce recombinant LEDGF/p75 IBD and its interaction counterparts JPO2, PogZ and MLL1 in sufficient yields and purity for further experiments.
- Characterize the protein properties and reconstitute the protein-protein complexes of IBD and its respective interaction partners. Determine the interacting regions of IBD and its binding partners on amino acid level and compare these results with previous yeast two-hybrid studies.
- Determine the structures of these protein complexes. Compare the resulting structural information with available structures of IBD–HIV IN with regard to potential inhibitor specificity.

# 3. LITERATURE REVIEW

## 3.1 LEDGF

The lens epithelium derived growth factor (LEDGF) is a human nuclear protein acting as a chromatin reader. It is a member of the Hepatoma-Derived Growth Factor (HDGF) family encoded by the PC4-and SFRS1-interacting protein 1 (*PSIP1*) gene (OMIM ID: 603620; located 9p22.3). The *PSIP1* gene contains 18 exons (Entrez Gene) and encodes mRNA of two splicing isoforms with experimental evidence [1]. Exons 1 through 15 encode LEDGF/p75 (also PSIP1) mRNA translated into a 60.1 kDa protein (UniProt identifier: O75475-1). Exons 1 through 9 and part of intron 9 encode LEDGF/p52 (also PSIP2) translated into a 37.7 kDa protein (UniProt identifier: O75475-2). Both isoforms share amino-terminal (N-terminal) section up to amino acid (aa) residue 325 which contains the common feature of all members of the HDGF family: conserved N-terminal Pro-Trp-Trp-Pro (PWWP) domain (Figure 1). This domain specifically binds tri-methylated lysine 36 of histone 3 (H3K36me3) which is a hallmark of transcriptionally active chromatin [2, 3]. It was shown to be essential for the chromatin scanning dynamics of LEDGF/p75 [4]. Towards the carboxyl-terminus (C-terminus), the PWWP domain is followed by nuclear localization signal (NLS) sequence, two AT-hooks and a supercoil-recognition domain (SRD) [5, 6]. Like the PWWP, the SRD and AT hook domains also represent chromatin interacting regions [5, 7].

LEDGF/p75 and LEDGF/p52 differ at the C-terminal region, which is extended in p75 by a 205 aa residue region including the integrase binding domain (IBD) [8]. The p75 splice isoform was shown to be more abundant in cells [9].

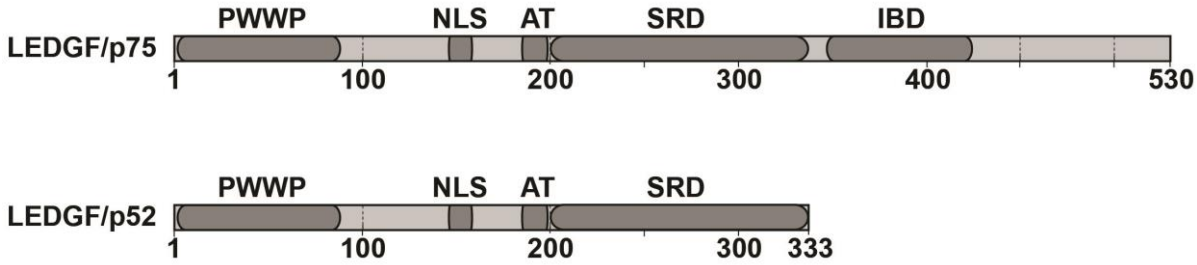


**Figure 1: Domain organization of LEDGF/p75 and LEDGF/p52**

Sequences of both splicing isoforms are schematically depicted with highlighted domains and boundaries. The common part of both isoforms is responsible for chromatin binding and nuclear localization (aa 1–325) and comprises the Pro-Trp-Trp-Pro (PWWP; aa 1–91), the nuclear localization signal (NLS; aa 148–156), the AT-hooks (AT; aa 178–198) and the supercoil recognition domain (SRD; aa 200–336 in p75). Both isoforms differ at the C-terminus. While LEDGF/p52 contains only 8 unique residues originating from intron 9, the unique C-terminus of LEDGF/p75 spans from aa 325 to 530 and comprises an integrase binding domain (IBD; aa 345–427), which is responsible for several protein-protein interactions [10].

### 3.1.1 Physiological roles of LEDGF/p75

Although the name lens epithelium derived growth factor suggests both function and expression profile, it is a misnomer, as the protein is expressed ubiquitously and has no direct role in the development of lens epithelium. Historically, a number of studies defined the role of LEDGF/p75 as a stress-related transcriptional co-activator [8, 11]. LEDGF/p75 promotes cell survival by co-activating the expression of stress response genes [12-14] and has been shown to be upregulated by simian virus 40 promoter transcription factor 1 (Sp1) under stress conditions [15]. Knockout of *PSIP1* gene leads to perinatal mortality and severe developmental defects in mice, further highlighting the importance of the LEDGF protein [16]. These developmental defects include severe homeotic skeletal malformations resembling the effects of mutations in *HOX* genes. It was shown later that LEDGF is really implicated in *HOX* genes regulation via interactions with trithorax protein family member MLL1 (myeloid/lymphoid or mixed-lineage leukemia 1) and Polycomb complex protein BMI1 [17]. *Hox* genes encode transcription factors crucial for embryonic development, body plan formation and adult hematopoiesis (reviewed by Mallo *et al.* [18] and Alharbi *et al.* [19]).

Another interaction with C-terminal binding protein interacting protein (CtIP) was shown to play important role in DNA double-strand break (DSB) homologous recombination repair pathway [20]. The interaction of LEDGF/p75 with CtIP is DNA damage-dependent and

decrease in cellular LEDGF/p75 level impairs the recruitment of CtIP to DNA DSBs and the subsequent DNA-end resection.

Several other human proteins have been identified as LEDGF/p75 binding partners, four of which interact with the IBD [21-24]. This suggests a role for LEDGF/p75 as a general chromatin-tethering factor in a variety of physiological and pathological processes [23]. In this regard, LEDGF/p75 would be an adaptor protein tethering other proteins or protein complexes to sites of transcriptionally active chromatin without having a single dominant defining function. In particular, LEDGF/p75 interacts through the IBD with mixed-lineage leukemia (MLL)/menin complex, Cdc7-activator of S-phase kinase (Cdc7-ASK), Myc-interacting protein JPO2 (also R1, RAM2, CDCA7L) and the domesticated transposase PogZ [21-24]. These interactions are the focal point of this thesis and will be described in detail hereinafter.

### 3.1.2 Pathophysiological roles of LEDGF/p75

While other physiological functions of LEDGF/p75 might yet to be discovered, its chromatin tethering properties are very important in pathological processes. Lentiviral integrases including the human immunodeficiency virus type 1 integrase (HIV IN) were first described to interact with the IBD of LEDGF/p75 [10]. LEDGF/p75 mediates several key steps in the process of integration of HIV DNA into the host cell genome. It targets integration into transcriptionally active sites of the host DNA through PWWP-mediated active chromatin recognition [25-28]. In addition, LEDGF/p75 efficiently protects HIV IN from proteasomal degradation and promotes its oligomerization and enzymatic activity [29-31]. Altogether, it was shown that LEDGF/p75 is a crucial cellular co-factor for HIV integration [32, 33]. As the p52 splice variant lacking the IBD cannot be utilized by HIV IN [33], residual integration can be mediated by hepatoma-derived growth factor related protein 2 (HRP-2), the only cellular protein besides LEDGF/p75 that contains both PWWP and IBD domains. However, HRP-2 only plays a minor role in HIV integration as it is less expressed than LEDGF/p75 and has lower affinity for HIV IN [34, 35].

Identification of the essential role of LEDGF/p75 in HIV integration and replication triggered the development of inhibitors that would target the LEDGF/p75–HIV IN interaction interface and effectively block HIV replication [36]. This effort was supported by determination of the structure of IBD bound to the dimer of the integrase catalytic core

domain (CCD; PDB code: 2B4J [37]), which revealed the IBD amino acid residues involved in the interaction with HIV IN (Figure 2).

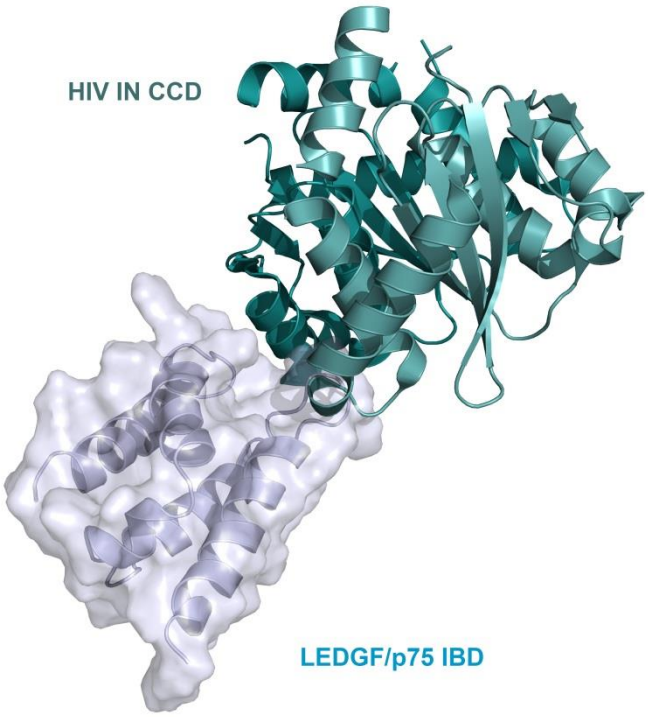


**Figure 2: Interaction of HIV IN CCD with LEDGF/p75 IBD**

The HIV IN CCD dimer interacts with interhelical loops of the IBD burying a surface of 1,280 Å$^2$. IBD residues critical for interaction with HIV IN as confirmed by mutation analysis are: Ile-365, Asp-366, Phe-406 and Val-408 [38]. Ile-365 and Asp-366 are anchored in a groove of the HIV IN CCD-CCD dimer interface. Although this interface is sufficient for effective binding to LEDGF/p75, the N-terminal domain (NTD) of HIV IN also contributes to the overall affinity [26, 39]. The IBD is a PHAT (Pseudo-HEAT repeat analogous topology) domain comprising two copies of a helix-hairpin-helix HEAT (Huntingtin/Elongation factor 3/subunit A of protein phosphatase 2A/yeast PI3K Tor1) repeat. Structure by Cherepanov *et al.*, 2005; PDB code: 2B4J [37]

Besides its role as a co-factor of lentiviral integration, LEDGF/p75 has been reported to play a role in other pathological processes including cancer and autoimmunity. Overexpression of LEDGF/p75 has been implicated in angiogenesis and lymphangiogenesis of ovarian carcinoma tumors and has also been reported in other types of cancer, such as prostate, breast, colon, liver, thyroid and uterine malignancies [40-42]. There are several roles of LEDGF/p75 in different types of cancer reported. In fusion with the nuclear pore protein 98 as a result of chromosomal translocation, LEDGF drives acute myeloid leukemic transformation of blood cells [43]. It was also shown to stabilize lysosomes and protect cancer cells from anticancer drugs triggering lysosomal cell death pathway [44]. Increased

expression of LEDGF/p75 was reported to provide survival advantage by inhibiting apoptosis (*in vitro*) and promoting chemotherapy resistance in human acute myelogenic leukemia cells [45]. Interestingly, opposite effect was observed for two fragments of LEDGF/p75 originating from its cleavage by caspase 3 and caspase 7. These protein fragments have pro-apoptotic effects and inhibit cell growth [46, 47]. Similar effect was reported for SUMOylated LEDGF/p75 [48].

The role of LEDGF/p75 in autoimmune diseases remain unclear as anti-LEDGF/p75 autoantibodies lack disease specificity and can be found both in patients with diverse conditions as well as healthy individuals [49-51]. Nonetheless, these autoantibodies have been implicated in the pathogenesis of autoimmune disorders including diverse set of inflammatory eye diseases including cataracts associated with atopic dermatitis, Vogt–Koyanagi–Harada disease, sympathetic ophthalmia, Behcet's disease, and atypical retinal degeneration (reviewed in Ganapathy *et al.*, 2004 [52] and Shinohara e*t al.*, 2002 [53]).

## 3.2 Interaction partners of LEDGF/p75 IBD

As mentioned above, the physiological function of LEDGF/p75 seems to be bringing other proteins to sites of transcriptionally active chromatin. In this regard, the IBD of LEDGF/p75 serves as a protein interaction hub for several cellular proteins and the HIV IN. The cellular interaction partners of LEDGF/p75 IBD are: JPO2, PogZ, Cdc7-ASK and the menin/MLL complex.

### 3.2.1 JPO2

JPO2 (also CDCA7L, RAM2 and R1) is a ubiquitously expressed protein encoded by *Cdca7l* (Cell Division Cycle Associated 7-Like) gene (OMIM ID: 609685, located 7p15.3). The *Cdca7l* gene contains 11 exons (Entrez Gene) and encodes mRNA of three experimentally confirmed protein-coding splicing isoforms (UniProt). The canonical isoform encodes the 454 aa residue (52.2 kDa) JPO2 protein (UniProt identifier: Q96GN5-1). It contains a predicted leucine zipper (Leu-zipper), a nuclear localization signal (NLS), and a zinc-binding RING finger-like domain (Figure 3).

It was first described in 2005 as an interaction partner of Myc proto-oncogene (c-Myc). JPO2 binds c-Myc *via* its Leucine-rich motif (amino acid residues 213–235) both *in*

*vivo* and *in vitro* and contributes to the c-Myc-mediated transformation in medulloblastoma [54]. Colony formation of medulloblastoma cell line was enhanced by JPO2 expression and inhibited by its knockdown. Analogous results were obtained more recently for hepatocellular carcinoma revealing the positive effect of JPO2 over-expression on tumor growth rate acceleration in mice [55].

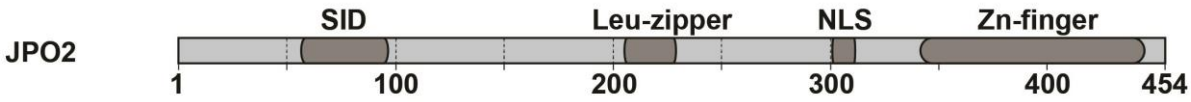


**Figure 3: Domain organization of JPO2**

JPO2 interacts with LEDGF/p75 via its specific interaction domain (SID) comprising aa residues 62–94 or 77–98, as published in two independent reports [23, 56]. The Leu-zipper (aa 213–235) has been shown to play a key role in the interaction with c-Myc [54]. The predicted nuclear localization signal (NLS; aa 301–318), and a zinc-binding RING finger-like domain (Zn-finger; aa 349–425) are depicted.

JPO2 also acts as a transcriptional repressor of monoamino oxidases A and B and regulates expression levels of these clinically important enzymes by binding to the Simian virus 40 promoter factor 1 (Sp1) sites [57-59]. Monoamino oxidase A, which primarily degrades serotonin, norepinephrine, and dopamine was shown to be involved the c-Myc-induced proliferative signaling pathway together with JPO2 [60]. Altogether, these results indicate the important role of JPO2 as a transcriptional regulator.

JPO2 has been shown to interact with the IBD of LEDGF/p75 both *in vitro* and in cell culture *via* a specific interaction domain (SID) comprising amino acid residues 62–94 or 77–98, as published in two independent reports [23, 56]. This interaction competes with LEDGF/p75 binding to HIV IN, although mutational analysis suggested that the mechanism of JPO2 binding to IBD is somewhat different from that of HIV IN [56]. In these studies, the stability of JPO2 association with chromatin was shown to be dependent on the interaction with LEDGF/p75. It was demonstrated later, that JPO2 can interact directly with chromatin on its own and only in presence of LEDGF/p75 it adopts the typical diffusion mode of PWWP-dependent chromatin scanning of LEDGF/p75 [61].

### 3.2.2 PogZ

PogZ (Pogo Transposable Element with Zinc Finger Domain) is a nuclear protein encoded by the *PogZ* gene (OMIM ID: 614787, located 1q21.3). The *PogZ* gene contains 24

exons (Entrez Gene) and encodes mRNA of at least seven possible splicing isoforms (UniProt), although most of these isoforms are lacking experimental evidence on protein level. The canonical isoform encodes 1410 aa residue (155.3 kDa) PogZ protein (UniProt identifier: Q7Z3K3-1).

PogZ was first described in 1997 as a putative polypeptide-encoding gene KIAA0461 through random isolation of a cDNA clone in size-fractionated cDNA libraries from human brain [62]. In 2000 it was still considered a putative protein as it was found to interact with the Sp1 transcription factor through its N-terminal part in yeast two-hybrid system [63].

In 2009 *in silico* analysis of its sequence identified PogZ as a domesticated transposase containing a predicted DDE (Asp-Asp-Glu) domain at its C-terminus (aa residues 1117–1323; Figure 4) [22]. This domain is typically present in enzymes that catalyze DNA cleavage at a specific site followed by a strand transfer reaction and is present in lentiviral integrases [64]. The domain contains three carboxylate residues that are responsible for coordinating metal ions necessary for catalysis. This catalytic triad is present in PogZ as amino acid residues D1126, D1229 and D1265. These residues fit the consensus $DD_{35}(D/E)$ motif that is invariant among retroviral integrases from retroviruses, yeast retrotransposons and some bacterial transposases and is crucial for enzyme activity [64]. In the same study, PogZ was identified as an interaction partner of LEDGF/p75 IBD by several independent methods. The specific IBD-interacting part of PogZ was described as a minimal interaction-sufficient region spanning aa residues 1139–1248 in the DDE domain (Figure 4) [22]. Interestingly, HIV IN also interacts with LEDGF/p75 IBD through its DDE-containing CCD [37]. Biochemical analysis confirmed that PogZ and HIV IN compete for the same binding interface on IBD [22]. In living HeLa cells, PogZ was affirmed to be tethered to chromatin by LEDGF/p75, which is probably a major determinant of its nuclear localization [61]. The function of LEDGF/p75 – PogZ interaction is yet to be discovered.

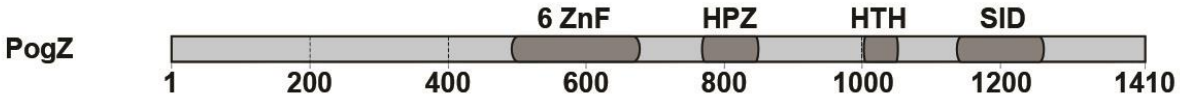


**Figure 4: Domain organization of PogZ**

The interaction with LEDGF/p75 IBD has been mapped to the specific interaction domain (SID; aa 1139–1248) which is encompassed by the DDE domain (aa 1117–1323; not depicted for clarity) [22]. PogZ contains a zinc finger array (6 ZnF; aa 473–675) and centromere protein B-like helix turn helix DNA-binding domain (HTH). The interaction with HP1α is maintained through HP1-binding zinc finger-like domain (HPZ; aa 791–850).

Intriguingly, PogZ was reported to be associated to trimethylated lysine 9 of histone 3 (H3K9me3) [65]. This epigenetic mark is generally associated with constitutive heterochromatin which is considered to be transcriptionally repressed [66, 67]. Several independent publications reported PogZ to interact with heterochromatin protein 1 alpha (HP1α) [65, 68, 69]. HP1α is an epigenetic reader with essential role in heterochromatin formation which recognizes H3K9me3 marks through its chromodomain [70, 71]. Whether the interaction between PogZ and H3K9me3 is direct or bridged by HP1α remains unclear.

Through the interaction with HP1α, PogZ destabilizes the association of HP1α with chromatin, which is required for correct chromosome segregation during mitosis [69]. Furthermore, PogZ is also essential for Aurora B activation, which by phosphorylating serine 10 of histone 3 further dissociates HP1α from chromatin during mitosis [69]. Additionally, POGZ was shown to interact with mitotic spindle checkpoint protein MAD2L2 underscoring its implication in cell-cycle regulation [65].

Recently, several mutations in *PogZ* gene were linked with autism spectrum disorder development [72, 73].

### 3.2.3 MLL1/MLL2

MLL1 (Myeloid/Lymphoid or Mixed-Lineage Leukemia Protein 1) is an essential transcriptional coactivator encoded by the *KMT2A* (Lysine-Specific Methyltransferase 2A) gene. The *KMT2A* gene (OMIM ID: 159555 located 11q23.3) contains 37 exons (Entrez Gene) and encodes mRNA of at least three possible splicing isoforms (UniProt). The canonical isoform encodes 3969 aa residue (431.7 kDa) MLL1 protein (UniProt identifier: Q03164-1). MLL1 has been studied in a variety of contexts and as a result it has been labelled by several aliases (HRX, HTRX1, WDST, ALL-1, CXXC7, TRX1 etc.).

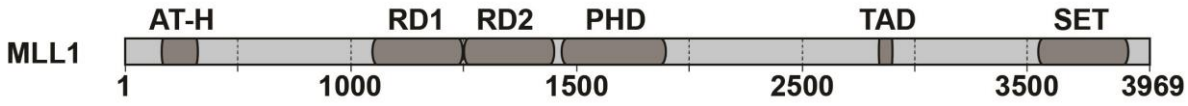


**Figure 5: Schematic domain organization of MLL1**

MLL1 contains AT-hook motifs (AT-H; aa 169–309) [74], transcriptional repression domains (RD1 and RD2; aa 1101–1250 and 1251–1400) [75] followed by four interspersed epigenetic reader PHD fingers (Plant HomeoDomain; aa 1431–1958) [76]. At the C-terminus the transactivation domain (TAD, aa 2829–2883) and the histone methyltransferase SET domain (aa 3829–3945) are located [77, 78].

MLL1 is post-translationally processed by threonine aspartase 1, which cleaves out a fragment between aa residues 2667 and 2718 [79]. Fragments 1–2666 and 2719–3969 form the functional and stable MLL1 heterodimer by non-covalent association [79, 80]. MLL1 contains several domains including transcription repression domains (RD1 and RD2) and transactivation domain (TAD) [75, 77]. While the RDs mediate repression of transcription via interactions with histone deacetylases 1 and 2, the TAD acts as a transcription activator by recruiting the CREB-binding protein [77, 81]. The SET domain (suppressor of variegation, enhancer of zeste and trithorax) possesses histone methyltransferase (HMT) activity which is pivotal for the function of MLL1. This domain methylates lysine 4 of histone 3 (H3K4) creating H3K4me2 and H3K4me3 epigenetic marks that represent a hallmark of transcriptionally active gene promoters [78, 82].

MLL1 is a vital regulator of homeobox (*Hox*) gene promoters and is essential in hematopoiesis and body development [83, 84]. It maintains the expression of *Hox* genes encoding transcription factors crucial for embryonic development, body plan formation and adult hematopoiesis (reviewed by Mallo *et al.* [18] and Alharbi *et al.* [19]). Knockout of MLL1 in mice leads to embryonic lethality and severe developmental defects underscoring its importance in morphogenetic processes [85]. Chromosomal translocations resulting in alterations of the MLL1-encoding *KMT2A* gene are causing aggressive acute lymphoblastic and myeloid leukemias (reviewed by Krivtsov *et al.* [86]). The *KMT2A* gene is susceptible to rearrangements as it contains a breakpoint cluster region accommodating many Topoisomerase II cleavage sites and nuclear matrix attachment points between exons 8 and 13 [87]. These leukemia-causing rearrangements most frequently result in expression of the N-terminal part of MLL1 (up to a breakpoint typically in the PHD fingers region) in fusion with an unrelated protein [88]. Surprisingly, only seven of these unrelated C-terminal fusion protein partners of MLL1 account for approximately 90% of acute leukemias in patients [88]. Interestingly, these major MLL fusions share a common pathway resulting in increased and

extended H3K79 methylation signatures [89]. The cause of rearrangements in the *KMT2A* gene is probably aberrant non-homologous end joining of DSBs in hematopoietic cells [90]. However, the exact molecular mechanisms of the causes and the leukemogenesis effects are different among the oncogenic fusion products and are yet to be fully elucidated.

Although MLL1 can interact with DNA on its own, it was shown to be tethered to chromatin by LEDGF/p75 in complex with tumor suppressor menin [21]. Menin is a putative tumor suppressor protein encoded by the Multiple Endocrine Neoplasia 1 (*Men1*) gene. Multiple endocrine neoplasia type 1 is a hereditary disease causing formation of multiple tumors throughout endocrine organs in patients (mostly parathyroid; reviewed by Giri *et al.* [91]). This severe syndrome is caused by germline inactivating mutations in the *Men1* gene scattered across the coding sequence [92]. The physiological role of menin is not clear, although it was revealed to regulate transcription of certain cyclin-dependent kinase inhibitors by recruiting MLL1 to promoter regions of these genes [93]. It was shown in the same report, that loss of function of menin causes deregulated cell growth. On the other hand, unmutated menin acts as an oncogenic agent in MLL1-fusion-associated leukemia [21, 94]. Menin thus exhibits an unusual ability to both promote leukemogenesis in the hematopoietic cell lineage and suppress tumorigenesis in the endocrine cell lineage. Analogously to menin, LEDGF/p75 is also indispensable for both MLL-dependent transcription and leukemogenesis [21, 95].

The structure of the LEDGF/p75–MLL1–menin ternary complex was studied by x-ray crystallography (PDB code: 3U88; [96]). Although only protein fragments and truncated constructs were used in this study, important structural features of this complex assembly were revealed (Figure 6). 580 out of 610 aa residues of menin are resolved in the structure revealing that this sequence comprises one large globular domain. An artificial 75 aa MLL1 peptide comprising aa residues 6–153 while excluding aa residues 16–22 and 36–102 was used for crystallization. This section of MLL1 remains largely unstructured upon the interaction with menin and regions 16–22 and 36–102 had to be excluded as they created a fuzzy complex and prevented crystallization. Only a 54 aa residue part of the MLL1 peptide is resolved in the structure, most of it stably anchored on the surface of menin. A small portion of this peptide is in contact with LEDGF/p75 IBD. The IBD–MLL1 interface is formed by MLL1 aa F129 and F133 and IBD aa I397, Y420, K424 and F427. Limited contacts between IBD and menin are maintained by IBD residues E414 and N425 and menin residues R92 and S104. Using these protein constructs, only a combination of both these interaction interfaces of menin and MLL1 could maintain the interaction with LEDGF/p75 IBD. [96]

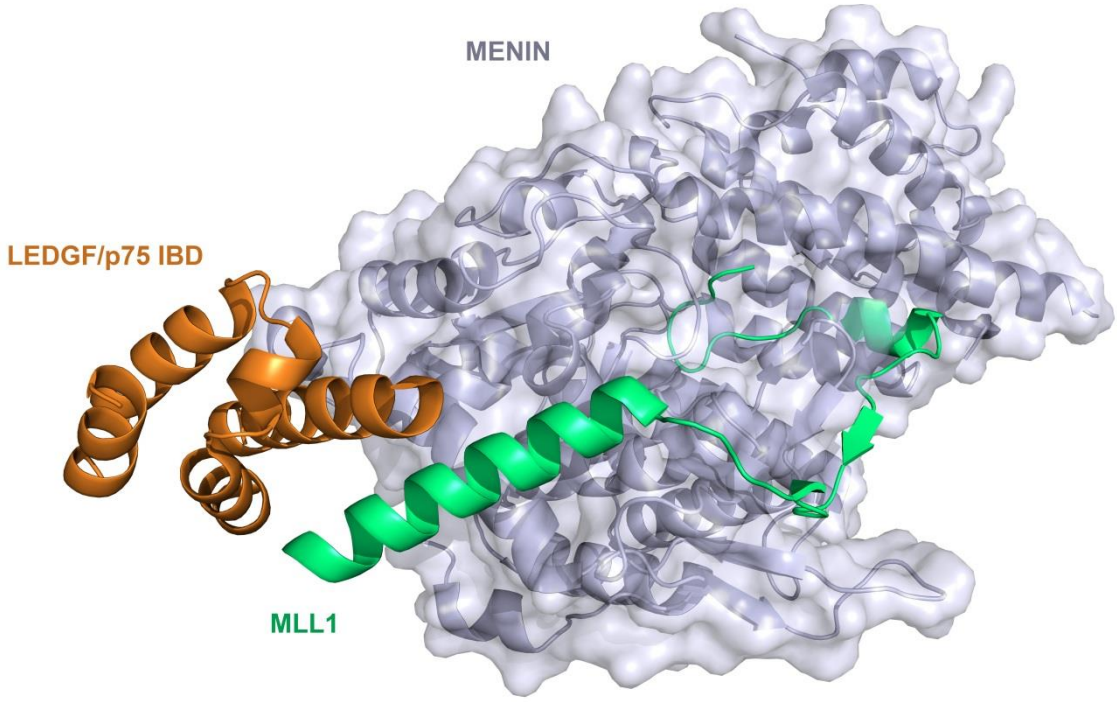


**Figure 6: Structure of the LEDGF/p75–MLL1–menin ternary complex**

Crystal structure as published by Huang *et al.* (PDB code: 3U88) [96]. The ternary complex consists of truncated constructs of the interaction partners: LEDGF/p75 IBD (here 67 aa residues from region spanning aa 353–429 in orange), MLL1 (here 54 aa residues from construct comprising aa residues 6–15 followed by aa 23–35 and 103–135 in green) and menin (here 550 aa residues from region spanning aa 2–581 in blue tint). Alpha helix 5 of IBD is anchored between MLL1 and menin.

Although the physiological role of LEDGF/p75 in the LEDGF/p75–MLL1–menin ternary complex is not clear, disrupting the LEDGF/p75–MLL1 interaction by inhibitors was shown to have anti-leukemic effect in cell cultures and in a mouse model [95, 97]. The chromatin-tethering function of LEDGF/p75 mediated by the PWWP domain is crucial in MLL1 fusion-driven leukemia [21]. The role of menin in this process is most probably only promoting the LEDGF/p75–MLL1 interaction [21].

MLL2 (Mixed-Lineage Leukemia Protein 2) is the closest paralog of MLL1 encoded by the *KMT2B* gene. The *KMT2B* gene (OMIM ID: 606834 located 19q13.12) contains 38 exons (Entrez Gene) and encodes mRNA of at least two possible splicing isoforms (UniProt). The canonical isoform encodes 2715 aa residue (293.5 kDa) MLL2 protein (UniProt identifier: Q9UMN6-1). There is a nomenclature confusion in the literature regarding MLL2, which is often referred to as MLL4 (also HRX2, KIAA0304, TRX2, WBP7).

In humans, there are at least eight H3K4-specific histone methyltransferases that include MLL1, MLL2, MLL3, MLL4, MLL5, hSet1A, hSet1B and ASH. Although MLL2

has analogous function as MLL1, it was demonstrated that these proteins act as distinct multiprotein complexes with several common subunits [84]. MLL2 was shown to interact with menin, which recruits it to promoter regions of estrogen-responsive genes and regulates their expression in an estrogen-dependent manner [98]. LEDFG/p75 represents another shared interaction partner of MLL1 and MLL2 [3, 21].

### 3.2.4 ASK

ASK (Activator of S-phase Kinase; also ZDBF1, DBF4A, and CHIF) is a regulatory subunit of cell division cycle 7 (CDC7) kinase encoded by the *DBF4* (Dumbbell Former 4) gene. The *DBF4* gene (OMIM ID: 604281, located 7q21.3) contains 12 exons (Entrez Gene) and encodes 2 splicing isoforms of ASK (UniProt). The canonical isoform encodes the 674 aa residue (76.8 kDa) ASK protein (Uniprot identifier: Q9UBU7-1).

ASK was originally described in 1982 as a yeast gene *DBF4* whose mutations inhibited DNA synthesis and arrested budding in a typical dumbbell morphology [99]. In 2000, sequence alignment of ASK-related proteins revealed only three short conserved stretches in the ASK polypeptide chain. These regions were called the N, M and C motif (Figure 7) [100]. The extent of these regions was surprising as DNA replication-related proteins are among the most conserved in eukaryotes. Motifs C and N were later revealed to be responsible for interaction of ASK with CDC7 as crystal structure of the CDC7–ASK complex was resolved by Hughes *et al.* [101]. Motif C is essential and sufficient to maintain this interaction and was identified as a zinc finger [101, 102].

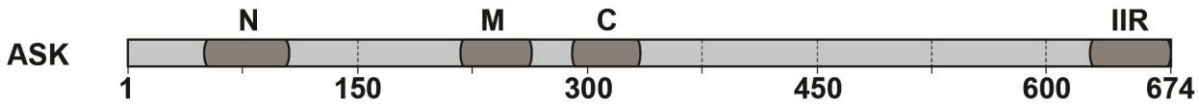


**Figure 7: Domain organization of ASK**

The ASK sequence contains three conserved motifs: motif-N (N; aa residues 48–92), motif-M (M; aa residues 214–265), and motif-C (C; aa residues 291–331) almost completely overlapping with later identified zinc finger (aa residues 290–335). The minimal region responsible for interaction with LEDGF/p75 IBD was narrowed down to the last 50 aa residues and is depicted as IIR (IBD Interaction Responsible)

CDC7 is a conserved serine/threonine kinase crucial for activation of DNA replication checkpoint, meiosis, chromosome cohesion and mitosis exit (reviewed by K. Labib [103] and

Matthews *et al.* [104]). CDC7 activity is also essential for ATR-dependent Chk1 activation in the DNA replication checkpoint [105, 106]. The essential roles of both CDC7 and ASK in mammalian cell growth and proliferation were demonstrated using conditional knockout cell lines. In both cases, DNA synthesis was arrested and cell death was induced in the mutant cells [107, 108].

The stably expressed CDC7 is activated by the ASK, whose expression levels are tightly controlled and correlate with late $G_1$ and S phase [109]. It was shown by Kitamura *et al.* that ASK is required for ATP binding to the catalytic subunit of CDC7 and also for substrate recognition. In many cancers and tumor cell lines, both CDC7 and ASK are overexpressed underlining the importance of the CDC7–ASK complex in cell proliferation [110]. CDC7–ASK has therefore become a drug development target for potential cancer therapy (reviewed by Montagnoli *et al.* [111]).

Interaction between CDC7–ASK heterodimer and LEDGF/p75 was thoroughly characterized by Hughes *et al.* in 2010 [24]. The IBD of LEDGF/p75 was found to be critical and sufficient to maintain the interaction and also to stimulate the kinase activity of CDC7–ASK *in vitro*. Phosphorylation of the CDC7–ASK heterodimer is necessary for CDC7–ASK–LEDGF/p75 complex formation. In a more detailed look, a basic patch on IBD comprising aa residues K360, K401, K402 and R405 is crucial for interaction with a presumably phosphorylated region of ASK. This region was narrowed down to last 50 C-terminal aa residues of the ASK. Interestingly, deletion of this region resulted in a hyperactive form of CDC7–ASK. This suggests an auto-inhibition role of the ASK C-terminus which is neutralized by the interaction with LEDGF/p75. Upon the formation of the CDC7–ASK–LEDGF/p75 complex, LEDGF/p75 is recognized as a substrate and phosphorylated by CDC7–ASK. There were four phosphorylation targets identified in LEDGF/p75 sequence *in vitro*: S273, S275, S206 and S208. Intriguingly, all these serines are localized in the largely unstructured SRD, which possesses a weak but specific affinity for negatively supercoiled DNA (Figure 1) [5]. However, out of these residues only S206 was confirmed to be phosphorylated by CDC7–ASK *in vivo*. Moreover, there was no phosphorylation detected in the first 145 N-terminal aa residues containing the PWWP domain as a major chromatin localization determinant of LEDGF/p75. Thus the role of these phosphoserines in LEDGF/p75 remains speculative along with the role of the CDC7–ASK–LEDGF/p75 complex in cells. [24]

Notably, there is a functional link between ASK and another interaction partner of LEDGF/p75 IBD. ASK was revealed to interact with the C-terminus of menin by Schnepp *et*

*al.* [112]. While not changing the kinase activity of CDC7–ASK, menin was shown to completely repress CDC7–ASK-induced cell proliferation [112].

### 3.2.5 HIV IN

The HIV IN belongs to the retroviral IN superfamily and a much larger DDE superfamily sharing the catalytic triad aa signature (Asp, Asp, Glu) and the ribonuclease H fold of the catalytic domain [113]. HIV IN consists of three domains: three-helical zinc-binding N-terminal domain (NTD), catalytic core domain (CCD) and C-terminal domain (CTD) with SH3 fold (reviewed by Li *et al.* [114]). It carries out one of the critical steps in HIV replication – integration of the viral linear DNA into the host genome. Only in the proviral state, efficient viral gene expression and persistent latent reservoir are achieved by the virus. The process *per se* consists of two endonucleolytic reactions. First, the HIV IN binds to the long terminal repeats (LTRs) at both ends of the viral DNA soon after it is synthesized by the reverse transcriptase in form of linear dsDNA. The LTRs are processed by the HIV IN in the first endonucleolytic reaction hydrolyzing 3′ phosphodiester bonds and yielding reactive adenosine 3′-hydroxyl groups. These hydroxyls are then used for nucleophilic attack on the phosphodiester bonds on opposing strands of target cellular DNA in the strand transfer reaction [115]. The resulting integration intermediate with free 5′ ends of viral DNA is then repaired by cellular enzymes producing fully integrated provirus. For a more detailed overview of HIV integration see Krishnan *et al.* [116] and Li *et al.* [114].

Since the 1990's, there was an effort to find inhibitors of HIV IN that would block the strand transfer reaction and prevent HIV integration. This effort was successful in 2000 when improved screening assay design led to the discovery of certain diketo acid compounds as potent inhibitors of the strand transfer reaction [117]. In 2007 the first HIV IN inhibitor Raltegravir was approved for use in patients improving therapy of HIV variants resistant to older drugs [118]. As of today, two other HIV IN inhibitors have been approved for therapy. In 2012 it was Elvitegravir and finally in 2013 Dolutegravir, a second generation HIV IN inhibitor. Thus far, Dolutegravir seems to be unique amongst all antiretrovirotics, as no *de novo* resistance has been reported in previously treatment-naïve patients (reviewed by Mesplede *et al.* [119]).

Analogously to other steps of HIV replication, integration also depends on a host of cellular co-factors. As mentioned hereinbefore, LEDGF/p75 is a prominent cellular co-factor

of HIV IN, whose crucial role in HIV integration was confirmed by many independent methods. This generated a broad interest in developing inhibitors that would block the interaction between LEDGF/p75 and HIV IN.

## 3.3 LEDGF/p75–HIV IN interaction inhibitors

Since the clinical characterization of acquired immune deficiency syndrome (AIDS) in 1981, it has caused an estimated 36 million deaths worldwide (as of 2012 according to UNAIDS). Although HIV replication can be stably repressed by optimal highly active antiretroviral therapy (HAART), there is no definite cure for AIDS. Therefore, novel antivirals that would act complementarily to current inhibitors of HIV reverse transcriptase, protease and integrase are required.

As indicated hereinbefore, the interaction between LEDGF/p75 and HIV IN drew attention of both academic and industrial research as a potential therapeutic target in HIV replication. Several studies indicated that this interaction might indeed be a druggable target. In particular, reducing the level of LEDGF/p75 in cell by different methods lowered the efficiency of HIV integration as much as 10-fold [33, 35, 120]. Moreover, overexpression of IBD alone as a competing agent to native LEDGF/p75 reduced HIV integration over 100-fold [121]. These results served as a proof of concept and development of small-molecules that would target the well-defined LEDGF/p75-binding pocket in dimer interface of the HIV IN begun. Such molecules were named LEDGINs and their development was greatly facilitated by structural characterization of the LEDGF/p75-HIV IN interface allowing successful structure-based drug design [36]. Further improvement in LEDGIN potency revealed a dual mechanism of their HIV IN activity inhibition. Additionally to blocking the LEDGF/p75-HIV IN interaction, they stabilize an inactive dimeric form of HIV IN resulting in allosteric inhibition of its activity [122, 123].

Interestingly, a reciprocal approach was also utilized targeting the LEDGF/p75 IBD by peptides mimicking the interaction interface of HIV IN [124]. Although the interaction interface on IBD was proposed to be generally overlapping for its physiological binding partners such as JPO2 and PogZ, differences were also reported by mutation analyses [22, 23, 56]. Whether there is enough divergence in these interactions to allow creation of selective inhibitors will be discussed in more detail hereinafter.

As of today, the most potent LEDGIN developed (CX014442) exhibits antiviral activity in nanomolar concentration range while being highly selective [123]. Although HIV IN resistance mutations that render LEDGINs inactive were reported, there seems to be a great potential in combining them with inhibitors of strand transfer reaction (described in Chapter 3.2.5) [123]. It is logical that targeting two sites on HIV IN at the same time would create additional selection pressure and could increase the number of mutations needed for resistance development to the extent that would destabilize the CCD of HIV IN. On the other hand, as highly effective HAART is available, introduction of novel anti-HIV therapeutics and their combinations to the market is not currently prioritized by pharmaceutical companies for economic reasons.

# 4. MATERIALS AND METHODS

## 4.1 Materials

### Bacterial strains

**One Shot® Max Efficiency® DH5α™-T1$^R$ *Escherichia coli*** (Invitrogen).

Genotype:

F- Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk-,mk+)*pho*A *sup*E44*thi*-1 *gyr*A96 *rel*A1 *ton*A.

***Escherichia coli* BL21 (DE3)** (New England BioLabs)

Genotype:

*fhuA2 [lon] ompT gal (λ DE3) [dcm] Δ hsdS*

*λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*

Bacterial strains were cultivated under sterile conditions in lysogeny broth (LB) media (liquid or solid). Ampicillin antibiotic (Serva) in final concentration 100 μg/ml was used for colony selection.

## 4.2 Methods

### 4.2.1 Cloning

For expression in *Escherichia coli*, coding sequences of the LEDGF/p75 IBD (aa residues 345–426), PogZ (amino acid residues 1117–1410), IWS1 (amino acid residues 352–548) and JPO2 (aa residues 1–130) were amplified and cloned into a modified pMCSG7 vector (T7 promoter-driven) [125]. This vector encodes an N-terminal $His_6$ affinity tag followed by the tobacco etch virus (TEV) protease recognition site. In the IWS1 coding sequence, residue 547 (arginine) was mutated to tryptophan using a modified reverse amplification primer. Upon TEV cleavage, the encoded proteins retained a cloning artifact of 5 amino acid residues (SNAAS) at their N-termini. $PogZ_{1117–1410}$, $IWS1_{352–548}$ and $JPO2_{1–130}$ were overexpressed from modified pMCSG7 vectors.

Other prokaryotic expression constructs were prepared at KU Leuven. Preparation of Flag-LEDGF/p75, MBP-JPO2 and MBP-PogZ$_{1117–1410}$ expression constructs was described in detail before (pCPNatFlag-p75 [22], pMAL-RAM2 [26] and pMAL-NpogZ [22], respectively). To create the expression plasmid for His-IWS1, IWS1 was amplified from pCR-Blunt II-TOPO-IWS1 (Source BioScience) and cloned into pDONR221 using the Gateway system (Life Technologies) according to the manufacturer's protocol. IWS1 was subsequently transferred from the pDONR221 plasmid to pHXGWA using the Gateway protocol [126]. To clone the plasmids for MLL1-GST expression and purification, GST was cloned into pET-20b digested with *Eco*RI/*Xho*I. The resulting plasmid was subsequently digested with *Nde*I/*Bam*HI and used to insert the sequence coding for MLL1$_{1-160}$ (PCR-amplified from a MLL1 synthetic gene and digested with *Nde*I and *Bgl*II). The His-tagged HIV integrase construct was also described earlier (*pKB-IN$_6$H* [56]). Point mutations were introduced into plasmids by site-directed ligase-independent mutagenesis (SLIM) [127].

## 4.2.2 Expression and purification of proteins

### 4.2.2.1 Transformation of competent bacterial cells

Competent *Escherichia coli* BL21 (DE3) cells were transformed with 1 µl of expression vector (25 – 50 ng of DNA) per one aliquot containing 100 µl of competent cells thawed on ice. This mixture was incubated 30 minutes on ice and then heat-shocked by 42 °C for 30 seconds and immediately transferred back on ice. In the next step, 900 µl of LB medium was added to the mixture and it was incubated for one hour in 37 °C pre-heated incubator while shaking at 150 RPM. After the incubation, 500 µl of the culture was plated on a big (20 cm in diameter) Petri dish with LB agar containing 1% glucose and 100 µg/ml ampicillin. The dish was incubated overnight at 37 °C.

Bacterial colonies grown on the agar were washed out with 10 ml of LB medium. The resulting suspension was used as inoculum. Flasks containing 500 ml of LB media supplemented with 2.5 ml of 85% glycerol and 100 µg/ml ampicillin we inoculated by adding such amount of the inoculum suspension to achieve final $OD_{550}$ = 0.1 value. The OD550 values were measured using spectrophotometer Ultrospec 3100*pro*.

#### 4.2.2.2 Protein expression

Unlabeled IBD, $PogZ_{1117–1410}$, $IWS1_{352–548}$ and $JPO2_{1–130}$ proteins were overexpressed from modified pMCSG7 vectors in *E. coli* BL21 (DE3) cells grown at 30 °C in LB broth (Sigma) supplemented with 0.8% glycerol and 100 μg/ml ampicillin. When cultures reached an $OD_{550nm}$ of approximately 0.8, heterologous expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG). The IPTG concentration, duration and temperature of cultivation were optimized for individual proteins to achieve maximal protein yields. Optimal conditions were 0.25 mM IPTG, 5 h at 30 °C for JPO2; 0.2 mM IPTG, 4 h at 20 °C for the IBD; 0.25 mM IPTG, 4 h at 30 °C for $PogZ_{1117–1410}$ and 0.2 mM IPTG, 4 h at 30 °C for $IWS1_{352–548}$.

Full-length His-IWS1 was expressed in *E. coli* BL21 (DE3) at 30 °C. Protein expression was induced with 0.2 mM IPTG at an $OD_{600nm}$ of 0.8, and the culture was harvested after 3 h.

Uniformly $^{15}N$- and $^{15}N/^{13}C$-labeled IBD, $PogZ_{1117–1410}$ and $JPO2_{1–130}$ were produced from cells grown in minimal media containing $^{15}N$-ammonium sulfate and $^{13}C$-*D*-glucose as the sole nitrogen and carbon sources. The media were supplemented with 20 μg/ml ampicillin and cells were grown at 30 °C for 6 h before being transferred to 20 °C. After temperature stabilization, heterologous expression was induced by addition of IPTG. The final concentration of IPTG for both IBD and $JPO2_{1–130}$ expression was 0.2 mM with cultivation for 15 h at 20 °C; for $PogZ_{1117–1410}$ we used 0.2 mM IPTG, 16 h at 20 °C and for $IWS1_{352–548}$ 0.2 mM IPTG, 14 h at 18 °C. The purification protocol for labeled proteins was the same as for the corresponding recombinant protein fragments expressed by cells grown in LB medium.

Flag-tagged LEDGF/p75 or derived mutants were expressed and purified the same way as non-tagged LEDGF/p75 [56]. MBP-JPO2 [26], MBP-$PogZ_{1117–1410}$ [22] and HIV IN-His [56] were purified as described before. $MLL_{1–160}$-GST and mutant constructs were expressed in *E. coli* Rosetta2 (DE3) grown on LB medium supplemented with 20 μg/ml ampicillin. Bacterial cultures were grown at 37 °C. Protein expression was induced with 0.5 mM IPTG at an $OD_{600}$ of 0.6. Cultures were harvested after 5 hours. Pellets were washed in 20 ml STE buffer (10 mM Tris pH 7.3; 100 mM NaCl, 0.1 mM EDTA), and stored at -20 °C.

#### 4.2.2.3 Protein purification

Bacterial cells were harvested by centrifugation and re-suspended in lysis buffer. The lysis buffer for the IBD and $IWS1_{352–548}$ contained 25 mM sodium phosphate, pH 7.8, 1 M NaCl, 40 mM imidazole and 0.05% β-mercaptoethanol (BME). The composition of $PogZ_{1117–1410}$ lysis buffer was 25 mM sodium phosphate, pH 7.8, 1 M NaCl, 40 mM imidazole and 0.05% BME. For lysis of $JPO2_{1–130}$, buffer containing 25 mM sodium phosphate, pH 8.0, 1 M NaCl, 15 mM imidazole and 0.05% BME was used. EDTA-free cOmplete protease inhibitor cocktail (Roche) tablets were added to the cell suspensions to block protease activity. Bacterial cells were disrupted by three passages through an Emulsiflex cell disrupter at 30 kpsi. The lysate was centrifuged to remove insoluble materials, and the supernatant was used for further purification steps. The supernatant was clarified by filtration through a sterile Millex filter unit (porosity 0.8 µm; Millipore) and subjected to affinity chromatography using a HisTrap$^{TM}$ HP 5 ml column (GE Healthcare) column equilibrated with lysis buffer. The $His_6$-tagged protein was eluted in a gradient of lysis buffer supplemented with 400 mM imidazole, which was then eliminated from the sample by dialysis against lysis buffer without imidazole. The samples were incubated overnight with recombinant $His_6$-tagged TEV protease at 16 °C to remove the affinity tag. Following TEV cleavage, a second round of His-Select Nickel Affinity Gel affinity chromatography was performed to remove TEV protease and un-cleaved $His_6$-tagged protein. The sample was then concentrated using Amicon Ultra Concentrators (Millipore). The final purification step was size exclusion chromatography on a Superdex 200 10/300 GL column using an Äkta Basic FPLC system (Amersham Biosciences) equilibrated in lysis buffer without imidazole.

This purification protocol yielded 16.5 mg of purified IBD and 7.1 mg of purified $PogZ_{1117–1410}$ per liter of bacterial culture. The final yields of purified $JPO2_{1–130}$ and $IWS1_{352–548}$ were 5.1 mg and 9 mg, respectively, per liter of bacterial culture. The identities of purified $JPO2_{1–130}$ and $IWS1_{352–548}$ were confirmed by intact mass measurements (MALDI-TOF) and peptide mass fingerprinting (LC-MS), as there were irregularities in their gel migration behavior typical for unstructured proteins.

Cells expressing full-length His-IWS1 were washed in 20 ml STE buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA). Cell pellets were re-suspended in lysis buffer [50 mM Tris-HCl, pH 7.5; 150 mM NaCl, 20 mM imidazole, 0.1 µg/ml DNAse (Thermo Scientific) and protease inhibitor (cOmplete, EDTA-free, Roche)] and lysed by sonication. The lysate was cleared by centrifugation at 19800×g for 30 min at 4 °C and subjected to

affinity chromatography using His-Select Nickel Affinity Gel (Sigma) equilibrated with wash buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole). His-tagged proteins were eluted with wash buffer containing 250 mM imidazole. Fractions were analyzed by SDS-PAGE for protein content. Peak fractions were dialyzed against a 100-fold excess of 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 10 % (v/v) glycerol at 4 °C overnight.

Purification of $MLL_{1–160}$-GST was carried out by affinity chromatography on Glutathione Sepharose-4 Fast Flow (GE Healthcare). Cells were lysed in buffer containing 50 mM Tris-HCl pH 7.3, 150 mM NaCl, 1 mM DTT, 0.1 µg/ml DNAse (Thermo Scientific) and protease inhibitor (cOmplete, EDTA free, Roche). The resin was equilibrated with wash buffer (50 mM Tris-HCl pH7.5, 100 mM NaCl, 1 mM DTT) and GST-tagged proteins were eluted in wash buffer supplemented with 25 mM Glutathione. $MLL_{1–160}$-GST wild type as well as mutants were further purified over a HiTrap Heparin HP resin (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT and eluted with a salt gradient using FPLC Äkta Purifier system (GE Healthcare). Fractions were analyzed by SDS-PAGE for protein content. Peak fractions were supplemented with 10% (v/v) glycerol and stored at -80 °C.

### 4.2.3 Analytical size exclusion chromatography

Analytical SEC was performed using the Äkta Basic FPLC system (Amersham Biosciences) on a Superdex 200 10/300 GL Tricorn column (Pharmacia). The column was equilibrated with buffer containing 50 mM Tris-HCl, pH 7.8, 150 mM NaCl and 0.05% β-mercaptoethanol. The protein standards (Sigma-Aldrich) used for molecular weight estimation were blue dextran (2000 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome c (12.4 kDa) and aprotinin (6.5 kDa).

### 4.2.4 Limited trypsin proteolysis

Final concentrations of 0, 0.01, 0.1, 1 and 10 µg/ml trypsin were used to digest $PogZ_{1117–1410}$ (final concentration 1 mg/ml) for 30 min at 37 °C in reaction buffer containing 30 mM Tris-HCl, pH 7, 150 mM NaCl, 0.05% β-mercaptoethanol and 5 mM $MgCl_2$. Analyzed proteins were excised from 15% SDS-PAGE stained with SimplyBlue SafeStain (Invitrogen) and subjected to N-terminal protein sequencing. The reaction mixture was

analyzed by MALDI-TOF accordingly to previously described protocols [128, 129] in the Laboratory of Mass Spectrometry, IOCB, ASCR v. v. i., Prague, Czech Republic.

### 4.2.5 Mass spectrometry

For peptide mass fingerprinting using LC-MS, the proteins were excised from a 15% SDS-PAGE gel stained with SimplyBlue SafeStain (Invitrogen) and digested with trypsin. Over 90% sequence coverage with more than 95% identification confidence was achieved in the peptide mass fingerprinting of both $JPO2_{1–130}$ and $IWS1_{352–548}$. The LC-MS and MALDI-TOF analyses were performed according to previously published protocols [128, 129] in the Laboratory of Mass Spectrometry, IOCB, ASCR v. v. i., Prague, Czech Republic.

### 4.2.6 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) analysis using a real-time PCR LightCycler 480 II (Roche) was performed to assess the stabilizing effects of various buffers and their effects on complex formation. Sixteen buffers (100 mM) covering the near neutral pH range were tested in the presence or absence of 200 mM NaCl as described by Ericsson *et al*. [130]. The same set of buffers with 200 mM NaCl and additional additives was used to find optimal conditions for IBD–$JPO2_{1–130}$ complex formation with special regard to NMR measurements. These experiments were based on methods published by Kopec *et al*. [131]. Proteins (0.1 mg/ml) were assayed in the presence of 8× Sypro Orange dye (Invitrogen) in a total reaction volume of 25 µl. The plates were sealed with LightCycler 480 Sealing Foil (Roche), and a temperature gradient from 20 °C to 95 °C with a rate of 1 °C/min was applied. The fluorescence intensity of Sypro Orange was recorded at 580 nm, with an excitation wavelength of 465 nm.

All experiments were performed at least in duplicate. Melting temperature values were determined using the Roche LightCycler 480 SW 1.5 software as the minima of the negative first derivative curves calculated from the melting curves. Control experiments were carried out for IBD–$JPO2_{1–130}$ complex measurements to distinguish buffer and binding effects.

### 4.2.7 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a MicroCal Auto-iTC200 System (GE Healthcare) at 25 °C. The samples were prepared in 50 mM Tris-HCl, pH 7.8, 150 mM NaCl, 0.05% BME. In the case of IBD–$MLL_{140–160}$ titration, the samples were prepared in 50 mM Tris-HCl, pH 7.8, 150 mM NaCl, 2.17% dimethyl sulfoxide, 0.05% 2-mercaptoethanol. Protein concentrations were determined by amino acid analysis. For the IBD–$MLL_{140–160}$ complex, 2 µl aliquots of 650 µM $MLL_{140–160}$ were injected stepwise into a sample cell containing 200 µl of 50 µM IBD. For the IBD–$JPO2_{1–130}$ complex, 2 µl aliquots of 60 µM $JPO2_{1–130}$ were injected stepwise into a sample cell containing 200 µl of 10 µM IBD. For the IBD–$PogZ_{1117–1410}$ complex, 2 µl aliquots of 114 µM $PogZ_{1117–1410}$ were injected stepwise into a sample cell containing 200 µl of 10 µM IBD. The samples for IWS1 titration experiments were prepared in 50 mM Tris-HCl, pH 8.5, 150 mM NaCl, 0.05% BME. For the IBD–$IWS1_{352–548}$ complex, 2 µl aliquots of 145 µM $IWS1_{352–548}$ were injected stepwise into a sample cell containing 200 µl of 15 µM IBD.

Each assay was accompanied by a control experiment in which the binding buffer was titrated with the injected protein alone. The dilution heat values obtained for the titration to the buffer were then subtracted from those obtained for the complex formation. All experiments were performed at least in triplicate. The association constants and stoichiometry (N) were estimated using MicroCal Origin software (GE Healthcare). Known variances of each measurement were utilized to calculate the weighted average of the dissociation constant ($K_d$) and stoichiometry as a maximum likelihood estimator. The standard deviation was determined from three independent experiments.

### 4.2.8 NMR spectroscopy

All NMR spectra were acquired at 25 °C on a 600 MHz Bruker Avance spectrometer equipped with triple-resonance ($^{15}N/^{13}C/^{1}H$) cryoprobe. The sample volume was 0.35 ml, with protein concentrations ranging from 0.1 mM to 0.35 mM for free $JPO2_{1–130}$, IBD, $IWS1_{352–548}$ and $PogZ_{1117–1410}$ proteins as well as IBD–$JPO2_{1–130}$, IBD–$IWS1_{352–548}$ and IBD–$PogZ_{1117–1410}$ complexes in 25 mM HEPES, pH 7.0, containing 100 mM NaCl, 0.05% BME, 5% $D_2O$/95% $H_2O$. Essentially complete sequence-specific resonance backbone assignments for the free IBD were obtained from 3D HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO spectra. Side-chain chemical shifts were assigned from CC(CO)NH-TOCSY,

HC(CO)NH-TOCSY, and HCCHTOCSY. Assignment of aromatic side chains was achieved based on 2D TOCSY and 2D NOESY spectra.

The specific interactions of $MLL1_{111–139}$, $MLL1_{123–160}$, $MLL_{140–160}$, $PogZ_{1117–1410}$, $IWS1_{352–548}$ or $JPO2_{1–130}$ with IBD were monitored by changes in the positions of signals of $^{15}N/^{13}C$-labeled IBD in 3D HNCO spectra. From these changes, the most affected amino acid residues were determined using the minimal backbone chemical shift ($^{15}N$, $^{13}C'$ and $^{1}H^{N}$) approach. In this system, position of each signal was described by set of three coordinates in the 3D spectra. A matrix of all possible distances of these signals in the free versus bound state was then calculated and only the minimal distances were taken into account. These minimal 3D chemical shift perturbations of individual IBD aa residues were divided by standard deviation of the minimal distances values and visualized. Amino acid residues within regions that were perturbed at least 1.5-fold of the standard deviation were then mapped on the structure of the IBD (PDB code 2B4J [37]). The assignments for the bound $PogZ_{1389–1404}$ peptide were obtained using $^{13}C/^{15}N$ filtered homonuclear TOCSY and NOESY experiments. $^{1}H$-$^{1}H$ distance constraints required to calculate the structure of the IBD–$PogZ_{1389–1404}$ complex were derived from 3D $^{15}N/^{1}H$ NOESY-HSQC, $^{13}C/^{1}H$ HSQC-NOESY and in 2D $^{13}C/^{15}N$ filtered/edited NOESY spectra, which were acquired using NOE mixing time of 120 ms. The family of converged structures for the IBD–$PogZ_{1389–1404}$ complex was initially calculated using Cyana 2.1 [132]. The combined automated NOE assignment and structure determination protocol was used to automatically assign the NOE cross-peaks identified in NOESY spectra and to produce preliminary structures. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the program TALOS+ [133] were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce the sets of converged structures with no significant restraint violations (distance and van der Waals violations < 0.2 Å), which were further refined in explicit solvent in YASARA (http://www.yasara.org).

In the case of IBD–$MLL1_{140–160}$ complex, concentrations of 0.1 mM $^{15}N$ labelled LEDGF/p75 IBD for binding site mapping or 0.5 mM equimolar solution of $^{13}C/^{15}N$-labelled LEDGF/p75 IBD and un-labelled $MLL_{140–160}$ were used for resonance assignments and structural determination of the complex in the same buffer as before.

Comprehensive backbone and side-chain resonance assignments for the $MLL_{140–160}$-bound $^{13}C/^{15}N$-labeled LEDGF/p75 IBD and for the bound unlabeled $MLL_{140–160}$ peptide were obtained analogously to the IBD–$PogZ_{1389–1404}$ complex. The essentially complete $^{15}N$, $^{13}C$,

and $^{1}$H resonance assignments allowed automated assignment of the NOEs identified in 3D $^{15}$N/$^{1}$H NOESY-HSQC, $^{13}$C/$^{1}$H HSQC-NOESY, and in 2D $^{13}$C/$^{15}$N-filtered/edited NOESY spectra using the CANDID protocol implemented in Cyana [132]. The combined automated NOE assignment and structure determination protocol [134] was used to automatically assign the NOE cross-peaks identified in 2D NOESY spectra, and to produce preliminary structures. This yielded unique assignments for 95.7% (2,338/2,444) of the NOE peaks observed, providing 1313 non-redundant $^{1}$H-$^{1}$H distance constraints, including 1199 intra-IBD, 78 intra-$MLL_{140–160}$ and 36 intermolecular constraints. In addition, backbone torsion angle constraints, generated from assigned chemical shifts using the program TALOS+ [133] were included in the calculations. Subsequently, five cycles of simulated annealing combined with redundant dihedral angle constraints were used to produce fifty-nine converged structures for the IBD–$MLL_{140–160}$ complex with no distance violations >0.2 Å from 100 random starting conformations using 1,451 NMR-derived structural constraints (14.8 constraints per residue). The 20 lowest energy conformers were further refined in explicit solvent in YASARA (http://www.yasara.org). Analysis and validation of the family of structures obtained was carried out using the programs Molmol, iCING and PyMol [135].

### 4.2.9 CD spectroscopy

Far UV CD spectroscopy experiments were carried out using a Jasco 815 spectrometer at 293 K. The buffer used for free IBD and $JPO2_{1-130}$ was 10 mM sodium phosphate, pH 8.0, containing100 mM NaCl, with final protein concentrations of 0.125 mg/ml for $JPO2_{1-130}$ and 0.11 mg/ml for IBD. The buffer used for the IBD-$JPO2_{1-130}$ complex was 5 mM Tris-HCl, pH 8.0, containing 100 mM NaCl, with a final protein concentration of 0.15 mg/ml. The CD spectra signal was monitored from 190 nm to 300 nm in a 0.1 cm quartz cell. After baseline correction, the spectra were converted to molar ellipticity θ (deg $cm^2$ $dmol^{-1}$) per residue, and the shares of α-helical fractions were calculated.

### 4.2.10 AlphaScreen

AlphaScreen measurements were performed in a total volume of 25 µl in 384-well Optiwell microtiter plates (PerkinElmer). The optimal protein concentrations for each experiment were determined by cross-titration to avoid binding curve perturbation while still

yielding a high signal-to-noise ratio. All components were diluted in assay buffer (25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM DTT, 0.1% [v/v] Tween-20 and 0.1% [w/v] bovine serum albumin). For binding curve determinations, 1 nM MBP-JPO2 and MBP-$PogZ_{1117-1410}$ wild-type and/or mutants were titrated against 200 nM Flag-LEDGF/p75. $MLL_{1-160}$-GST wild type or its mutants were titrated against 0.3 nM Flag-LEDGF/p75. In out-competition experiments, 20 nM Flag-LEDGF/p75 was pre-incubated with 200 nM IWS1 for 1 h at 4 °C. Subsequently, either HIV IN CCD, $JPO2_{1–130}$, $PogZ_{1117–1410}$ or $MLL1_{123–160}$ was added at the indicated concentrations. To assess the interaction effects of LEDGF/p75 mutants, wild type or mutant LEDGF/p75 was titrated against 20 nM His-HIV-1 IN or His-IWS1 or 1 nM MPB-$PogZ_{1117–1410}$ or MBP-JPO2.

Plates were incubated for 1 h at 4 °C. Subsequently, a mix of AlphaScreen donor and acceptor beads (PerkinElmer) was added (final concentration 20 μg/ml), bringing all proteins to the indicated final concentrations. After 1 h of incubation at 20 °C, the plate was analyzed in an EnVision Multi-label Reader in AlphaScreen mode (PerkinElmer). Each titration was performed at least in duplicate, and assays were independently repeated three times. Results were analyzed in Prism 5.0 (GraphPad software) after non-linear regression with the appropriate equations.

### 4.2.11 Peptide synthesis

Peptides used in this study ($MLL_{111–139}$, $MLL_{123–160}$ and $MLL_{140–160}$) were synthesized by solid phase synthesis and purified by HPLC in the Laboratory of Medicinal chemistry, IOCB, ASCR v. v. i., Prague, Czech Republic.

# 5. RESULTS

## 5.1 Preparation of proteins for LEDGF/p75 interaction studies

To characterize interactions between IBD and its binding partners, we prepared protein fragments by recombinant expression in *Escherichia coli*. We designed and cloned numerous expression vectors encoding different regions of these proteins. Expressed regions were chosen based on previously published results that suggested potential IBD interaction domains within the JPO2 and PogZ sequences [22, 56] and on sequence-based predictions. We optimized expression and purification protocols for to obtain the high yields and purity required for detailed characterization of IBD interactions (Figure 8).

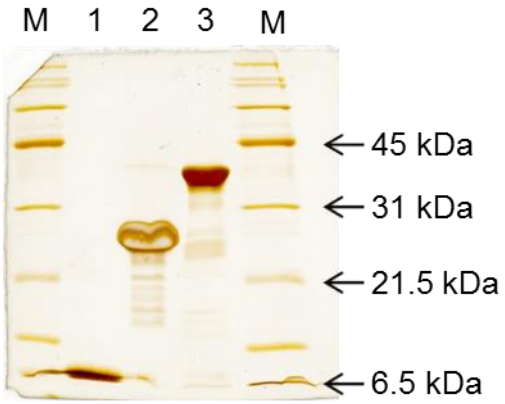


**Figure 8: SDS-PAGE analysis of purified protein fragments**

Silver-stained SDS-PAGE gel of purified IBD, $JPO2_{1–130}$ and $PogZ_{1117-1410}$. M = molecular weight standards; lane 1: IBD (theoretical molecular weight: 10.2 kDa); lane 2: $JPO2_{1-130}$ (theoretical molecular weight: 15 kDa); lane 3: $PogZ_{1117–1410}$ (theoretical molecular weight: 33.1 kDa)

The JPO2 construct encoding residues 1–130 ($JPO2_{1–130}$), which proved to be the most suitable for overexpression in *E. coli* included the specific interaction domain (SID) implicated in the interaction of JPO2 with IBD (amino acid residues 62–94 [23] or 77–98 [56]), as illustrated in Figure 9.

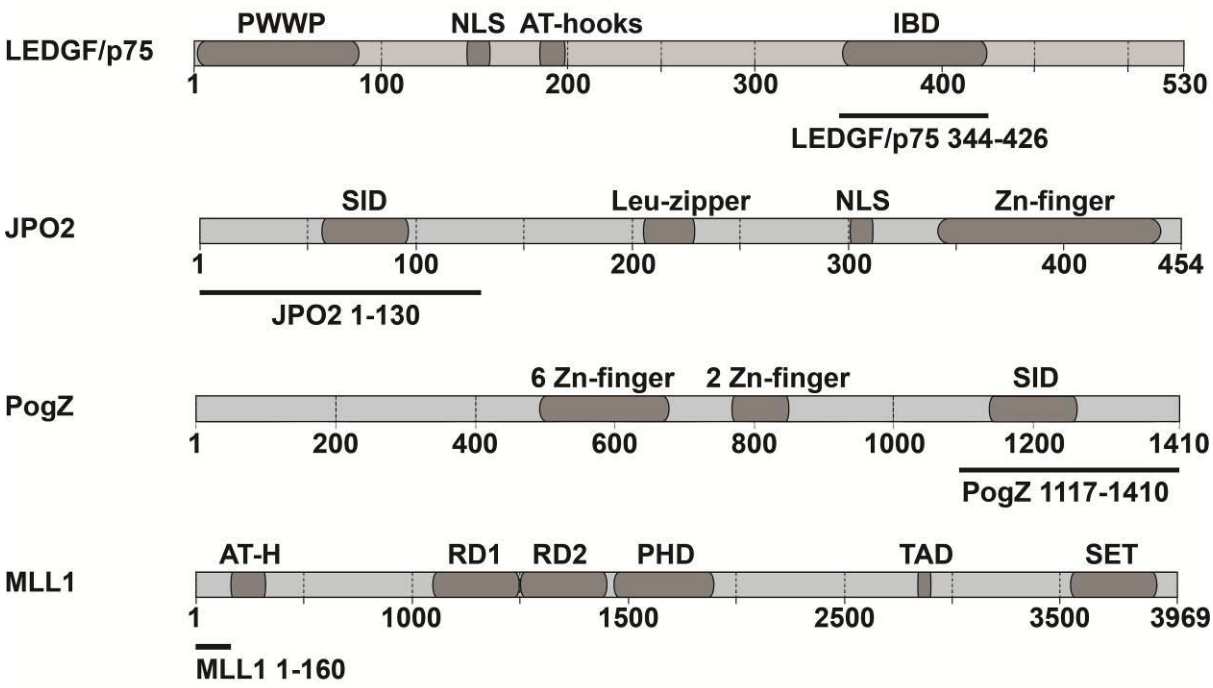


**Figure 9: Protein fragments used in this study**

Domain organization of LEDGF/p75 and its cellular binding partners is schematically depicted. Regions covered in protein constructs used in this work are represented by black bars below each protein sequence scheme. PWWP (Pro-Trp-Trp-Pro domain), IBD (integrase binding domain), TFIIS (transcription elongation factor S-II N-terminal domain), NLS (nuclear localization sequence) and SID (specific interaction domain) are labelled along with transcriptional repression domains (RD1 and RD2), PHD fingers (Plant HomeoDomain), transactivation domain (TAD) and histone methyltransferase SET domain of MLL1.

The identity of the purified $JPO_{1–130}$ protein was confirmed by LC-MS and MALDI-TOF, as there were irregularities in its gel migration behavior (Figure 10). Over 90% sequence coverage with more than 95% identification confidence was achieved in the peptide mass fingerprinting. The measured intact mass of $JPO2_{1–130}$ corresponded to the theoretical one. This confirmed the aberrant gel migration behavior of $JPO2_{1–130}$.

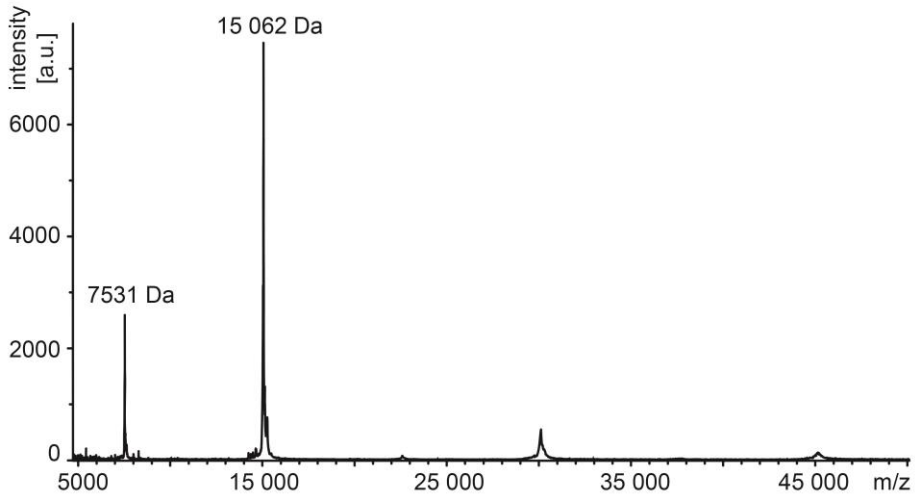


**Figure 10: Mass spectrometry analysis of $JPO2_{1–130}$**

The intact mass of $JPO2_{1–130}$ as determined by MALDI-TOF. The theoretical molecular mass of $JPO2_{1–130}$ (15.066 kDa) was verified.

The optimized PogZ construct spanning residues 1117–1410 ($PogZ_{1117–1410}$) included the predicted DDE domain, which was previously identified as a region responsible for interaction with IBD [22] (Figure 9). Domain boundaries were estimated based on secondary structure predictions and sequence homology between the DDE domain of PogZ and Mos1, a transposase with known crystal structure (PDB code 2F7T; [136]).

The optimized IBD construct used in this work contains LEDGF/p75 residues 345–426 (Figure 9). Compared to the IBD constructs used in previous structural studies [37, 38], our construct is truncated at the C-terminus. The C-terminal boundary was optimized based on secondary structures present in the crystal structure of IBD in complex with HIV IN (PDB code 2B4J; [37]). Inclusion of N425 and M426 at the C-terminus dramatically improved protein expression yields.

The MLL1 construct was designed based on previous publications to test the interaction with LEDGF/p75. Specifically, there was a discrepancy regarding the proposed menin-independent interaction between MLL1 and LEDGF between results published by Mereau *et al.* [95] and in previous studies. The MLL1 fragment implicated in the proposed menin-independent interaction spanned aa residues 1–160 (Figure 9). Therefore, we optimized expression and purification of this fragment ($MLL_{1–160}$), to be used in our detailed characterization of the IBD–MLL1 complex.

## 5.2 Characterization of MLL1 interaction with LEDGF/p75

### 5.2.1 Identification of the MLL1–LEDGF/p75 interaction interface

Previous studies on identification of the MLL1–LEDGF/p75 interaction interface reported ambiguous results. As mentioned before, Huang *et al.* claimed that neither menin nor MLL1 alone could stably associate with LEDGF/p75 [96]. Contradictorily to this assertion, Mereau *et al.* later described menin-independent interaction of MLL1 aa fragment 1–160 ($MLL_{1–160}$) [95]. This discrepancy was probably caused by using a crystallization-optimized peptide comprising 75 non-sequential aa residues of MLL1 region 6–153 in the older study (see Chapter 3.2.3 for details). Additionally, the MLL1 Phe-129 (F129) aa residue was previously identified as crucial to support the MLL1–LEDGF/p75 interaction in co-immunoprecipitation experiments by Yokoyama and Cleary [21].

When we introduced F129A mutation into recombinant $MLL_{1–160}$-GST it failed to significantly reduce the interaction signal in an *in vitro* AlphaScreen assay [137]. Moreover, substitution of the two critical aa residues of MLL1 which are in contact with LEDGF/p75 in the resolved crystal structure (F129A, F131A) did not abolish the interaction with LEDGF/p75 (Figure 11). On the other hand, mutations outside of the MLL1 region resolved in the crystal structure, specifically E145Q, E146Q and F148A, almost abolished the interaction with LEDGF/p75. These results supported the existence of an important additional interface between LEDGF/p75 and MLL1. As the last C-terminal residue of MLL1 resolved in the crystal structure is E135, this additional menin-independent interface was indicated in the region between aa residues 136 and 160. [137]

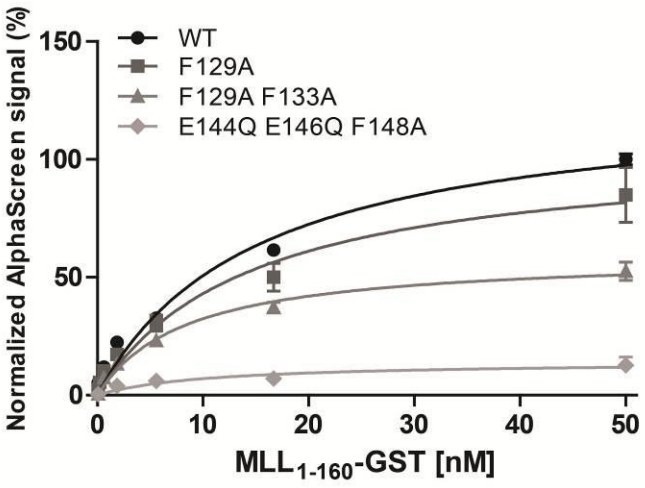


**Figure 11: AlphaScreen assay of interactions between LEDGF/p75 and $MLL_{1–160}$**

Interactions of wild-type $MLL_{1–160}$ (WT) and its mutant forms were tested by titration experiments monitored by AlphaScreen signal. Flag-tagged LEDGF/p75 (0.3 nM) and GST-tagged $MLL_{1–160}$ were used. Standard deviations were calculated from three independent experiments performed in duplicate. AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

In order to characterize the MLL–LEDGF/p75 interface in detail we produced three MLL1-derived peptides for nuclear magnetic resonance (NMR) spectroscopic analysis. These peptides comprised different stretches of MLL1 aa residues: 111 – 139 ($MLL_{111–139}$), 123 –160 ($MLL_{123–160}$) and 140 – 160 ($MLL_{140–160}$). We performed measurements of 2D $^{15}N/^{1}H$ heteronuclear single quantum correlation (HSQC) NMR spectra of $^{15}N$-labelled LEDGF/p75 IBD, alone or in presence of each MLL1-derived peptide in equimolar concentration (Figure 12).

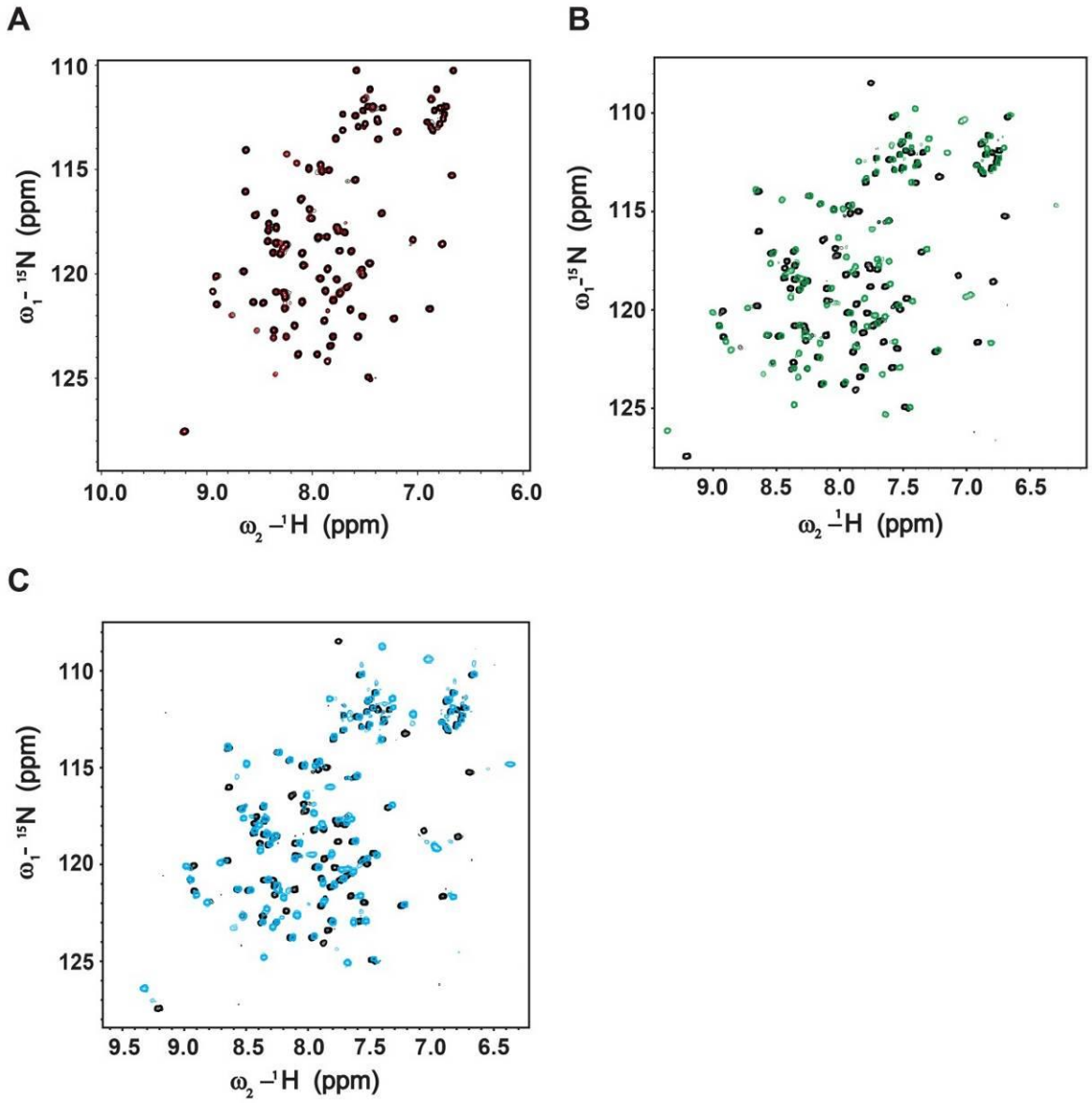


**Figure 12: Interactions of IBD with different MLL1-derived peptides monitored by NMR**

2D $^{15}$N/$^1$H heteronuclear single quantum correlation (HSQC) NMR spectra of $^{15}$N-labelled IBD alone (black crosspeaks) or in presence of unlabeled synthetic MLL1-derived peptide (colored crosspeaks). **(A)** $^{15}$N/$^1$H HSQC spectra of IBD alone (100 µM; black) and after addition of unlabeled $MLL_{111–139}$ (100 µM; red). **(B)** $^{15}$N/$^1$H HSQC spectra of IBD alone (100 µM; black) and after addition of unlabeled $MLL_{123–160}$ (100 µM; green). **(C)** $^{15}$N/$^1$H HSQC spectra of IBD alone (100 µM; black) and after addition of unlabeled $MLL_{140–160}$ (100 µM; blue)

These spectra were then compared. While $MLL_{111–139}$ did not induce any significant shifts in the positions of backbone or sidechain amide crosspeak signals of IBD, the opposite was observed for both $MLL_{123–160}$ and $MLL_{140–160}$. Major chemical shift perturbations of the IBD backbone amide signals induced by addition of either $MLL_{123–160}$ or $MLL_{140–160}$ are obvious in the overlaid spectra (Figure 12). Substantial shifts in the positions of backbone

amide signals indicated specific interactions of IBD with MLL1 region 140–160. To further investigate this interaction, we used isothermal titration calorimetry (ITC). In this experiment, IBD was titrated with injections of $MLL_{140–160}$ and the thermodynamic parameters of this interaction were monitored by calorimetric measurement (Figure 13). The results clearly confirmed the interaction between IBD and $MLL_{140–160}$ with stoichiometry close to 1:1 ($N$ = 1.15). Dissociation constant ($K_d$ = 86.3 ± 21.3 μM) of this interaction was calculated from three experiments. These results unambiguously confirmed an additional menin-independent interaction interface between LEDGF/p75 IBD and MLL1.

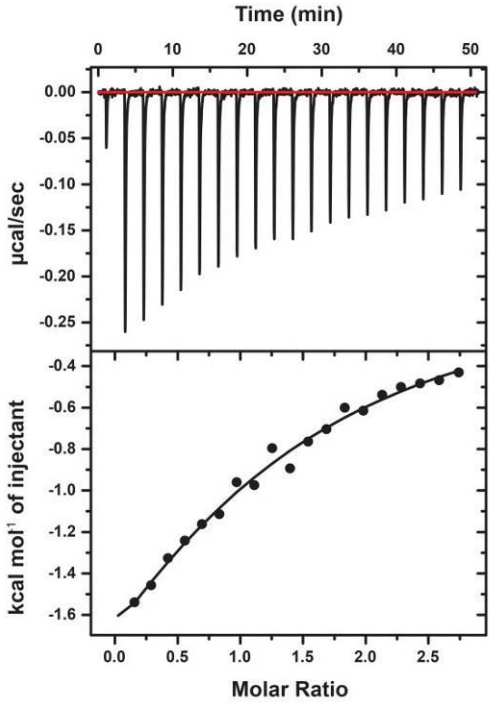


**Figure 13: Isothermal titration calorimetry (ITC) measurement of the IBD–$MLL_{140–160}$ interaction**

ITC data obtained for the IBD–$MLL_{140–160}$ interaction. Upper graph: experimental data; lower graph: fit (line) of the dilution-corrected integrated heats (full circles) from each injection of 650 μM $MLL_{140–160}$.

To map this additional MLL1 interaction interface on the IBD, we utilized NMR spectroscopy. We produced $^{15}N/^{13}C$ uniformly labelled IBD and performed measurements of a standard set of triple-resonance NMR experiments for backbone signals assignment. We successfully assigned 95.4% of $^{1}H$, $^{15}N$, $^{13}C'$ and $^{13}C^{\alpha/\beta}$ atoms of the IBD. Backbone amide signals ($^{15}N$ and $^{1}H$) were assigned for all residues except for four amino acids (E345, M348, Q410 and T417; Figure 14).

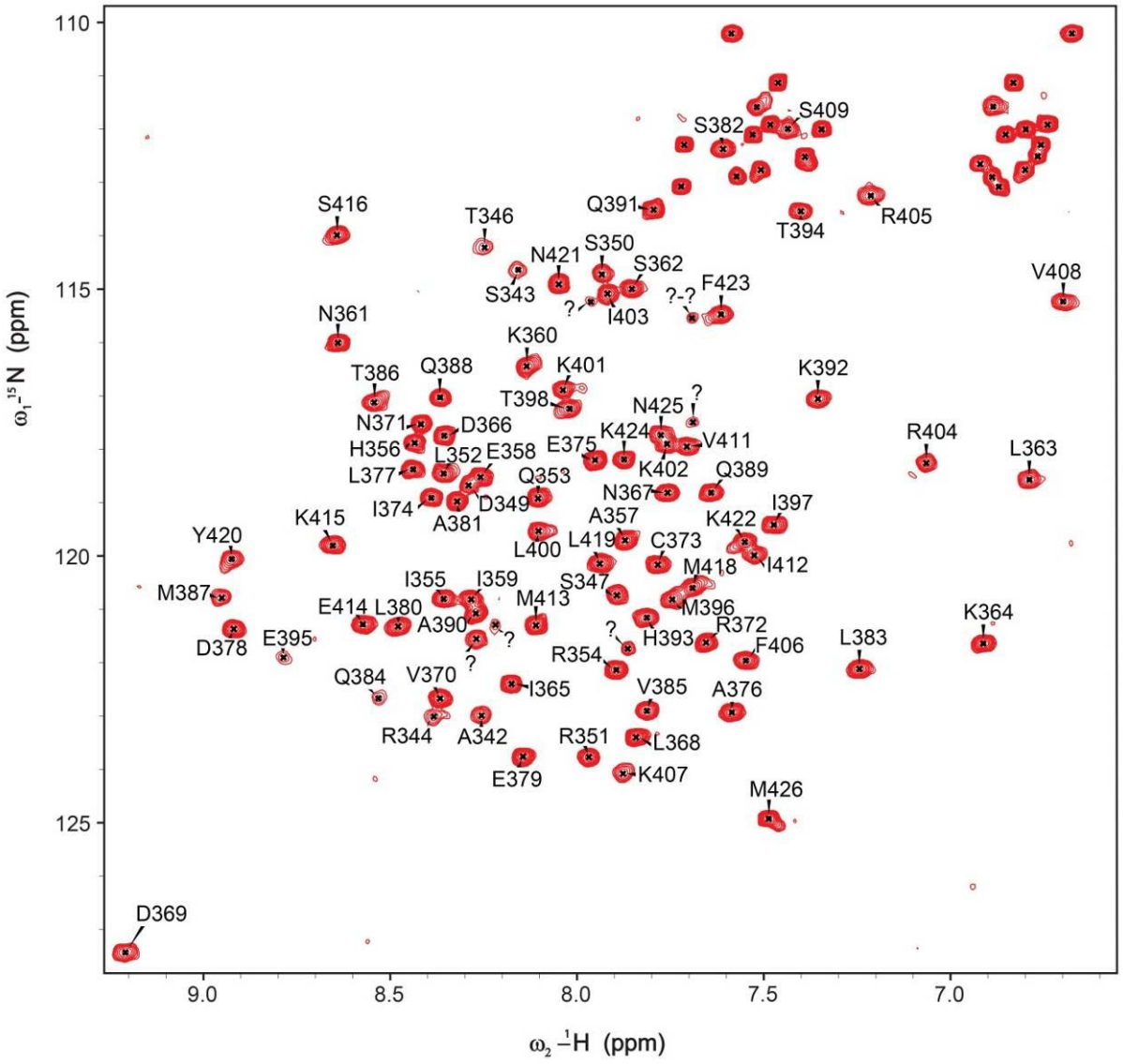


**Figure 14: Assigned $^{15}$N/$^{1}$H HSQC NMR spectra of IBD (aa 345–426)**

Result of backbone assignment represented on 2D $^{15}$N/$^{1}$H heteronuclear single quantum correlation HSQC NMR spectra of $^{15}$N-labelled LEDGF/p75 IBD (aa 345–426). Backbone amide signals ($^{15}$N and $^{1}$H) were assigned for all residues except for four aa residues (E345, M348, Q410 and T417). Every red crosspeak represents one backbone or side chain amide. Unassigned backbone amides are marked with question marks.

We monitored the structural changes of IBD upon formation of the $MLL_{140-160}$–IBD and $MLL_{123–160}$–IBD complexes by comparing the backbone signals in its free and bound state. Chemical shift perturbations of IBD backbone signals ($^{1}$H, $^{15}$N and $^{13}$C′) were analyzed in 3D HNCO spectra. In this system, position of each signal was described by set of three coordinates in the 3D spectra. A matrix of all possible distances of these signals in the free versus bound state was then calculated and only the minimal distances were taken into

account. These minimal 3D chemical shift perturbations of individual IBD aa residues were divided by standard deviation of the minimal distance values and visualized (Figure 15). There were two regions identified in the IBD sequence where the interaction with both $MLL_{123–160}$ and $MLL_{140–160}$ induced significant changes (over 1.5× standard deviation). These two regions (aa I359 to D369 and K402 to M413) highly overlapped with two interhelical loops bridging IBD alpha helices α1–α2 and α4–α5 respectively (Figure 15).

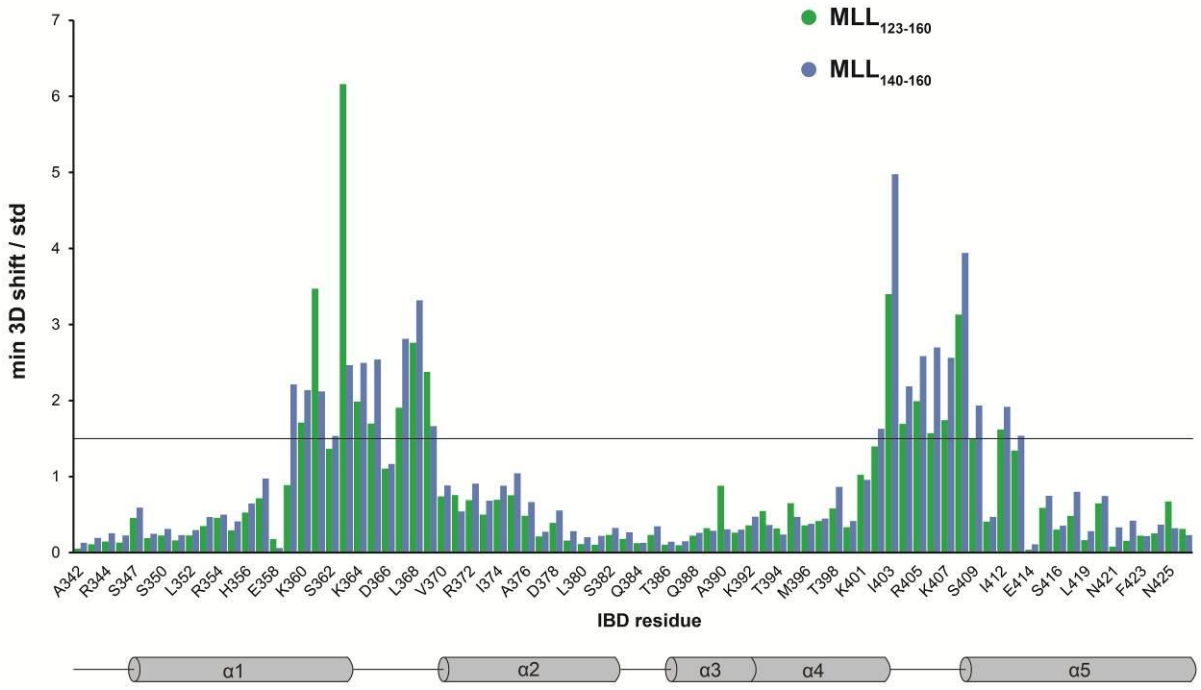


**Figure 15: Mapping of the menin-independent MLL1–LEDGF/p75 interaction interface**

Comparison of the minimal 3D chemical shift values of IBD backbone signals ($^{1}H$, $^{15}N$ and $^{13}C'$) in 3D HNCO spectra upon interaction with either $MLL_{140–160}$ or $MLL_{123–160}$. The minimal chemical shift values were divided by the corresponding standard deviation of all minimal 3D chemical shifts.
Data obtained for interaction with $MLL_{123–160}$ are depicted in green. Data obtained for interaction with $MLL_{140–160}$ are depicted in blue. Threshold of significance was determined as 1.5× standard deviation of corresponding dataset and is depicted as solid line. Schematic representation of IBD secondary structures is depicted below the x-axis accordingly to amino acid sequence.

These two regions are adjacent to each other and solvent-exposed on the surface of IBD in the structure of IBD (Figure 16) as solved by Cherepanov *et al.* [38]. Additionally, parts of IBD alpha helices 1 and 5 were also identified to be strongly affected by the interaction with both $MLL_{123–160}$ and $MLL_{140–160}$.

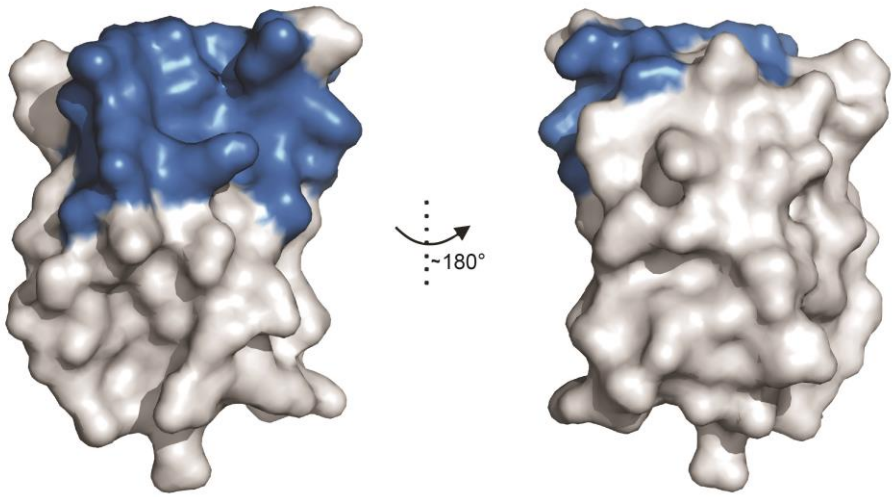


**Figure 16: Structural overview of IBD interaction interface for $MLL_{140–160}$ and $MLL_{123–160}$**

Specific changes induced in the IBD backbone resonances by interaction with $MLL_{140–160}$ or $MLL_{123–160}$ (blue, overlaying). Residues with substantial chemical shifts (over 1.5× standard deviation) are highlighted on the IBD structure surface representation.

Taken together, these results again suggested the menin-independent IBD-binding interface of MLL1 to be localized in the 140–160 region as both peptides induced analogous changes both in 2D HSQC NMR spectra and in minimal 3D chemical shift perturbations analysis.

### 5.2.2 Structural characterization of menin-independent IBD-binding interface

Although the minimal 3D chemical shift backbone perturbations were highly indicative of the actual MLL1 binding site on the IBD, further characterization was needed to reveal detailed involvement of individual MLL1 residues in the interaction. Therefore, we further employed NMR spectroscopy techniques and determined the solution structure of the IBD in complex with the $MLL_{140–160}$ peptide. We measured and analyzed a broad set of triple-resonance experiments to achieve backbone and side chain resonance assignments using $^{15}N/^{13}C$-labelled IBD in interaction with unlabeled $MLL_{140–160}$. Backbone and sidechain resonances of the unlabeled peptide were assigned using $^{13}C/^{15}N$-filtered homonuclear total correlation spectroscopy (TOCSY) and nuclear Overhauser effect spectroscopy (NOESY) experiments. This information was utilized in identification of $^{1}H–^{1}H$ distance constraints

using a series of 3D NOESY experiments. Overall, we were able to obtain 1,199 intra-IBD distance constraints, 78 intra-$MLL_{140–160}$ distance constraints, and 36 intermolecular constraints. These results were used for computational generation of structural models with minimal distance violations. These structures were further refined by calculations of lowest energy in solvent (for details see Chapter 4.2.8).

The resulting set of water-refined structures revealed important details of the menin-independent interaction between IBD and $MLL_{140–160}$. The IBD was found to be folded in a helix bundle almost identical to the previously published structures of IBD in free form or in complex with HIV integrase (PDB codes: 1Z9E [38] and 3U88 [37]). $MLL_{140–160}$ was unstructured and anchored between the two interhelical loops of IBD identified in minimal 3D chemical shift mapping (Figure 17A). The modest number of intermolecular distance constraints was mainly related to two aa residues of $MLL_{140–160}$: phenylalanines 148 (F148) and 151 (F151). These two residues are the major determinants of this interaction interface, anchoring the $MLL_{140–160}$ in two hydrophobic pockets on the surface of IBD (Figure 17B). The aromatic side-chain ring of F148 is in contact with IBD aa residues L363, L368, I403, F406, K407, and V408 (Figure 17C). Position of F151 in a deeper hydrophobic pocket was described by more intermolecular $^{1}H–^{1}H$ NOE contacts with IBD aa residues I359, K360, L363, T399, K402, and I403 (Figure 17C). The structure was deposited to the Protein Data Bank with the accession code 2MSR. NMR constraints and resonance assignments are available in the Biological Magnetic Resonance Data Bank (accession number 25130).

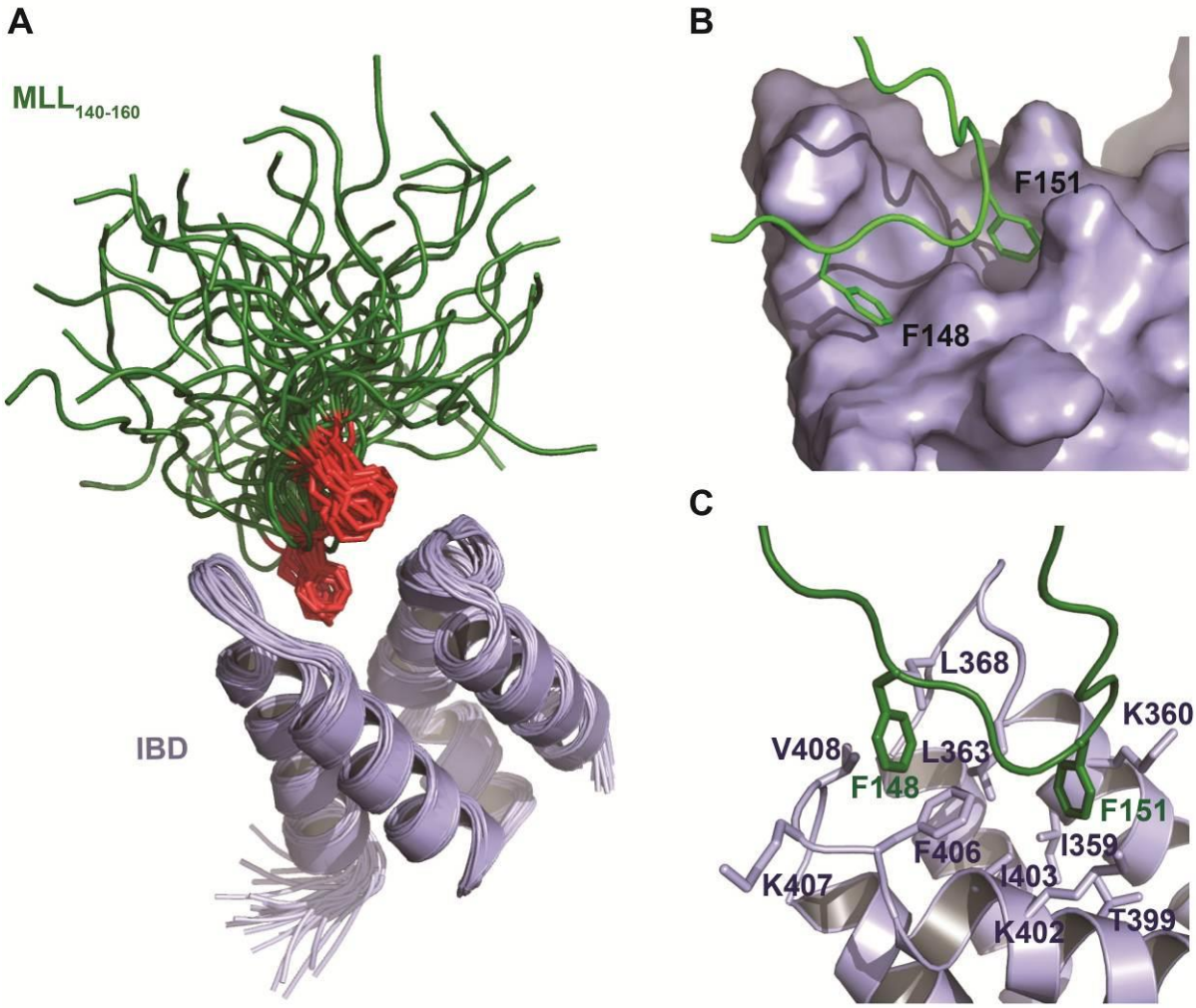


**Figure 17: Solution structure of IBD–MLL140–160 interaction**

**(A)** Converged set of structures of IBD (blue-white tint) interacting with $MLL_{140–160}$ (green) as determined by NMR (PBD code: 2MSR). The two key phenylalanines of $MLL_{140–160}$ (F148 and F151) are highlighted as red sticks. **(B)** F148 and F151 anchor the $MLL_{140–160}$ peptide in two hydrophobic pockets on the surface of IBD. One representative structure is depicted. **(C)** Detailed overview of the interaction interface. The F148 of $MLL_{140–160}$ interacts with IBD aa residues L363, L368, I403, F406, K407, and V408. The F151 of $MLL_{140–160}$ interacts with IBD aa residues I359, K360, L363, T399, K402, and I403. One representative structure is depicted.

As a verification of the structural data, we again employed mutation analysis monitored by AlphaScreen assay. We wanted to evaluate the importance of this newly identified interface in a MLL1 construct containing also the menin-dependent IBD-interaction site. Single mutant forms of $MLL_{1–160}$-GST were produced, containing either F148A or F151A. Additionally, a double-mutant form of $MLL_{1–160}$-GST was also prepared, containing both F148A and F151A. These mutant forms along with the wild-type one were tested for binding affinities to the full-length Flag–LEDGF/p75 (Figure 18). All three mutant forms of

$MLL_{1–160}$-GST exhibited dramatically reduced binding to full-length Flag–LEDGF/p75 (in comparison to wild-type $MLL_{1–160}$-GST), although the interaction was not fully abolished.

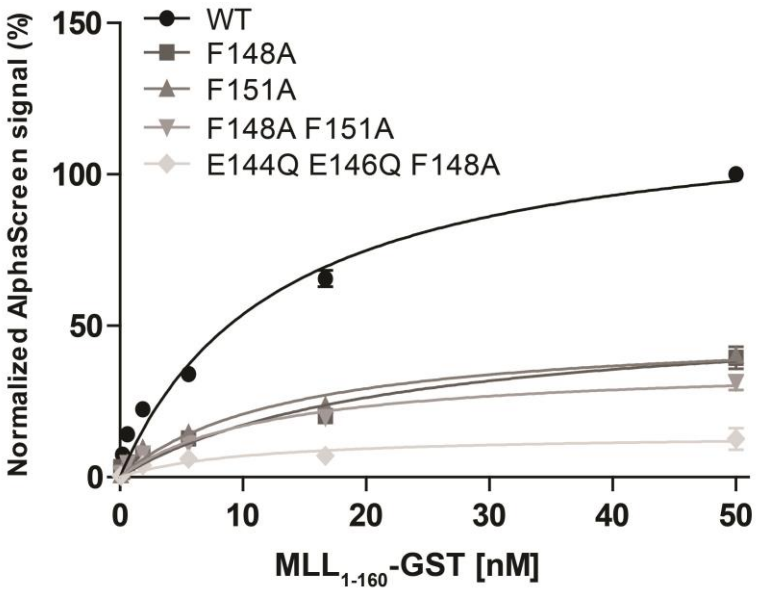


**Figure 18: AlphaScreen assay of interactions between LEDGF/p75 and $MLL_{1–160}$-GST**

Interactions of wild-type $MLL_{1–160}$-GST (WT) and its mutant forms were tested by titration experiments monitored by AlphaScreen signal. Flag-tagged LEDGF/p75 (0.3 nM) and GST-tagged $MLL_{1–160}$ were used. Standard deviations were calculated from three independent experiments performed in duplicate.
Adopted from Čermáková *et al.* [137]. AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

Only after mutation of three aa residues (E144Q/E146Q/F148A), was the interaction almost fully abolished. This showed the importance of glutamic acids in positions 144 and 146. Contacts of these residues with the IBD were not detected in our NMR experiments, probably due to deprotonated character of their carboxyl groups in solution. Carboxylate anions lack protons that would allow $^{1}H–^{1}H$ NOE contacts detection.

Combining all the available structural information together with results from mutation analyses, we created an expanded model of the IBD–MLL1 interaction (Figure 19). In this model, both interaction interfaces are present: the previously identified menin-dependent interface (for details see Chapter 3.2.3) along with the newly identified menin-independent one. In the menin-independent interface, four aa residues seem to play critical role in maintaining the interaction: Glutamic acids E144 and E146 and phenylalanines F148 and F151.

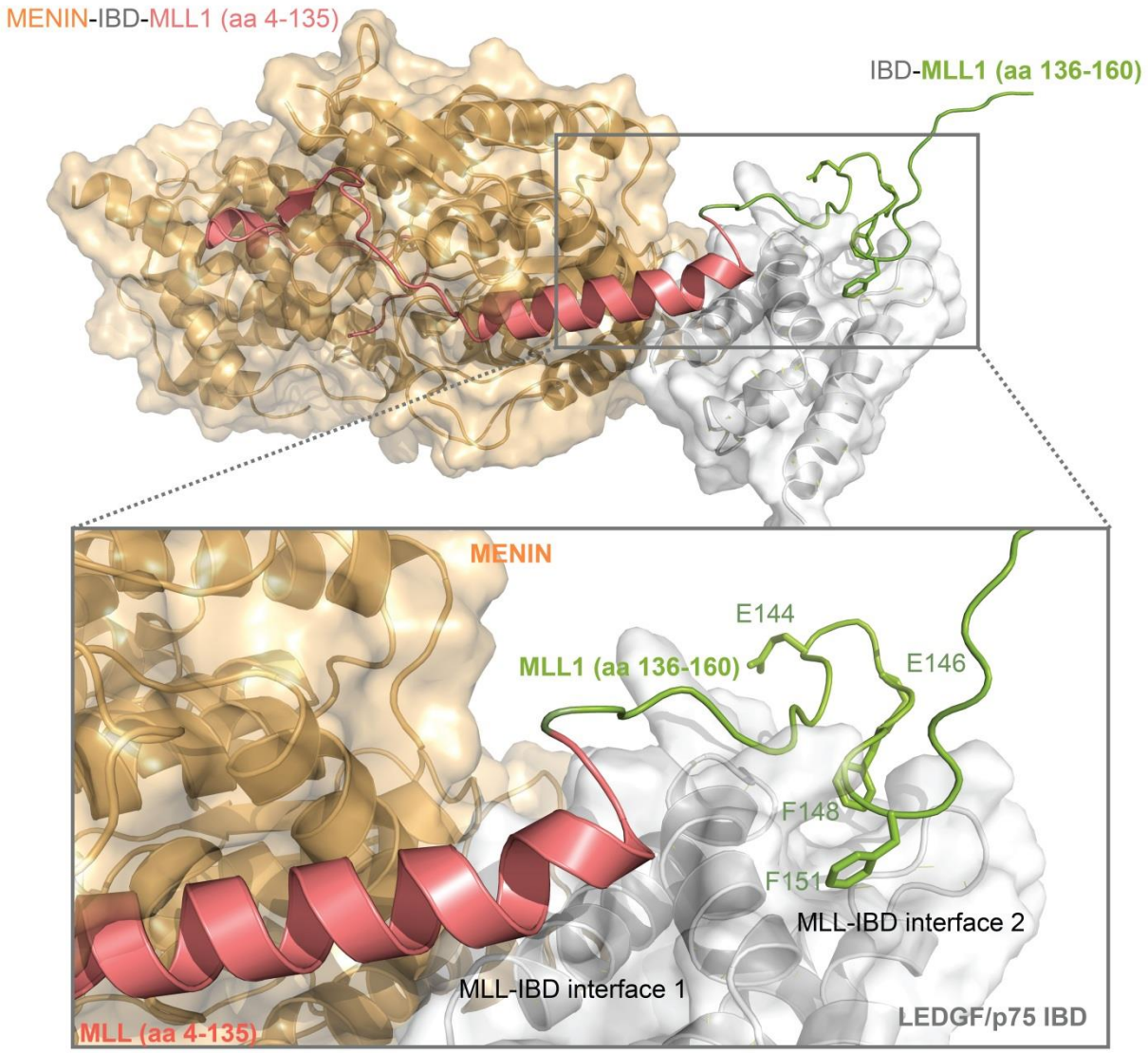


**Figure 19: Combination of structural data on available for IBD–MLL1–menin complex**

The IBD of LEDGF/p75 is depicted in gray and menin in orange. Portion of MLL1 resolved by Huang *et al.* [96] (PDB code: 3U88; marked as interface 1) is colored red and the additional portion published in Čermáková *et al.* [137] (PDB code: 2MSR; marked as interface 2) is colored green. Glutamic acids in positions 144 and 146 and phenylalanines in positions 148 and 151 are depicted as sticks.

### 5.2.3 The menin-independent interface is important in MLL1 fusion-driven leukemia

Knockdown of LEDGF/p75 expression was previously shown to inhibit clonogenic growth of MLL1 fusion-driven leukemic cells [21, 95]. After identifying the menin-independent interface for MLL1–LEDGF/p75 interaction, we wanted to test its importance in MLL1 transformation. Therefore, we utilized a previously developed cyclic peptide inhibitor that disrupted the interaction between LEDGF/p75 and HIV IN by binding to LEDGF/p75

IBD [124]. This cyclic peptide 65 (CP65) inhibitor was successfully tested for inhibiting the interaction between LEDGF/p75 and $MLL1_{1–160}$ while not interfering with the interaction between MLL1 and menin. When expressed in primary murine leukemic blasts isolated from mice, the CP65 reduced their colony-forming ability by 66%. Expression of mutant IBD-binding defective CP65 did not have any significant effect on colony formation. Moreover, the phenotype of reduced colony formation could be rescued by WT LEDGF/p75 but not by mutant MLL1-binding defective LEDGF/p75 (L368A and K407D). These results indicated that the menin-independent interface of MLL1–LEDGF/p75 interaction is indeed critical for LEDGF/p75-dependent MLL-mediated leukemic transformation and represents a potential therapeutic target. [137]

These results were obtained by our colleagues at KU Leuven and UKBB ZLF Basel and are in detail described in our joint publication Čermáková *et al.* [137].

## 5.3 Characterization of JPO2 and PogZ interactions with LEDGF/p75

Although LEDGF/p75 was intensively studied in contexts of HIV integration and MLL fusion-driven leukemia, little was known about its physiological interactions. These interactions might be the key to uncovering a dominant physiological role of LEDGF/p75 besides being a chromatin-tethering factor for other proteins. As the IBD of LEDGF/p75 became a potential therapeutic target, knowledge of its physiological interactions and their structural features gained in importance. Therefore, we chose to investigate the physiological interactions of LEDGF/p75 IBD with two know binding partners: JPO2 and PogZ.

### 5.3.1 Characterization of JPO2

After expression and purification of the $JPO_{1–130}$ fragment in high yields and purity, we characterized it by several biochemical and biophysical methods. During this characterization, several results hinted at disordered nature of $JPO2_{1–130}$.

First, we noticed irregularities in the gel migration behavior of this protein fragment. Therefore, the identity of purified $JPO2_{1–130}$ had to be verified by mass spectrometry analysis (Figure 10). Aberrant gel migration is considered to be an indication of protein disorder (http://www.disprot.org/view_detection.php). Moreover, circular dichroism spectra obtained

for $JPO2_{1–130}$ did not suggest any regular secondary structure (Figure 20B). The characteristic result for unstructured protein is apparent when compared with CD spectra of structured IBD (Figure 20A).

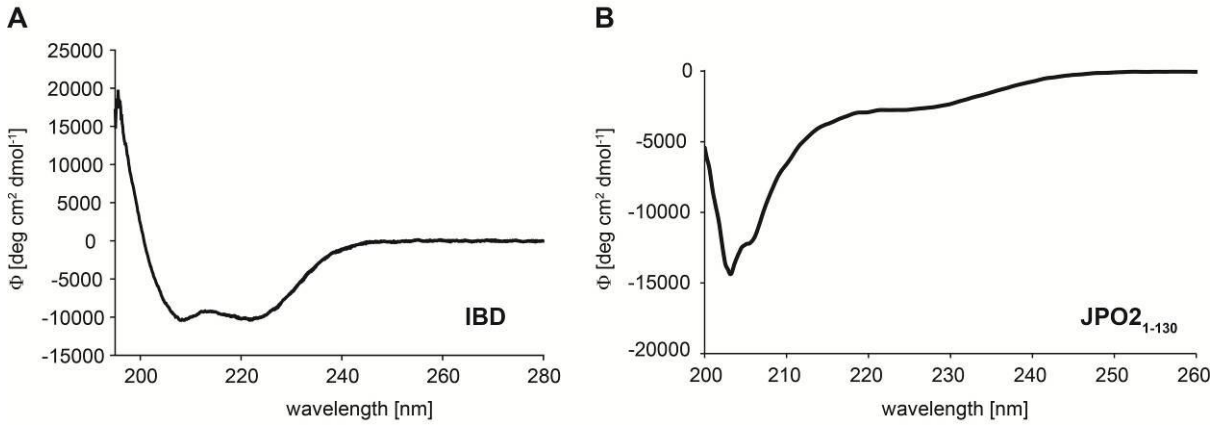


**Figure 20: Circular dichroism (CD) spectra of $JPO2_{1–130}$ and IBD**

**(A)** CD spectrum of IBD manifesting the typical double minimum of alpha-helical secondary structures (estimated as 81% of α-helices in the sequence of IBD). **(B)** CD spectrum of disordered $JPO2_{1–130}$.

Such result could be obtained due to denaturation by unsuitable buffer composition or pH. Therefore, we screened the thermal stability of $JPO2_{1–130}$ in various buffers using differential scanning fluorimetry (DSF). This screening revealed absence of a hydrophobic core in $JPO2_{1–130}$ in all 46 conditions tested (Figure 21). The fluorescence signal for $JPO2_{1–130}$ was highest at the beginning, with a slow decline over the course of the temperature gradient, showing no thermal unfolding. Absence of a hydrophobic core in the tested protein prevented determination of its melting temperature.

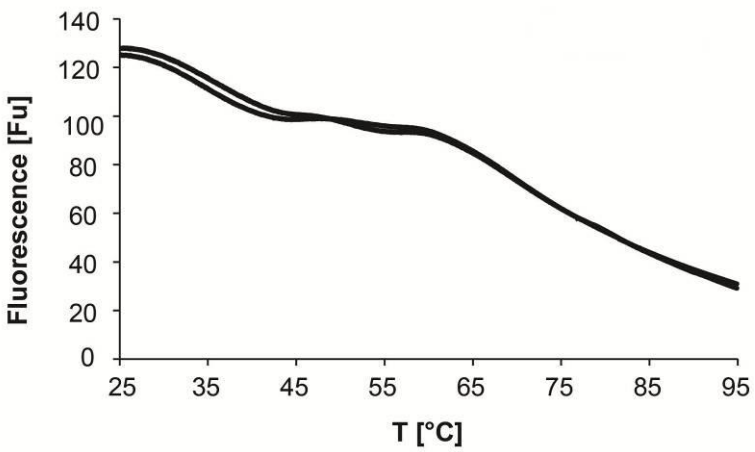


**Figure 21: Typical result of DSF experiment for JPO2$_{1–130}$**

Representative JPO2$_{1–130}$ unfolding curves determined by DSF. Two fluorescence intensity curves corresponding to experimental duplicates are shown.

Finally, the 1D $^{1}$H NMR spectra of JPO2$_{1–130}$ revealed poor dispersion of all groups of signals, in particular of amide hydrogens and side-chain methyl groups (at 7.5–8.5 ppm and 0.5–1.0 ppm), which are usually well-dispersed in a structured globular protein such as IBD (Figure 22).

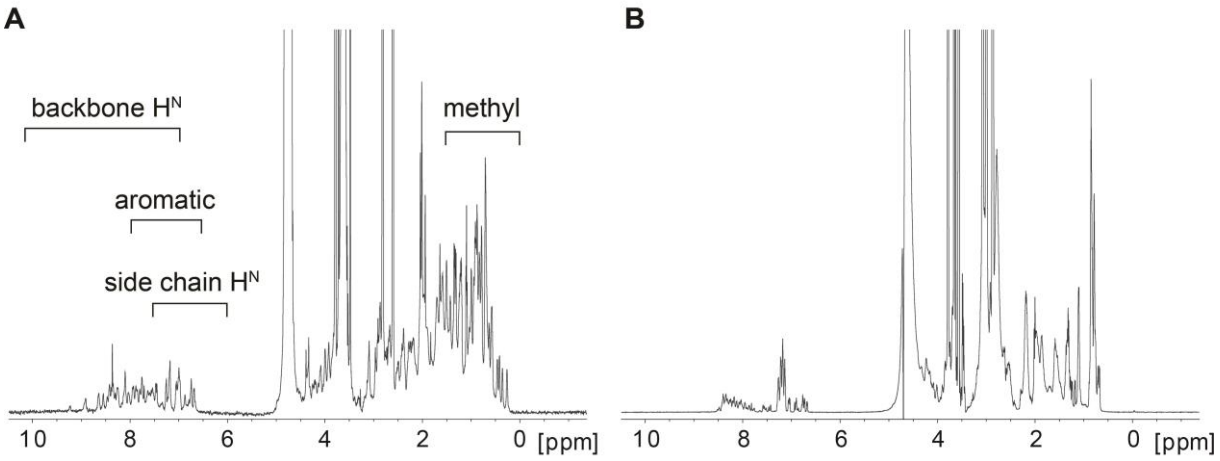


**Figure 22: 1D $^{1}$H NMR spectra obtained for IBD and JPO2$_{1–130}$**

**(A)** 1D $^{1}$H NMR spectra of free IBD. Characteristic chemical shift ranges for different protein $^{1}$H populations are marked. **(B)** 1D $^{1}$H NMR spectra of free JPO2$_{1–130}$.

The low dispersion of signals was also observed in 2D $^{15}$N/$^{1}$H HSQC NMR spectra of $^{15}$N-labelled JPO2$_{1–130}$ (Figure 23). Again, this represented a characteristic result for a protein lacking a stable tertiary structure caused by extensive signal-overlap and exchange

broadening of amide proton signals. The comparison of 2D $^{15}N/^{1}H$ HSQC NMR spectra of $^{15}N$-labelled $JPO2_{1–130}$ and IBD clearly shows the difference in signal dispersion (Figure 23).

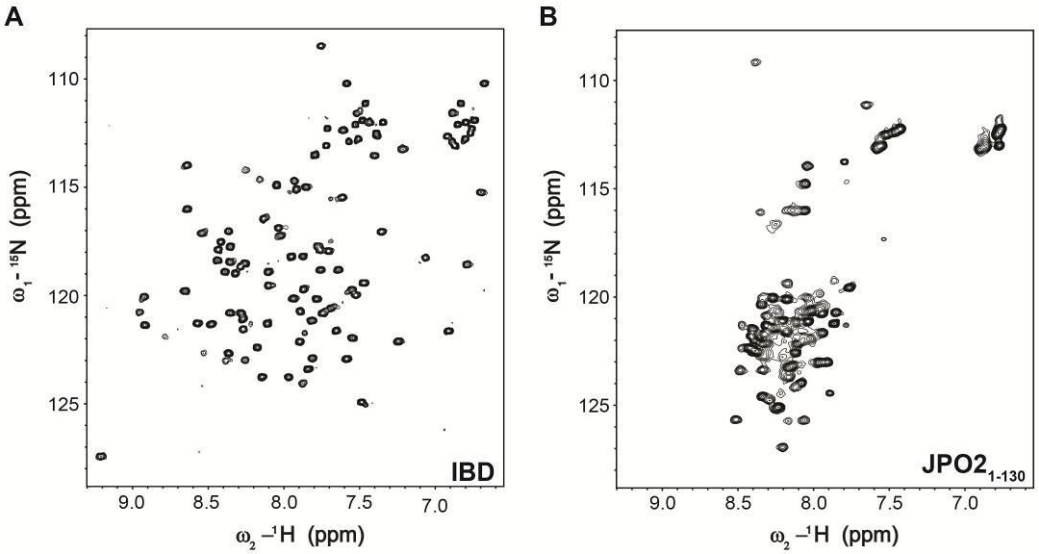


**Figure 23: Comparison of 2D $^{15}N/^{1}H$ HSQC NMR spectra of $JPO2_{1–130}$ and IBD**

**(A)** 2D $^{15}N/^{1}H$ HSQC NMR spectra of $^{15}N$-labelled IBD (88 aa residues) **(B)** 2D $^{15}N/^{1}H$ HSQC NMR spectra of $^{15}N$-labelled $JPO2_{1–130}$ (134 aa residues)
Crosspeak signals in these HSQC spectra represent backbone or sidechain amides.

The lack of experimental evidence for a regular structure of $JPO2_{1–130}$ raised a question whether this is due to the natural propensity of the JPO2 sequence or rather an effect of inappropriate experimental setup. Several lines of evidence suggested that the N-terminal part of JPO2 is intrinsically disordered. Firstly, our experimental data agreed with bioinformatic analysis of the $JPO2_{1–130}$ amino acid sequence. Protein disorder prediction tools FoldIndex [138] and PONDR-FIT [139] did not predict any significant regions of stable secondary structure within this protein fragment as well as the majority of following JPO2 sequence (Figure 24). Moreover, we have covered a broad range of buffer composition and pH (from 4.5 to 9.0) in our DSF experiments and none hinted at any change in the unstructured character of JPO2. To test the integrity of our purified JPO2 construct, we applied the only functional assay available: its ability to bind the IBD of LEDGF/p75.

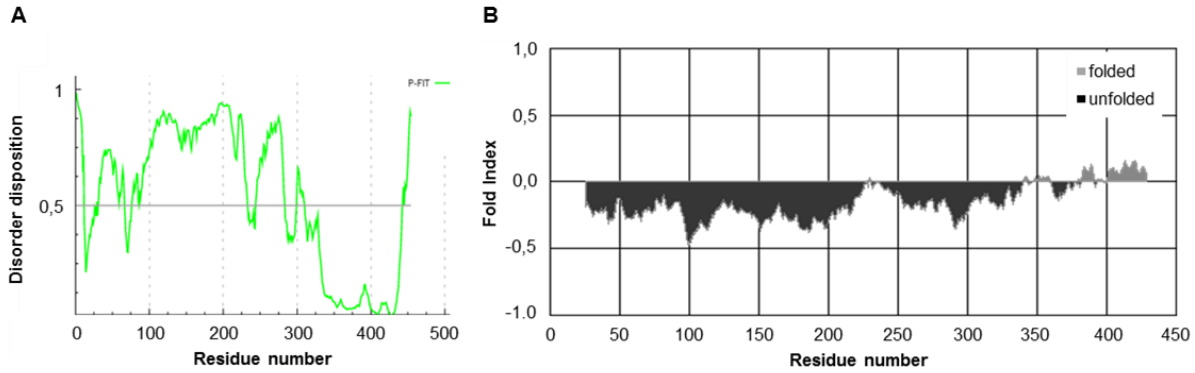


**Figure 24: Disorder predictions for JPO2**

**(A)** Protein disorder prediction tool PONDR-FIT [139] suggested only short possibly structured regions within the $JPO2_{1–130}$ sequence. **(B)** Protein disorder prediction tool FoldIndex© [138] suggest unstructured character of the whole $JPO2_{1–130}$ sequence except the very C-terminus.

### 5.3.2 $JPO2_{1–130}$ interacts with IBD via an intrinsically disordered region

We evaluated the ability of $JPO2_{1–130}$ to interact with the IBD by two approaches: isolation of reconstituted IBD–$JPO2_{1–130}$ complex using size exclusion chromatography (SEC) and thermodynamic characterization of the interaction by isothermal titration calorimetry (ITC). Despite being disordered, $JPO2_{1–130}$ readily interacted with IBD with sufficient affinity to allow the IBD– $JPO2_{1–130}$ complex to be isolated by SEC (Figure 25).

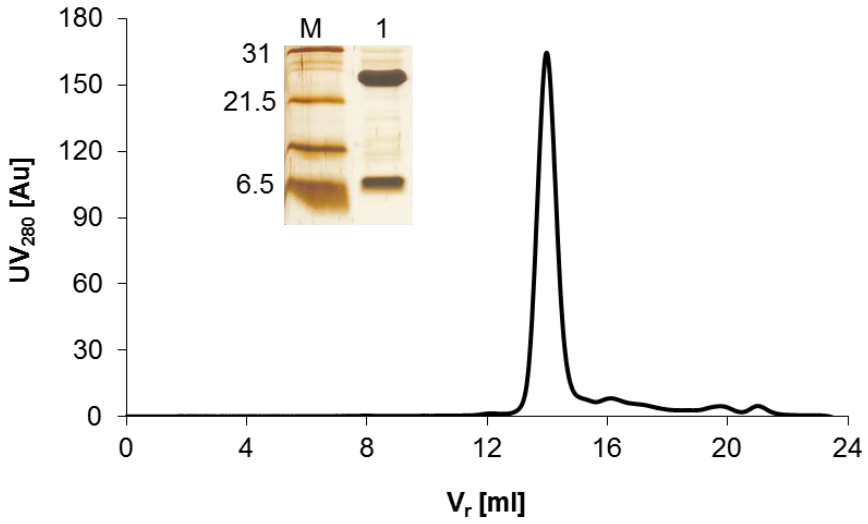


**Figure 25: Size exclusion chromatography (SEC) analysis of IBD–JPO2$_{1–130}$ complex.**

The chromatogram is overlaid with a silver stained SDS-PAGE gel of the isolated IBD–JPO2$_{1–130}$ complex (1). Protein molecular weight markers (M) are labelled in kDa. For individual proteins, see Figure 8.
The complex was reconstituted by overnight incubation of purified JPO2$_{1–130}$ with IBD in 50 mM Tris-HCl, pH 8, 150 mM NaCl and 0.05% β-mercaptoethanol at 4 °C.

ITC experiments revealed the affinity of this complex to be in low micromolar range ($K_d$ = 1.9 ± 0.2 µM). Interestingly, the measured stoichiometry suggested two molecules of IBD interacting with one molecule of JPO2$_{1–130}$ (average N = 1.9; Figure 26).

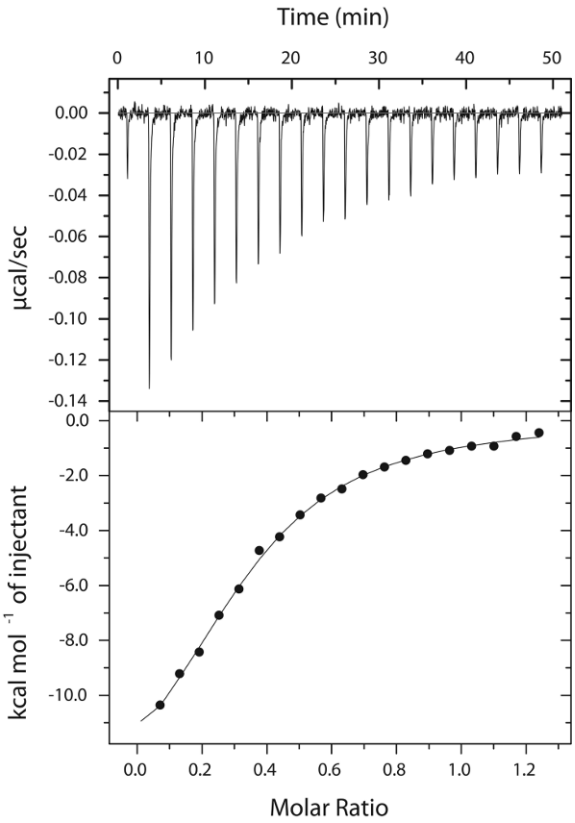


**Figure 26: ITC characterization of the IBD–JPO2$_{1–130}$ interaction.**

ITC data obtained for the IBD–JPO2$_{1–130}$ interaction. Upper graph: experimental data; lower graph: fit (line) to the integrated heats (full circles) from each injection of 60 µM JPO2$_{1–130}$ with correction of the dilution heat in the buffer.

The purified IBD–JPO2$_{1–130}$ complex was subjected to comprehensive buffer screening using DSF to determine optimal conditions for its stability. Interestingly, the fluorescence profile of the complex was equivalent to the one of a single protein with a hydrophobic core (Figure 27). The profile corresponded to a two-state model of thermal unfolding [131] without the high initial fluorescence signal of free JPO2$_{1–130}$. This allowed for an accurate determination of melting temperatures ($T_m$) in various buffers. In suboptimal conditions, the complex seemed to dissociate. This was easily distinguishable, as the strong initial fluorescence signal of free JPO2$_{1–130}$ in the mixture covered any potential melting curves of the complex or free IBD (Figure 27). Such behavior suggested the formation of a hydrophobic core upon interaction with IBD and a possible structural re-arrangement in JPO2$_{1–130}$ under optimal conditions.

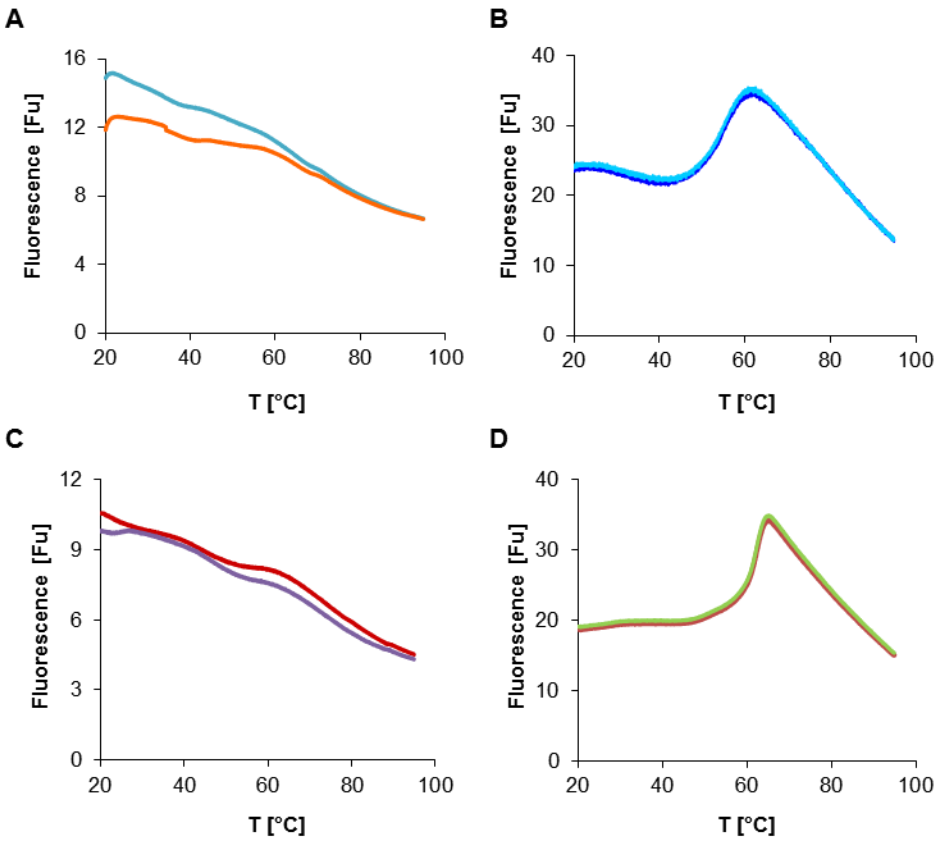


**Figure 27: Fluorescence levels during DSF experiments**

**(A)** For $JPO2_{1–130}$, the fluorescence signal was highest at the beginning with a slow decline over the course of the temperature gradient. **(B)** IBD–$JPO2_{1–130}$ complex in conditions optimal for complex formation (25 mM HEPES, pH 7, 100 mM NaCl, 0.05% β-mercaptoethanol) displayed a fluorescence profile typical for a single protein with a hydrophobic core. **(C)** IBD–$JPO2_{1–130}$ complex in conditions not optimal for complex formation showed the high initial fluorescence signal characteristic of free $JPO2_{1–130}$ with an indication of the overlaid melting curve of free IBD. **(D)** Free IBD in conditions optimal for complex formation (see legend panel B legend) shows a melting curve typical of a single protein with a hydrophobic core.
Two curves corresponding to experimental duplicates are shown.

Next, we investigated the possible structural rearrangements of JPO2 upon binding to IBD by comparison of the 1D $^{1}H$ NMR spectra acquired for the IBD– $JPO2_{1–130}$ complex and free proteins. We focused on the intensities of poorly dispersed signals of the side chain amide and aromatic proton region from unstructured $JPO2_{1–130}$. These signal intensities were reduced by approximately 25% in the complex relative to those of the free form, suggesting that part of $JPO2_{1–130}$ might adopt a defined conformation (Figure 28). However, a significant region within $JPO2_{1–130}$ preserved its disordered character, as the spectra of free and IBD-bound $JPO2_{1–130}$ largely overlap.

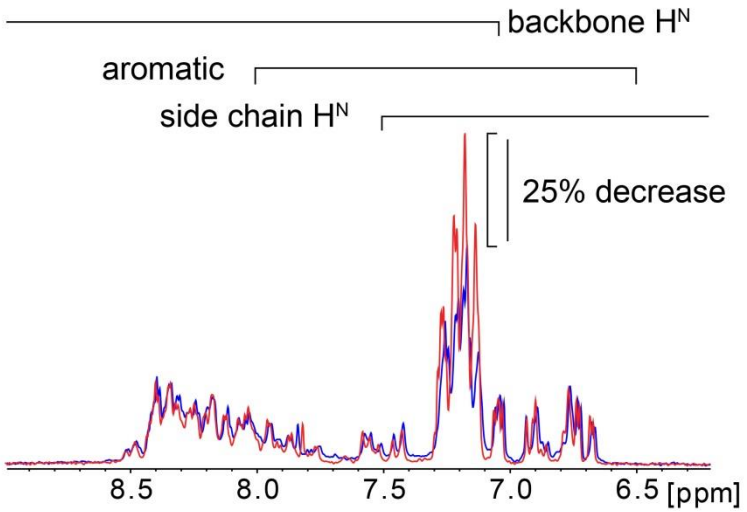


**Figure 28: 1D $^{1}$H NMR spectra obtained for IBD–JPO2$_{1–130}$ complex**

Comparison of the side chain amide and aromatic proton region of the $^{1}$H NMR spectra acquired for the free (red) and IBD-bound JPO2$_{1–130}$ (blue).

In addition to the changes in 1D $^{1}$H NMR spectra, we also observed significant differences between the CD spectra obtained for the IBD–JPO2$_{1–130}$ complex and the sum of the spectra obtained for IBD and JPO2$_{1–130}$ individually (Figure 29). This again indicated a structural change upon the interaction of JPO2$_{1–130}$ with IBD. Interestingly, the measured CD spectrum of the complex shows a significantly greater shift towards the unstructured character than expected from a computational projection. Taken together, these results indicated that JPO2$_{1–130}$ remains largely unstructured upon the interaction with IBD analogously to MLL$_{140–160}$.

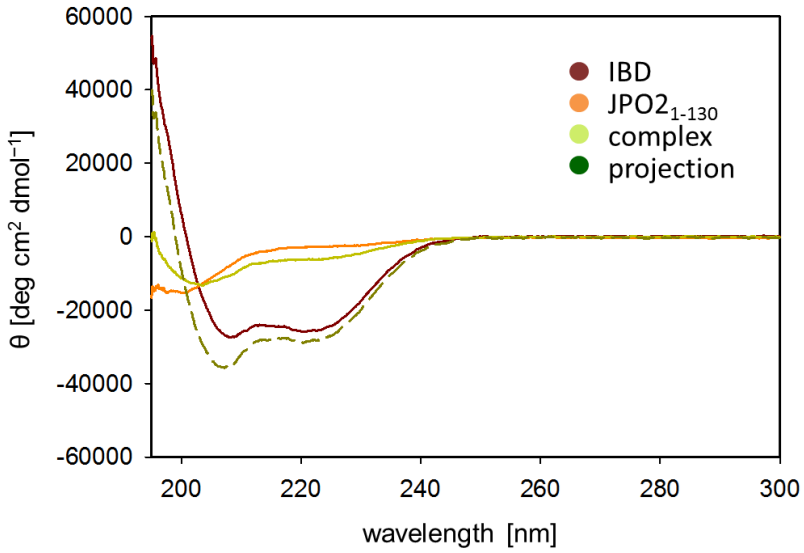


**Figure 29: Circular dichroism (CD) spectra of $JPO2_{1–130}$, IBD and IBD–$JPO2_{1–130}$ complex**

Results of far UV spectroscopy measurements of IBD (red), $JPO2_{1–130}$ (orange) and IBD–$JPO2_{1–130}$ complex (light green) in comparison with a computational projection of the actual IBD–$JPO2_{1–130}$ complex spectrum (dark green).

Altogether, several results strongly indicate that the structural flexibility of $JPO2_{1–130}$ is not an artefact but an inherent property of the protein fragment. The fact that the IBD–$JPO2_{1–130}$ complex can be isolated by SEC and exhibits sufficient stability with a low micromolar $K_d$, together with the evidence of structural changes within $JPO2_{1–130}$ upon IBD binding, supports successful production of functional JPO2 construct.

### 5.3.3 Characterization of PogZ

While $JPO2_{1–130}$ proved to be intrinsically disordered, biophysical characterization of $PogZ_{1117–1410}$ suggested that this protein adopts a well-defined conformation. As mentioned before, $PogZ_{1117–1410}$ was predicted to contain a DDE domain (aa residues 1117–1323) present also in lentiviral integrases. Structural information of homologous DDE domain-containing transposase Mos1 (PDB code: 2F7T [136]) along with secondary structure predictions indicated regular structure in the $PogZ_{1117–1410}$ fragment. Structured character of $PogZ_{1117–1410}$ was confirmed experimentally by DSF and CD spectra measurements. The fluorescence profile of $PogZ_{1117–1410}$ corresponded to the one of a folded protein with a hydrophobic core (Figure 30). This allowed for measurements of melting temperatures and screening for

conditions optimal for $PogZ_{1117–1410}$ stability (Tris 50 mM, pH 7.8, NaCl 150 mM, bme 0.05%).

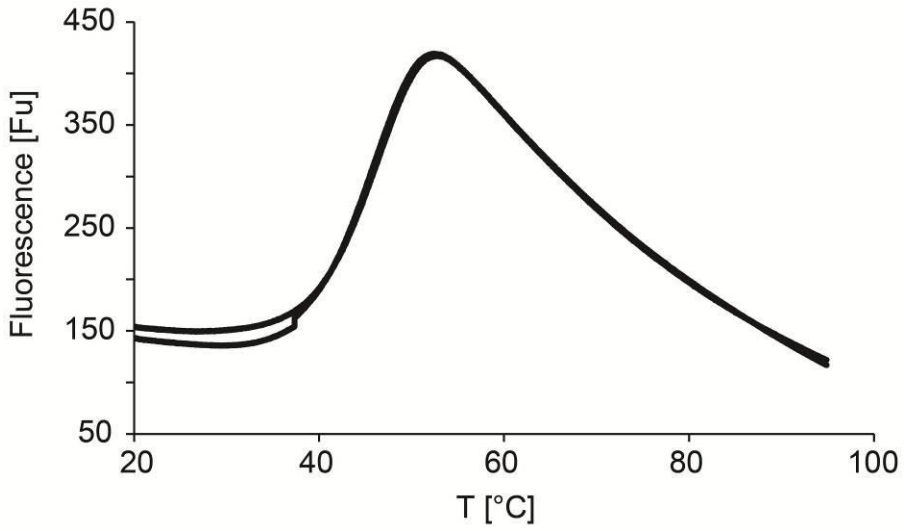


**Figure 30: Typical result of DSF experiment for $PogZ_{1117–1410}$**

Representative $PogZ_{1117–1410}$ unfolding curves determined by DSF. Two fluorescence intensity curves corresponding to experimental duplicates are shown.

The CD spectra measurements revealed a dominant representation of regular secondary structures in $PogZ_{1117–1410}$ with a significant proportion of β strands (Figure 31). This also corresponded to the predicted character of the DDE domain in $PogZ_{1117–1410}$.

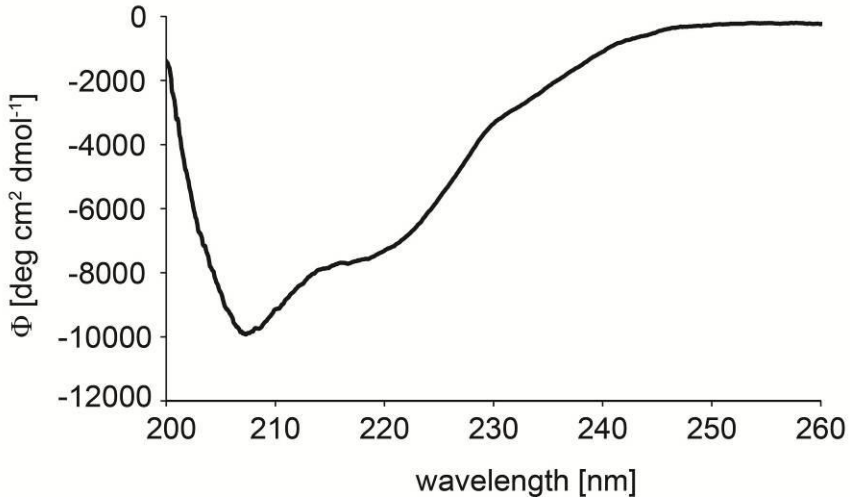


**Figure 31: CD spectrum of $PogZ_{1117–1410}$**

CD spectrum of $PogZ_{1117–1410}$ converted to molar ellipticity per residue manifesting the typical double minimum of alpha-helical secondary structures significantly affected by β-strands (estimated as 25% of α-helices and 40% β-sheets in the sequence of $PogZ_{1117–1410}$).

Although the CD spectra and DSF analyses showed the structured character of $PogZ_{1117–1410}$, a more detailed insight was gained by utilizing 2D $^{15}N/^{1}H$ HSQC NMR spectrum measurement. In this spectrum, a short unstructured region at the very C-terminus of $PogZ_{1117–1410}$ was revealed (after assignment recognized as aa residues 1368–1410; Figure 32). Crosspeaks of the backbone amide signals in this region are notably sharper and stronger than the amide signals from the structured part of $PogZ_{1117–1410}$.

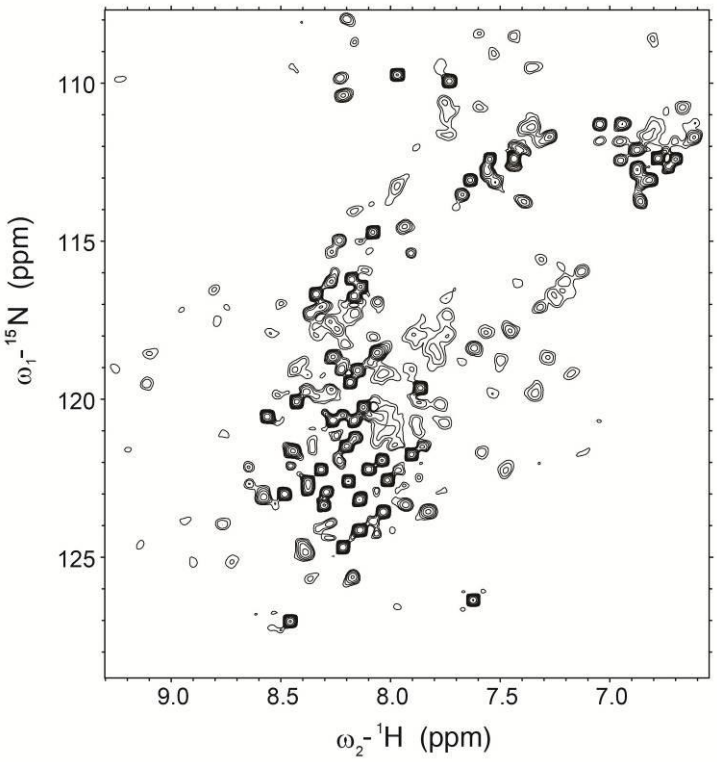


**Figure 32: 2D $^{15}N/^{1}H$ HSQC NMR spectrum of $PogZ_{1117–1410}$**

2D $^{15}N/^{1}H$ heteronuclear single quantum correlation (HSQC) NMR spectra of $^{15}N$-labelled $PogZ_{1117–1410}$. Backbone amide correlations of the unstructured region yielded notably sharper and stronger signals (crosspeaks).

Moreover, most amide signals of the 294 aa residues long fragment are not detectable in the spectrum. As the theoretical molecular weight of $PogZ_{1117–1410}$ is 33.11 kDa, this suggested a higher oligomeric assembly increasing the overall mass over the detection limit of this experiment. Signals from the flexible C-terminus would be less affected by the molecular weight increase, explaining their high intensity.

The presence of this flexible C-terminus was supported by results of limited trypsin proteolysis followed by N-terminal sequencing and mass-spectrometry analysis of the resulting products. This assay revealed a highly accessible cleavage site at the C-terminus following aa residue 1372 (Figure 33).

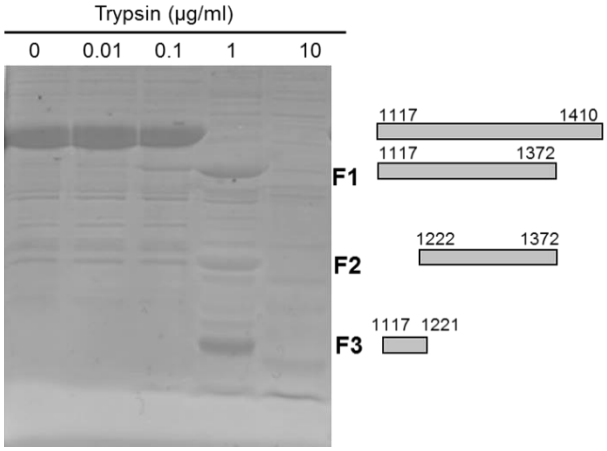


**Figure 33: Limited trypsin proteolysis of $PogZ_{1117–1410}$.**

Limited trypsin proteolysis of $PogZ_{1117–1410}$ generated stable fragments containing residues 1117–1372 (F1), 1222–1372 (F2) and 1117–1221 (F3) as determined by N-terminal protein sequencing and MALDI-TOF. Amino acid residue ranges of these fragments are depicted next to their respective migration bands on SDS-PAGE gel.
This analysis was performed by Kateřina Procházková at IOCB AS CR, Prague.

Moreover, a higher oligomeric assembly of $PogZ_{1117–1410}$ was indicated by analytical SEC. The retention volume of $PogZ_{1117–1410}$ ($V_r$ = 14.24 ml) corresponded to a molecular weight of 64.6 kDa, suggesting that $PogZ_{1117–1410}$ exists as a dimer in solution (Figure 34). In an analogous experiment, the theoretical molecular mass of IBD monomer closely correlated to molecular mass suggested by analytical SEC (11.5 kDa vs. 11.24 kDa; Figure 34).

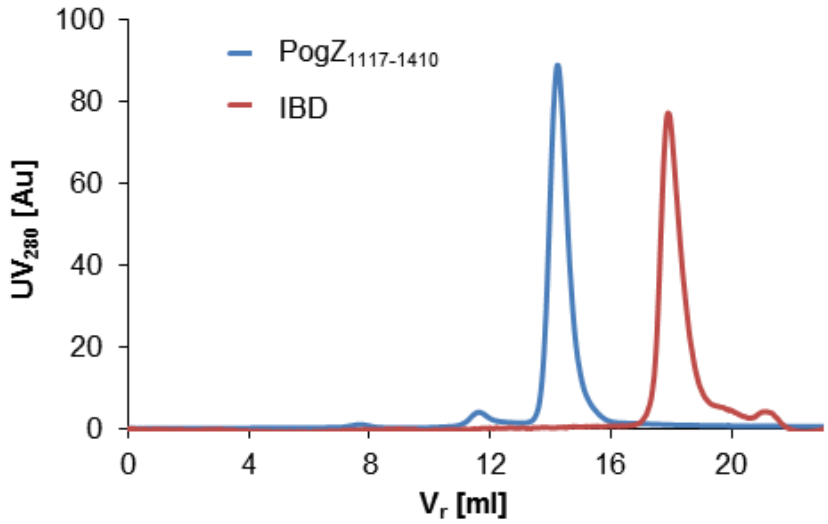


**Figure 34: Size exclusion chromatography (SEC) analysis of $PogZ_{1117–1410}$ and IBD.**

Analytical SEC of $PogZ_{1117–1410}$ (blue). The retention volume ($V_r$ = 14.24 ml) corresponded to an apparent molecular weight of 64.6 kDa according to calibration trend line equation. This suggests dimeric assembly of $PogZ_{1117–1410}$ (theoretical molecular weight: 33.11 kDa).
Analytical SEC of IBD (red). The retention volume ($V_r$ = 17.91 ml) corresponded to an apparent molecular weight of 11.5 kDa according to calibration trend line equation. This suggests monomeric assembly of IBD (theoretical molecular weight: 10.24 kDa).

### 5.3.4 Interaction of $PogZ_{1117–1410}$ with IBD

To assess the stoichiometry and affinity of the IBD–$PogZ_{1117–1410}$ complex, we used ITC (Figure 35). IBD readily interacted with $PogZ_{1117–1410}$ with a $K_d$ of 1.6 ± 0.2 µM. The ITC results also revealed the interaction stoichiometry to be close to 1:1 (N = 0.94).

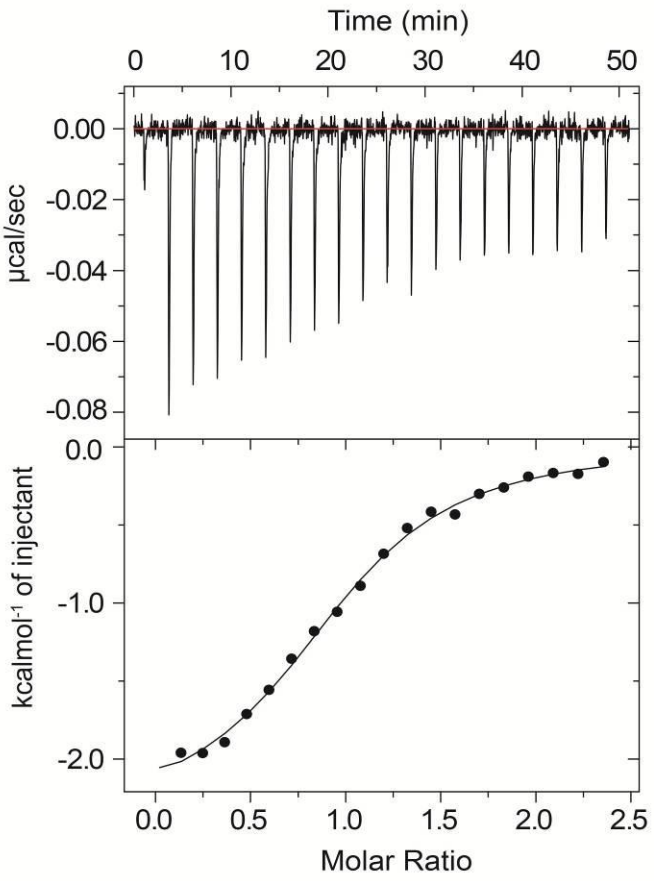


**Figure 35: ITC measurement of the IBD–$MLL_{140–160}$ interaction.**

ITC data obtained for the IBD–$PogZ_{1117–1410}$ interaction. Upper graph: experimental data; lower graph: fitting of the integrated heats (full circles) from each injection of 114 µM $PogZ_{1117–1410}$ with correction of the dilution heat in the buffer.

Taken together with the SEC analyses (Figure 34), we assumed that in the IBD–$PogZ_{1117–1410}$ complex a dimer of $PogZ_{1117–1410}$ interacts with two monomeric molecules of IBD.

### 5.3.5 Mapping of JPO2 and PogZ binding epitopes on IBD

To map the interaction interfaces of JPO2 and PogZ on IBD, we employed NMR spectroscopy approach. Analogously to the mapping of interaction interface of $MLL_{140–160}$ (Chapter 5.2.1), we employed the minimal chemical shift perturbations approach on IBD backbone signals in the free and bound state (Figure 36). Therefore, we could also utilize the assigned 2D and 3D spectra of $^{15}N/^{13}C$ double-labelled IBD (Figure 14), mapping the chemical shift perturbations to its individual aa residues. We could clearly observe significant shifts in the positions of backbone amide signals in the 2D $^{15}N/^{1}H$ HSQC spectra of both

complexes (Figure 36 A and B). To gain a more detailed insight into the interaction interfaces and reduce the chance of false negative minimal shifts, we again analyzed the $^{1}H$, $^{15}N$ and $^{13}C'$ chemical shift perturbations in 3D HNCO spectra. Only the minimal signal distances between free and bound state HNCO spectra we used in the analysis. These minimal distances were divided by standard deviation of the corresponding dataset. These specific changes affecting individual aa residues of IBD as induced by either $JPO2_{1–130}$ or $PogZ_{1117–1410}$ were plotted together in a graph (Figure 36C).

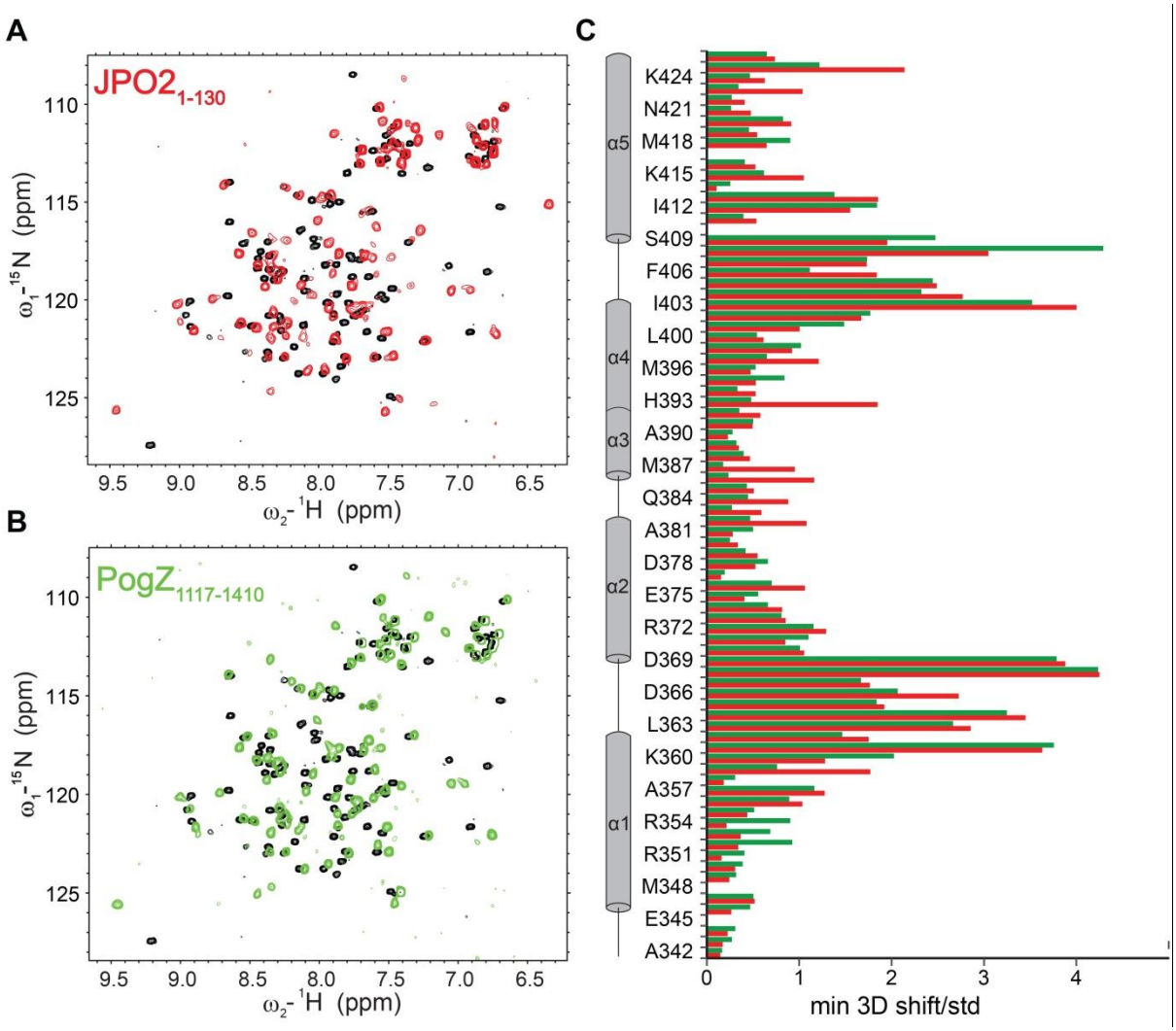


**Figure 36: Cellular partner-induced changes in the backbone of IBD**

**(A)** Comparison of the 2D $^{15}N/^{1}H$ HSQC spectra of the free (black) and cellular partner bound (red) IBD. The red spectrum in was obtained from the complex of $^{15}N$-labelled IBD with unlabeled $JPO2_{1–130}$. **(B)** Comparison of the 2D $^{15}N/^{1}H$ HSQC spectra of the free (black) and cellular partner bound (green) IBD. The green spectrum was obtained from the complex of $^{15}N$-labelled IBD with unlabeled $PogZ_{1117–1410}$. **(C)** The column graph summarizes the minimal chemical shift values divided by corresponding standard deviations for backbone resonances ($^{15}N$, $^{13}C'$ and $^{1}H^{N}$) calculated for both complexes (red for IBD–$JPO2_{1–130}$ complex; green for IBD–$PogZ_{1117–1410}$ complex). Alpha helices of IBD (α1-α5) are schematically illustrated.

Interestingly, interaction of IBD with either $JPO2_{1–130}$ or $PogZ_{1117–1410}$ seemed to induce very similar chemical shift perturbations when plotted together. Both $JPO2_{1–130}$ and $PogZ_{1117–1410}$ induced substantial changes (over 1.5× standard deviation) in two aa regions of IBD (residues I359–D369 and K402–M413). These regions cover two interhelical loops that are located close to each other on the surface of the IBD and were identified as interaction interface of $MLL_{140–160}$ in Chapter 5.2.1 (Figure 37).

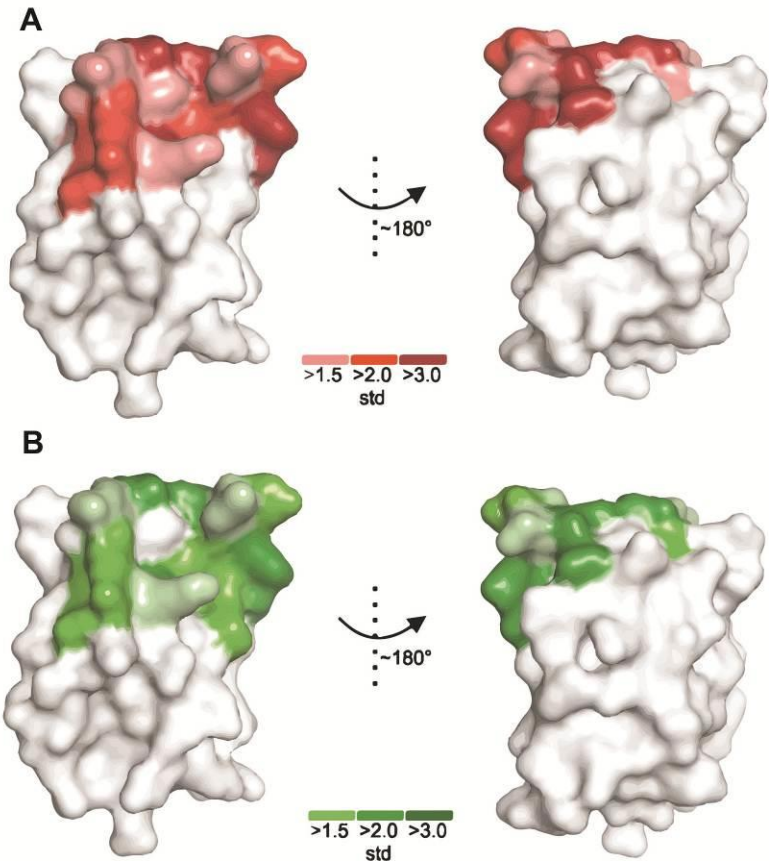


**Figure 37: Structural overview of IBD interaction interfaces for $JPO2_{1–130}$ or $PogZ_{1117–1410}$**

**(A)** Specific changes induced in the IBD backbone resonances by interaction with $JPO2_{1–130}$ (red). **(B)** Specific changes induced in the IBD backbone resonances by interaction with $PogZ_{1117–1410}$ (green).
Residues with substantial chemical shifts are highlighted on the IBD structure. The color saturation reflects the relative degree of chemical shift perturbations as a multiple of standard deviation (std.) of the corresponding data set.

Additionally, there was a subset of residues showing significantly higher perturbations within the two regions interacting with $JPO2_{1–130}$ (N361, L363, K364, D366, L368, D369, I403, R404 and V408). We presumed that these residues are directly implicated in the interaction with $JPO2_{1–130}$. In the case of $PogZ_{1117–1410}$, the IBD residues most affected by interaction were all present also in the case of $JPO2_{1–130}$ (N361, L363, K364, L368, D369, I403 and V408). This indicated that residues N361, L363, K364, L368, D369, I403 and V408 are implicated in direct interaction with both $JPO2_{1–130}$ and $PogZ_{1117–1410}$ and residues D366 and R404 might be in closer contact with $JPO2_{1–130}$ specifically.

### 5.3.6 Identification of IBD-interaction interfaces on JPO2 and PogZ

Interestingly, the IBD interaction interface for $MLL_{140–160}$, $JPO2_{1–130}$ and $PogZ_{1117–1410}$ was found to be highly overlapping. Next, we were interested in detailed characterization of IBD-interaction interfaces in $JPO2_{1–130}$ and $PogZ_{1117–1410.}$ As described in Chapters 3.2.1 and 3.3.2, minimal IBD-interacting constructs of JPO2 and PogZ were originally identified by yeast two-hybrid system experiments in previous studies [22, 23, 56]. These aa residue stretches were called specific interaction domains (SIDs). There were two SIDs identified in JPO2 by two independent studies (aa residues 62–94 or 77–98) [23, 56]. In PogZ, the SID was defined as aa residues 1139–1248 lying within the DDE domain.

As we strived for a detailed characterization of IBD-interacting epitopes, we employed the NMR-based mapping analogously to defining interaction interfaces on IBD. For this purpose, $^{15}$N-labelled $JPO2_{1–130}$ and $PogZ_{1117–1410}$ were produced. Analysis of the 2D $^{15}N/^{1}H$ HSQC NMR spectra obtained for free and IBD-bound $^{15}$N-labelled $JPO2_{1–130}$ revealed significant specific shifts in the backbone amide signal positions (Figure 38 A). Following this interaction confirmation, we produced $^{15}N/^{13}C$ uniformly labelled $JPO2_{1–130}$ and performed measurements of a set of triple-resonance NMR experiments for backbone signals assignment. We successfully assigned 94.4% of $^{1}$H, $^{15}$N, $^{13}C'$ and $^{13}C^{\alpha/\beta}$ backbone signals of $JPO2_{1–130}$. This allowed us to utilize the chemical shift perturbations approach in 3D HNCO spectra of free and IBD-bound $JPO2_{1–130}$. The specific changes in 3D signal positions were quantified as minimal chemical shift values in backbone HNCO resonances ($^{15}$N, $^{13}C'$ and $^{1}H^{N}$) divided by the corresponding standard deviation. In the highly overlapping spectra of unstructured $JPO2_{1–130}$, we were able to discern seven significantly (over 1.5× standard deviation) interaction-affected residues (V27, G28, F29, T66, G86, F87 and N93; Figure 38 B). Intriguingly, these residues were located in two separate regions along the $JPO2_{1–130}$ sequence. Moreover, three out of seven of these residues were located outside of the presumed SID.

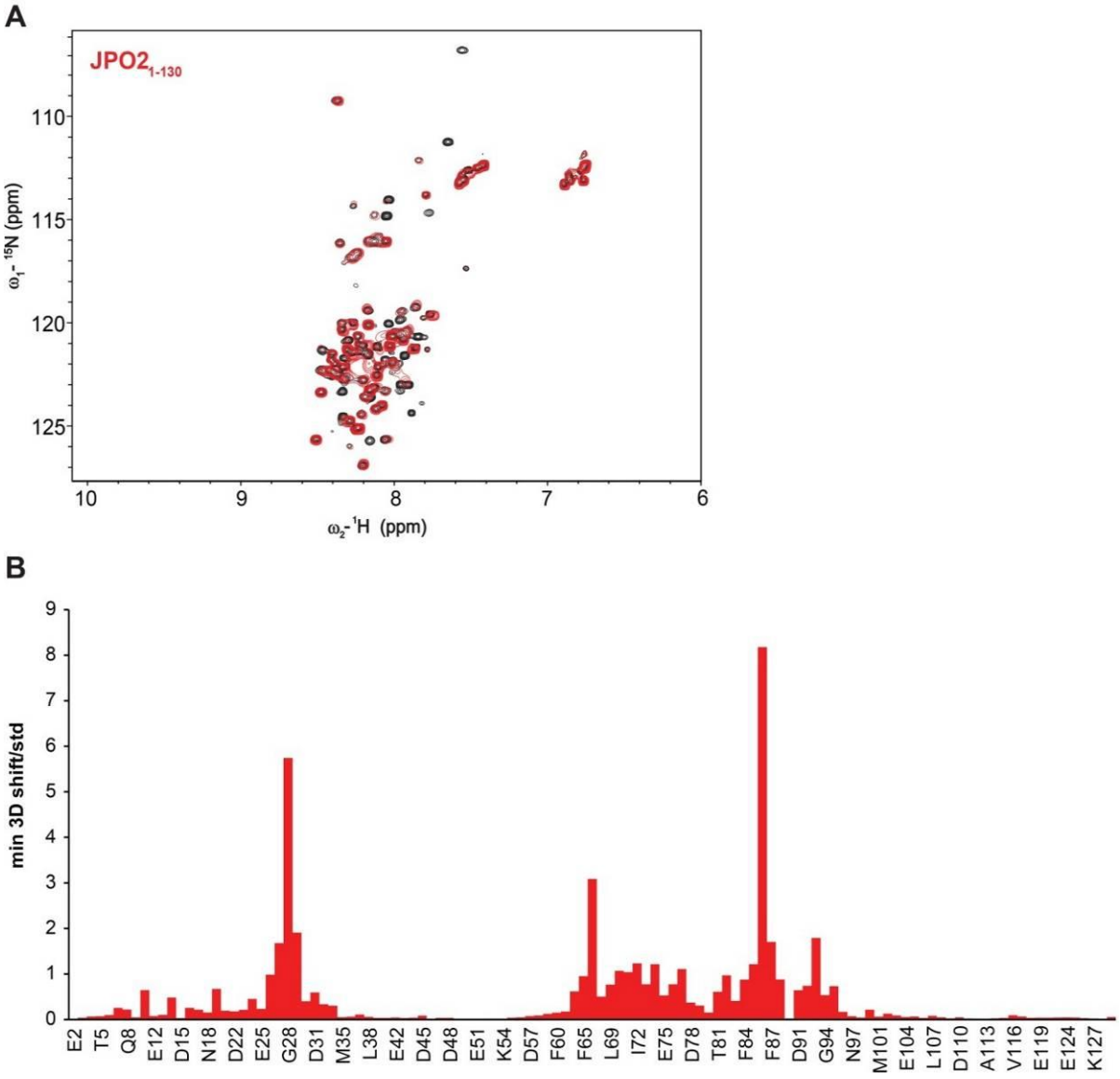


**Figure 38: IBD-induced changes in the backbone of JPO2$_{1–130}$**

**(A)** Comparison of the 2D $^{15}N/^{1}H$ HSQC spectra of the free (black) and IBD-bound (red) JPO2$_{1–130}$. The red spectrum in was obtained from the complex of $^{15}N$-labelled JPO2$_{1–130}$ with unlabeled IBD. **(B)** The column graph summarizes the minimal chemical shift values divided by corresponding standard deviation for backbone resonances ($^{15}N$, $^{13}C'$ and $^{1}H^{N}$) calculated for IBD–JPO2$_{1–130}$ complex.

In the case of PogZ$_{1117–1410}$, analysis of the 2D $^{15}N/^{1}H$ HSQC NMR spectra obtained in the free and IBD-bound also showed specific interaction-induced shifts in the backbone amide signal positions (Figure 39A). These shifts seemed to be limited to the fraction of noticeably more intensive signals, which we assumed to be representing a flexible part of PogZ$_{1117–1410}$. Analogously to JPO2$_{1–130}$, we proceeded by producing $^{15}N/^{13}C$ uniformly labelled PogZ$_{1117–1410}$ in order to get more detailed information. After measuring a set of triple-resonance NMR experiments for backbone signals assignment, we successfully assigned

practically all $^{1}H$, $^{15}N$, $^{13}C'$ and $^{13}C^{\alpha/\beta}$ signals of the flexible region. This region proved to be at very C-terminus of the PogZ protein (aa residues 1368–1410). Utilizing the assigned 3D HNCO spectrum, the minimal shift approach was employed. In this case, all the significantly (over 1.5× standard deviation) interaction-affected residues (F1397, Y1398, G1399, F1400 and A1403) were identified in a single stretch at the flexible C-terminal part of $PogZ_{1117–1410}$ (Figure 39B).

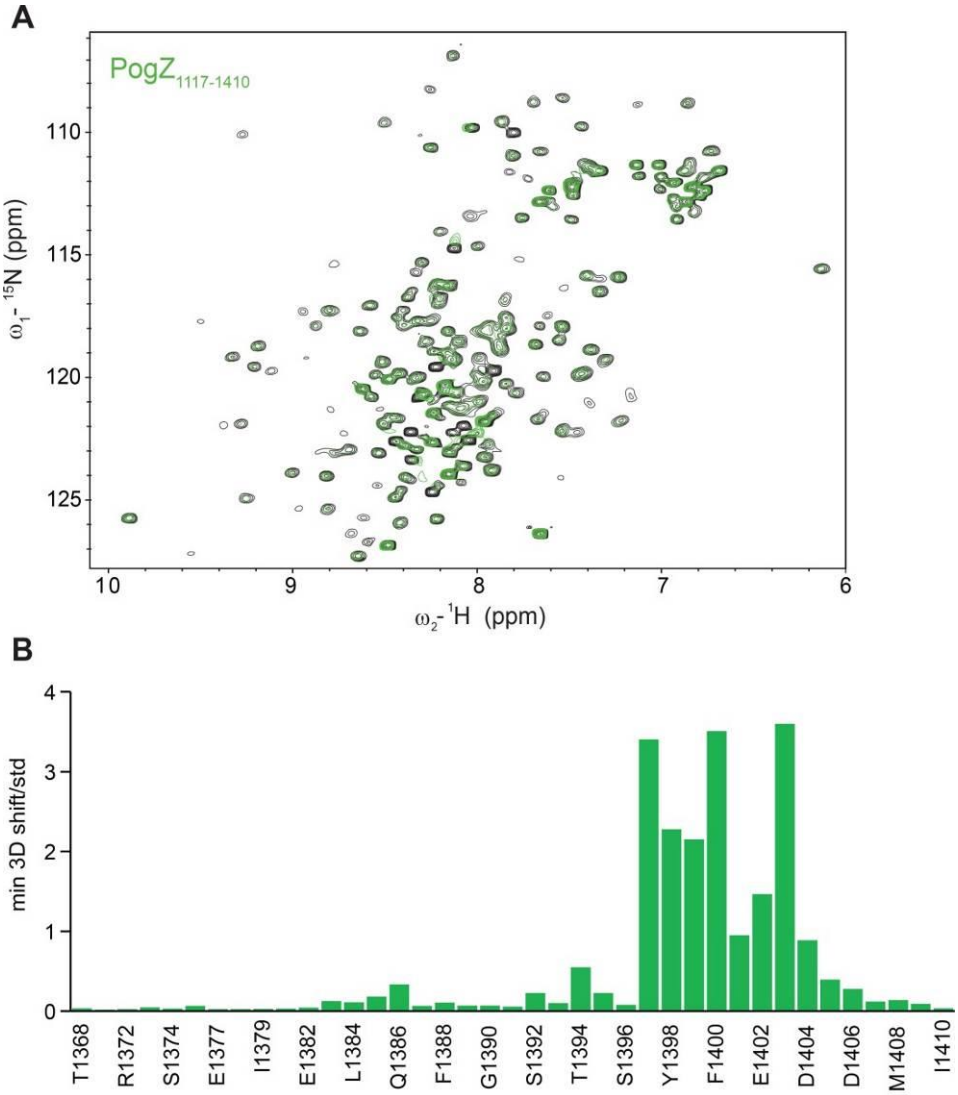


**Figure 39: IBD-induced changes in the backbone of $PogZ_{1117–1410}$**

**(A)** Comparison of the 2D $^{15}N/^{1}H$ HSQC spectra of the free (black) and IBD-bound (green) $PogZ_{1117–1410}$. The green spectrum in was obtained from the complex of $^{15}N$-labelled $PogZ_{1117–1410}$ with unlabeled IBD. **(B)** The column graph summarizes the minimal chemical shift values divided by corresponding standard deviation for backbone resonances ($^{15}N$, $^{13}C'$ and $^{1}H^{N}$) calculated for IBD– $PogZ_{1117–1410}$ complex.

## 5.4 Identification of a linear consensus IBD-binding motif

Taking the available information together including the previously described menin-independent IBD-binding interface of $MLL_{140–160}$, we revealed a striking level of identical residues shared among IBD-interaction interfaces. Indeed, by aligning the IBD-interaction regions of $MLL_{140–160}$, $JPO2_{1–130}$ and $PogZ_{1117–1410}$ a consensus motif was recognized. This consensus motif is intrinsically disordered in all these IBD interaction partners and contains sequence of eight aa residues: [E/D]-X-E-X-F-X-G-F (Figure 40).

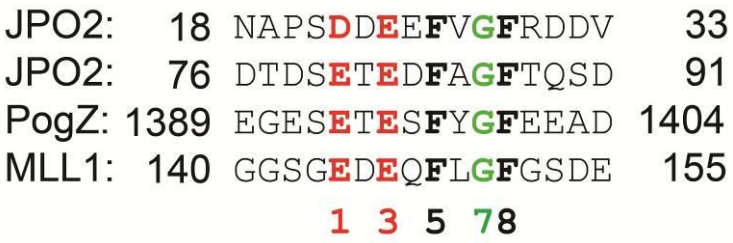


**Figure 40: Partial alignment of sequences containing the IBD-interaction motif (IBM)**

Alignment of IBM-containing sequences: $JPO2_{18–33}$, $JPO2_{76–91}$, $PogZ_{1389–1404}$ and $MLL1_{140–155}$. Amino acid residues of the consensus motif are numbered and colored. Acidic IBM residues are depicted in red, glycine residues in green and phenylalanines in bold black.

This consensus motif will be further referred to as the IBD-interaction motif (IBM). Intriguingly, there are two copies of IBM in the sequence of JPO2 (aa residues D22–F29 and E80–F87). Only one copy of IBM could be found in the respective sequences of both MLL1 (E144–F151) and PogZ (E1393–F1400). The two phenylalanines identified to be critical for $MLL_{140–160}$ interaction with IBD are completely conserved in all IBMs together with the glycine at position seven. This glycine is most probably necessary to increase degree of freedom of the second phenylalanine at position eight, allowing it to occupy the hydrophobic pocket on the surface of IBD as revealed in the previously presented structure of the IBD–$MLL_{140–160}$ complex. In the case of the first IBM in the sequence of JPO2, there is a substitution of glutamic to aspartic acid in position one. Nonetheless, these amino acid residues are highly similar in their physicochemical properties.

### 5.4.1 Validation and structural characterization of the IBM

In order to validate the newly identified IBM, we performed a series of mutation analyses of PogZ monitored by AlphaScreen. Analogously to the case of the IBD–$MLL_{140–160}$

complex, we introduced single and double point mutations (F1397A and F1397A/F1400A) into maltose binding protein (MBP)-tagged $PogZ_{1117–1410}$. (Figure 41, Table 1).

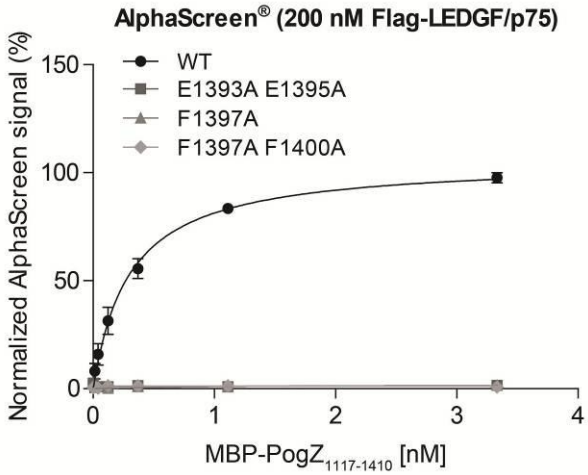


**Figure 41: AlphaScreen assay of interactions between LEDGF/p75 and $PogZ_{1117–1410}$**

Interaction of wild-type $PogZ_{1117–1410}$ (WT) and its mutant forms were tested by titration experiments monitored by AlphaScreen signal. Flag-tagged LEDGF/p75 (0.3 nM) and MBP-tagged $PogZ_{1117–1410}$ were used. Standard deviations were calculated from three independent experiments performed in duplicate. AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

Strikingly, mutation of a single phenylalanine of the IBM (F1397) completely abrogated the interaction of PogZ with LEDGF/p75. Actually, all mutations introduced in the IBM of $PogZ_{1117–1410}$ completely abrogated the interaction with LEDGF/p75 including the E1393A/E1395A double mutant. These conserved glutamic acid residues at positions one and three in the IBM were shown to be involved in the interaction with IBD also in the case of JPO2. As there are two IBM copies in the sequence of JPO2, the mutation analysis was more complicated (for details see Table 1). Both copies of the IBM were proved to be engaged in the interaction of JPO2 with LEDGF/p75 *in vitro*. In each IBM copy, the two phenylalanines at positions five and eight were critical for the interaction.

**Table 1: The effect of mutations introduced into the IBM consensus motifs of JPO2 or PogZ on the interaction with LEDGF/p75 as measured by AlphaScreen.**

| MBP-JPO2 | Normalized AlphaScreen signal (%)* |
|---|---|
| wild type | 100 |
| D22A E24A | 56 |
| E80A E82A | 99 |
| D22A E24A E80A E82A | 39 |
| F26A | 24 |
| F26A F29A | 22 |
| F84A | 44 |
| F84A F87A | 41 |
| F26A F84A | 1 |
| F26A F29A F84A F87A | 1 |

| MBP-PogZ $_{1117–1410}$ | Normalized AlphaScreen signal (%)* |
|---|---|
| wild type | 100 |
| E1393A E1395A | 2 |
| F1397A | 1 |
| F1397A F1400A | 1 |

*Normalized values were determined from AlphaScreen signals measured for 200 nM Flag-LEDGF and 1.6 nM MBP-JPO2 or 3.3 nM MBP-PogZ$_{1117–1410}$. AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

Taken together, the relevance of the IBM in interactions of PogZ and JPO2 with IBD was validated by mutation analysis. As the IBM consensus is by definition the same in JPO2, PogZ and MLL1, we presumed the structural features of the respective interaction interfaces would also correspond to each other. To test this hypothesis, we utilized NMR spectroscopy and determined the solution structure of the IBD in complex with a PogZ-derived peptide. This peptide spanned 16 aa residues from E1389 to D1404 (PogZ$_{1389–1404}$). We measured and analyzed a set of triple-resonance experiments to achieve backbone and side chain resonance assignments using $^{15}$N/$^{13}$C-labelled IBD in interaction with unlabeled PogZ$_{1389–1404}$. These assignments were then utilized for specific identification of $^{1}$H–$^{1}$H distance constraints in 3D NOESY experiments. Overall, 94.9% of all observed NOE peaks were assigned yielding 1018 unique $^{1}$H–$^{1}$H distance constraints including 35 intermolecular. After multiple structure calculations, 29 conformation variants with lowest free energy were refined in water. The resulting array of water-refined structures confirmed the initial hypothesis. As expected, IBD was again in the previously described helix bundle fold with PogZ$_{1389–1404}$ anchored between

the two interhelical loops of IBD suggested by all previous experiments (Figure 42A). Analogously to the case of $MLL_{140–160}$, the two critical phenylalanines (F1397 and F1400) anchored the unstructured $PogZ_{1389–1404}$ peptide in the same hydrophobic pockets on the surface of IBD (Figure 42B). The structure is available in the Protein Data Bank (PDB code: 2N3A). NMR constraints and resonance assignments are available in the Biological Magnetic Resonance Data Bank (accession number 25639).

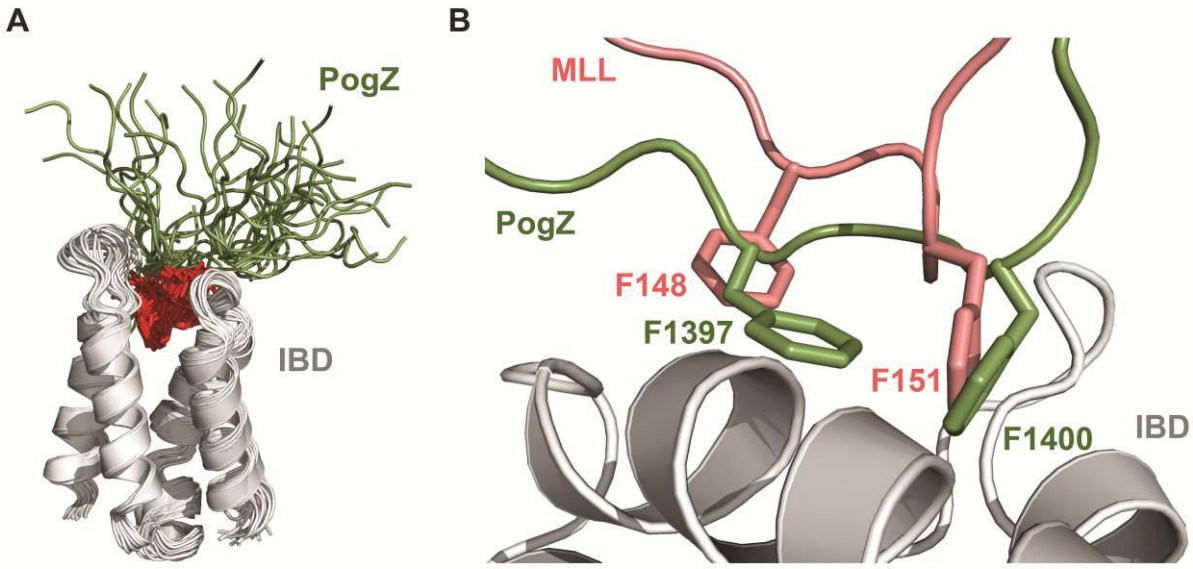


**Figure 42: Solution structure of IBD– $PogZ_{1389–1404}$ interaction**

**(A)** Converged set of structures of IBD (grey tint) interacting with $PogZ_{1389–1404}$ (green) as determined by NMR (PBD code: 2N3A). The two key phenylalanines of $PogZ_{1389–1404}$ (F1397 and F1400) are highlighted as red sticks. **(B)** Comparison of the critical phenylalanine residues of $MLL_{140–160}$ (F148 and F151; salmon) and $PogZ_{1389–1404}$ (F1397 and F1400; green) anchoring both peptides in two hydrophobic pockets on the surface of IBD. One representative structure per complex is depicted.

In conclusion, we functionally validated the newly identified consensus motif utilized by the physiological interaction partners of LEDGF/p75 IBD and characterized it on structural level.

### 5.4.2 IBM in the sequences interaction partners of IBD

After thorough validation of the IBM, we investigated sequences of known interaction partners of IBD, namely ASK and MLL2 and searched for its presence. The IBM was identified in MLL2, close homolog of MLL1, at position 119–126 (Figure 43). In the case of ASK, the IBM was found at the C-terminus spanning residues 626 to 633 in agreement with previously published data by Hughes *et al.* [24]. This study narrowed down the region of ASK responsible for interaction with IBD to the last 50 C-terminal residues (residues 624–674). Interestingly, phosphorylation of this 50 C-terminal region of ASK by CDC7 was necessary for the interaction. This is unique among the physiological interaction partners of the IBD. A possible explanation was uncovered when we compared the identified IBMs. There was a mismatch in the IBM sequence from ASK at position 3 (Figure 43), where glutamic acid was substituted by serine. Should this serine be phosphorylated, its physicochemical properties would be comparable to those of the otherwise conserved glutamic/aspartic acid. Indeed, glutamic or aspartic acids are generally used as phosphoserine-mimicking aa residues. Taking this information into account, we discovered another potential copy of an atypical IBM at the very C-terminus of ASK (residues 667–674). In this case, both IBM acidic residues at positions one and three are substituted with serines. Again, these serines would have to be phosphorylated to substitute for the acidic residues of the consensus.

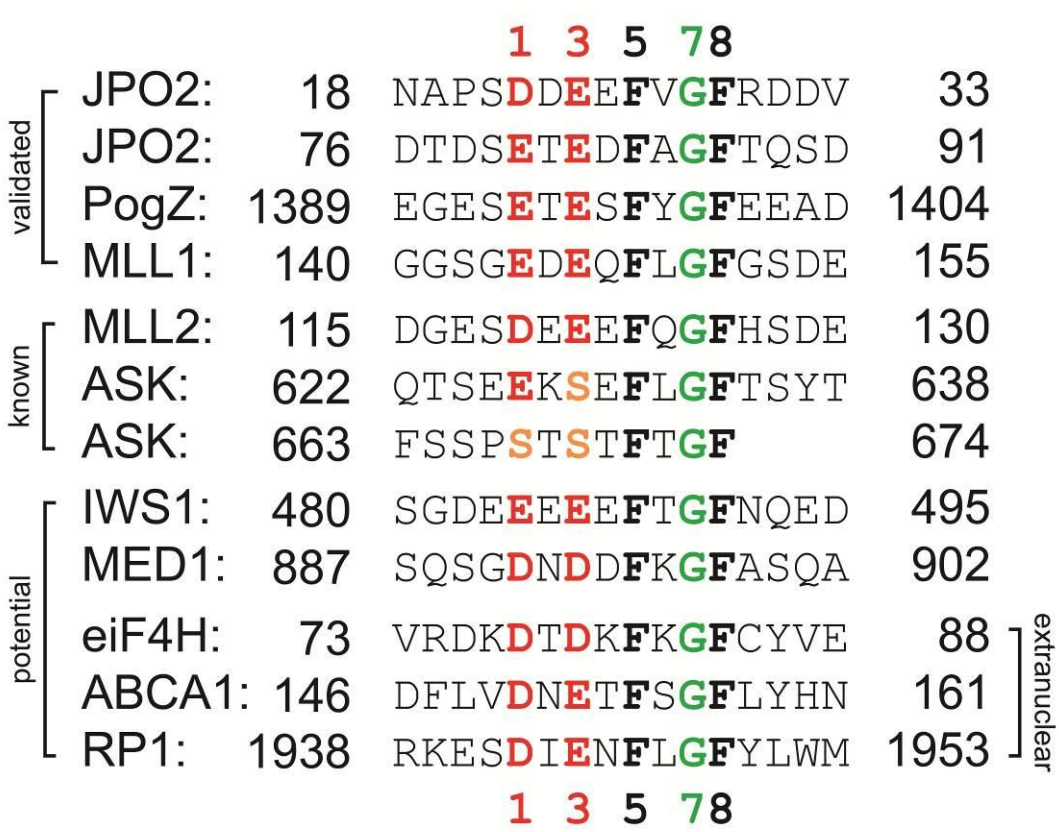


**Figure 43: Partial alignment of sequences containing the IBD-interaction motif (IBM)**

Partial alignment of IBM-containing sequences in various cellular proteins. Proteins that interact with IBD through IBM are marked as validated. Known interaction partners of LEDGF/p75 IBD lacking validation of the IBM role are marked as known. Potential interaction partners without any information regarding their possible interaction with IBD are marked as potential. Potential interaction partners described to be located outside of cell nucleus are marked as extra-nuclear. Amino acid residues of the consensus motif are numbered and colored. Acidic IBM residues are depicted in red, glycine residues in green and phenylalanines in bold black. Serine residues that would simulate physicochemical properties of acidic consensus residues when phosphorylated are colored in orange.

After identifying the IBMs in known interaction partners of LEDGF/p75 IBD, we were also intrigued whether we could use the motif to identify any unknown potential binding partners of LEDGF/p75. We used the ScanProsite database search tool [140] to search for a [E/D]-X-[E/D]-X-F-X-G-F motif in all human proteins. Besides all the known interaction partners of IBD, the IBM was found in several other proteins: mediator complex subunit 1 (Med1), eukaryotic translation initiation factor 4H (eiF4H), ATP-binding cassette transporter 1 (ABCA1), retinitis pigmentosa 1 protein (RP1) and interacts with Spt6 1 (IWS1) (Figure 43). As a candidate protein should be localized in cell nucleus, cytoplasmic or membrane-associated proteins eiF4H, ABCA1 and RP1 were excluded from further studies. Thus, IWS1 and MED1 were identified as potential IBD-interaction candidates. Based on the

characteristics of its sequence, nuclear localization and physiological function, we selected IWS1 for further studies aimed at validation of IWS1–LEDGF/p75 interaction.

## 5.5 IWS1 as a novel interaction partner of IBD

IWS1 is a human transcription factor which plays a key role in defining the composition of the RNA polymerase II elongation complex [141]. It interacts with the H3K36 (histone 3, lysine 36) methyltransferase Hypb/Set2 [141]. As mentioned before, trimethylated H3K36 represents a signature chromatin mark of active transcription and is recognized by PWWP domain of LEDGF/p75 [142, 143]. This would suggest a possible functional link between LEDGF/p75 and IWS1.

To investigate the IWS1 as a potential interaction partner of IBD, we cloned, expressed and purified a fragment encoding residues 352–548 of IWS1 ($IWS1_{352–548}$; Figure 44). This fragment contains the IBM (aa residues 484–491). Fragment boundaries were based on several computational analyses including domain and repeat profiling, secondary structure, disorder and hydrophobicity predictions (data not shown).

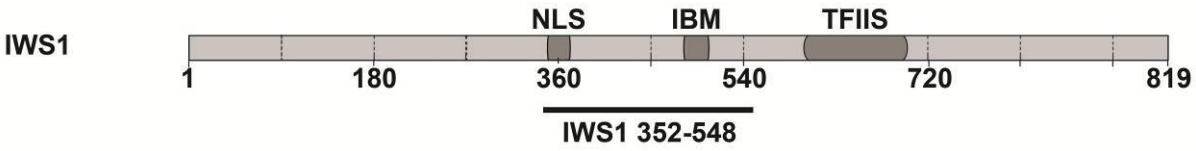


**Figure 44: Domain organization of IWS1**

Sequence of IWS1 with schematically depicted domains and motifs. The IBD-binding motif (IBM) and nuclear localization signal (NLS) are depicted along with the only known structured domain of IWS1, the transcription elongation factor S-II. The expressed fragment $IWS1_{352–548}$ is illustrated by black bar below the scheme.

The $IWS1_{352–548}$ protein manifested aberrant gel migration behavior (Figure 45), similarly to $JPO2_{1–130}$.

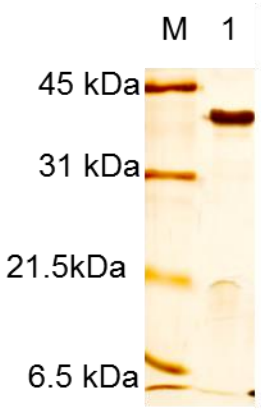


**Figure 45: SDS-PAGE analysis of purified IWS1$_{352–548}$**

Silver-stained SDS-PAGE of purified IWS1$_{1–130}$ (lane 1, theoretical molecular weight: 22.88 kDa). M = molecular weight standards

Although this is a typical trait of intrinsically disordered proteins, we needed to verify the identity of the purified protein in order to proceed with further analyses. The identity of IWS1$_{352–548}$ was verified by LC-MS and MALDI-TOF (Figure 46) mass spectrometric measurements. Over 90% sequence coverage with more than 95% identification confidence was achieved in the peptide mass fingerprinting. The measured intact mass of IWS1$_{352–548}$ corresponded to the theoretical molecular weight of 22.88 kDa.

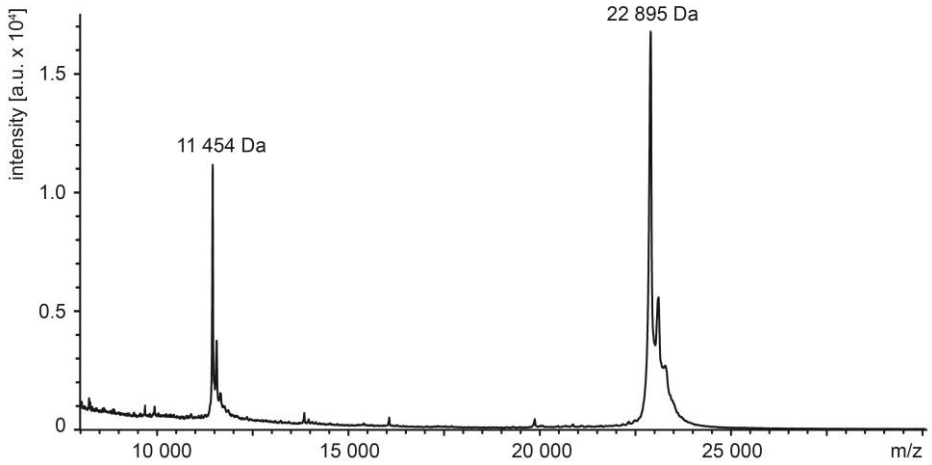


**Figure 46: Mass spectrometry analysis of IWS1$_{352–548}$**

The intact mass of IWS1$_{352–548}$ as measured by MALDI-TOF. The theoretical molecular mass of IWS1$_{352–548}$ (22.88 kDa) was verified.

Majority of the IWS1 sequence was predicted to be disordered by protein disorder prediction tools FoldIndex [138] and PONDR-FIT [139] (Figure 47). The only predicted

structured region overlapped with the TFIIS domain (Figure 44 and 47). These predictions also indicated that a highly acidic glutamate-rich region of IWS1 comprising aa residues 83–509 should be unstructured.

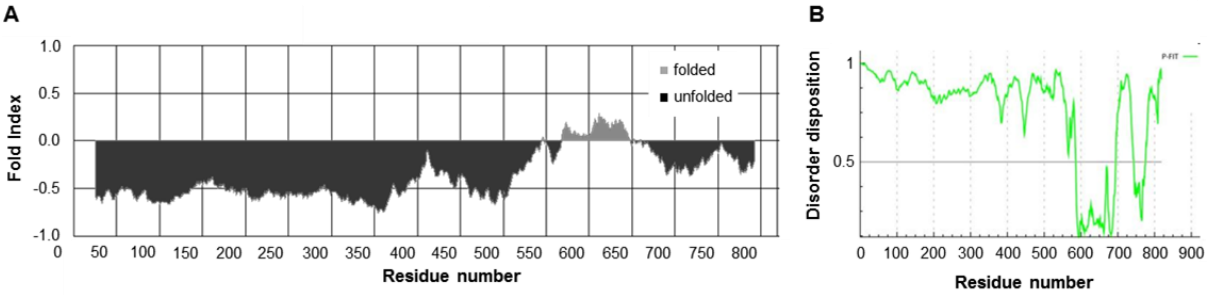


**Figure 47: Protein disorder predictions of IWS1**

**(A)** FoldIndex [138] disorder prediction of IWS1. **(B)** PONDR-FIT [139] disorder prediction of IWS1. Protein disorder prediction tools suggest a large unstructured region in the IWS1 sequence spanning residues 1 to approximately 550. The TFIIS domain is clearly distinguishable due to its structured four-helix bundle character.

We experimentally confirmed the predicted unstructured character of $IWS1_{352–548}$ by multiple independent methods. First, the circular dichroism spectra obtained for $IWS1_{352–548}$ were typical for unstructured protein and did not suggest any regular secondary structure (Figure 48).

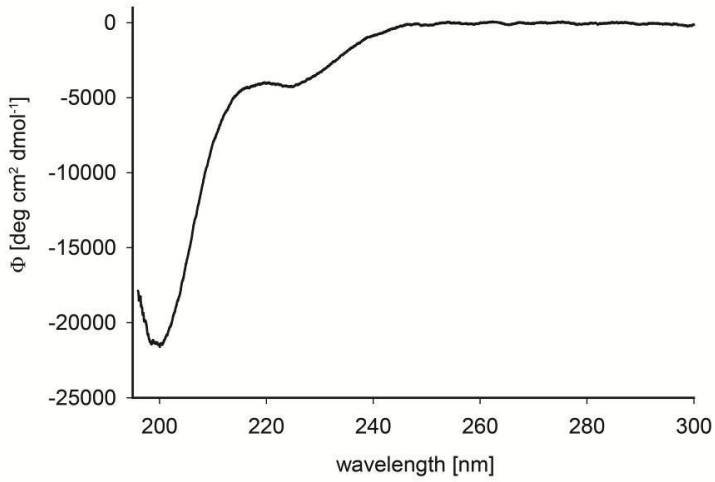


**Figure 48: CD spectrum of $IWS1_{325–548}$**

CD spectrum of $IWS1_{352–548}$ converted to molar ellipticity per residue characteristic for unstructured proteins.

This was not caused by unsuitable buffer composition or pH as the $IWS1_{352–548}$ protein fragment behaved the same in all 46 conditions tested by DSF. Thermal stability screening by DSF did not reveal presence of any hydrophobic core in $IWS1_{352–548}$ and determination of its melting temperature was therefore impossible (Figure 49).

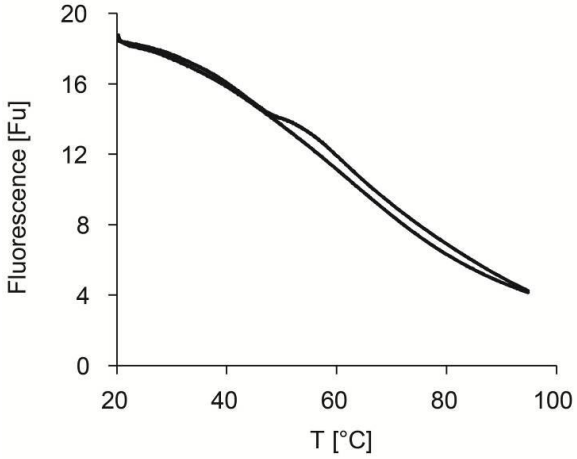


**Figure 49: Typical results of $IWS1_{352–548}$ thermal stability screening monitored by DSF**

Fluorescence intensity of $IWS1_{352–548}$ sample during DSF experiments. Two curves corresponding to experimental duplicates are shown. The fluorescence signal for $IWS1_{352–548}$ was highest at the beginning, with a slow decline over the course of the temperature gradient, showing no thermal unfolding.

Finally, both 1D $^{1}H$ and 2D $^{15}N/^{1}H$ HSQC NMR spectra showed poor dispersion of signals. In particular, the 1D $^{1}H$ NMR spectrum of $IWS1_{352–548}$ amide hydrogens and side-chain methyl groups (at 7.5–8.5 ppm and 0.5–1.0 ppm) manifested very poor signal dispersion (Figure 50A). In the 2D $^{15}N/^{1}H$ HSQC NMR spectrum of $^{15}N$-labelled $IWS1_{352–548}$, backbone and sidechain amide signals were also poorly dispersed and overlapping (Figure 50B).

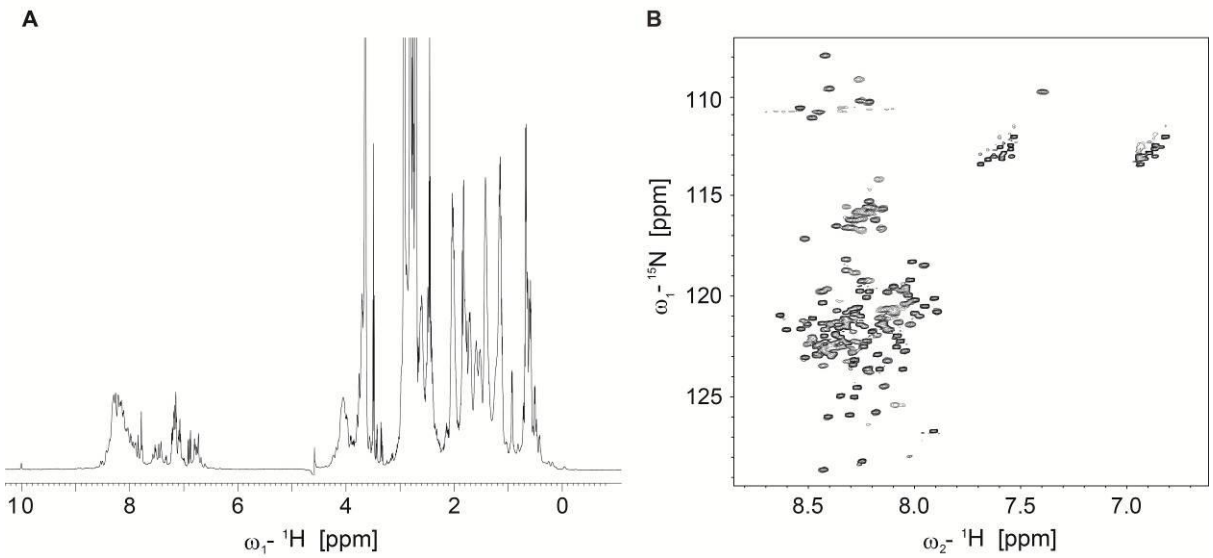


**Figure 50: NMR spectra of IWS1$_{352–548}$**

**(A)** 1D $^{1}$H NMR spectrum of IWS1$_{352–548}$ revealed poor dispersion of all groups of signals, in particular of amide hydrogens and side-chain methyl groups (at 7.5–8.5 ppm and 0.5–1.0 ppm, respectively). **(B)** 2D $^{15}$N/$^{1}$H HSQC NMR spectrum of IWS1$_{352–548}$ showed poor dispersion of amide signals. Both spectra suggested unstructured character of IWS1$_{352–548}$.

In order to test the functional integrity of IWS1$_{352–548}$ and confirm its intrinsically disordered character, we tested its ability to interact with the IBD. We used size exclusion chromatography (in Tris 25 mM, pH 7.8, NaCl 150 mM, 0.05% BME) and observed the isolated IBD– IWS1$_{352–548}$ complex by SDS-PAGE (Figure 51). This indicated relatively high affinity between these two interaction partners.

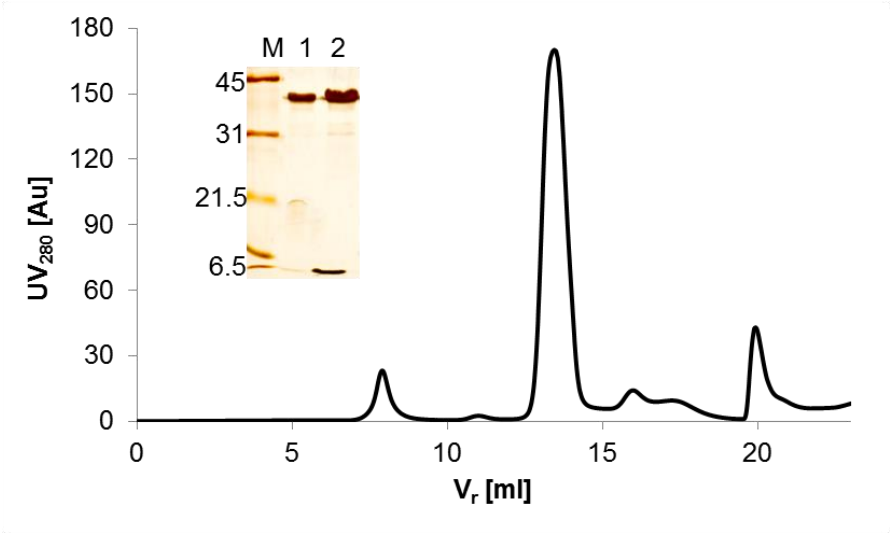


**Figure 51: Size exclusion chromatography of the IBD–IWS1$_{352–548}$ complex**

The chromatogram is overlaid with a silver-stained SDS-PAGE gel of the isolated IBD–IWS1$_{352–548}$ complex (lane 2) and purified IWS1$_{352–548}$ alone (lane 1 for comparison). The IBD–IWS1$_{352–548}$ complex was eluted in the largest peak ($V_r$ = 13.5 ml) The complex was reconstituted prior to chromatography by overnight incubation of equimolar purified IWS1$_{352–548}$ and the IBD in 25 mM Tris-HCl, pH 7.8, 150 mM NaCl, and 0.05% β-mercaptoethanol at 4 °C. The protein molecular weight markers (M) are labeled in kDa.

This interaction was confirmed by ITC, allowing a precise quantification of affinity and thermodynamic parameters (Figure 52). ITC experiments revealed the affinity of this complex to be in low micromolar range ($K_d$ = 6.7 ± 2 µM) with 1:1 stoichiometry ($N$ = 0.96).

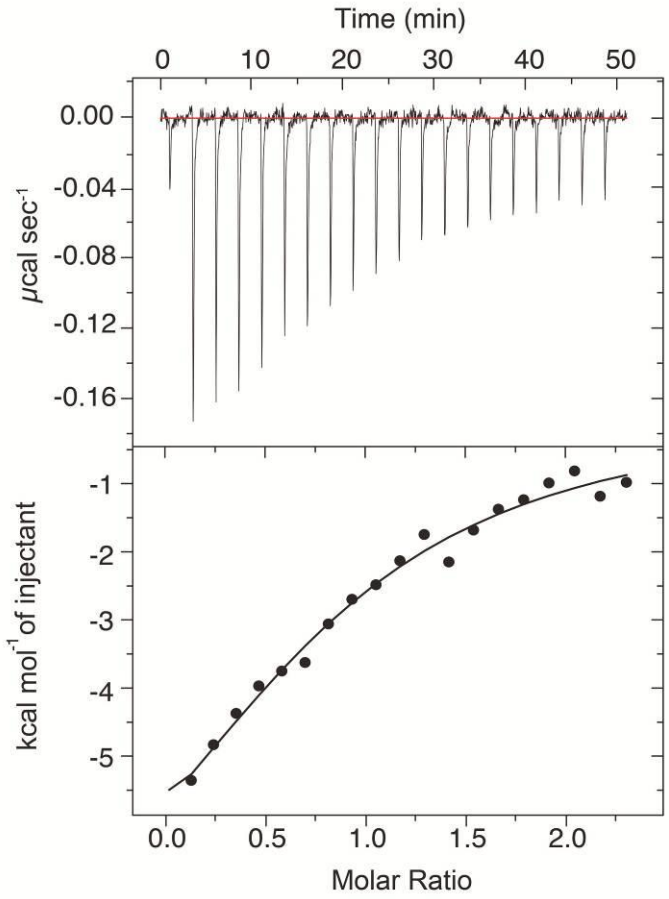


**Figure 52: ITC measurement of the IBD– IWS1$_{352–548}$ interaction**

ITC data obtained for the IBD– IWS1$_{352–548}$ interaction. Upper graph: experimental data; lower graph: fit (line) of the dilution-corrected integrated heats (full circles) from each injection of 145 µM IWS1$_{352–548}$.

Analogously to previous experiments with MLL, JPO2 and PogZ constructs, we evaluated the effect of IBD–IWS1$_{352–548}$ complex formation on IBD backbone in NMR spectra. The minimal chemical shift perturbations approach between free and IWS1$_{352–548}$-bound state was applied, utilizing the assigned 2D and 3D spectra of $^{15}$N/$^{13}$C double-labelled IBD. In the 2D $^{15}$N/$^{1}$H HSQC spectra, we observed significant shifts in the positions of IBD backbone amide signals upon the interaction (Figure 53A). To gain a more detailed information localizing these specific changes, we again analyzed the $^{1}$H, $^{15}$N and $^{13}$C′ chemical shift perturbations in 3D HNCO spectra. Only the minimal signal distances between free and bound state HNCO spectra we used in the analysis. These minimal distances were divided by standard deviation of the dataset comprising all minimal distances. The resulting processed distance values affecting individual aa residues of IBD induced by interaction with IWS1$_{352–548}$ were plotted in a graph (Figure 53B).

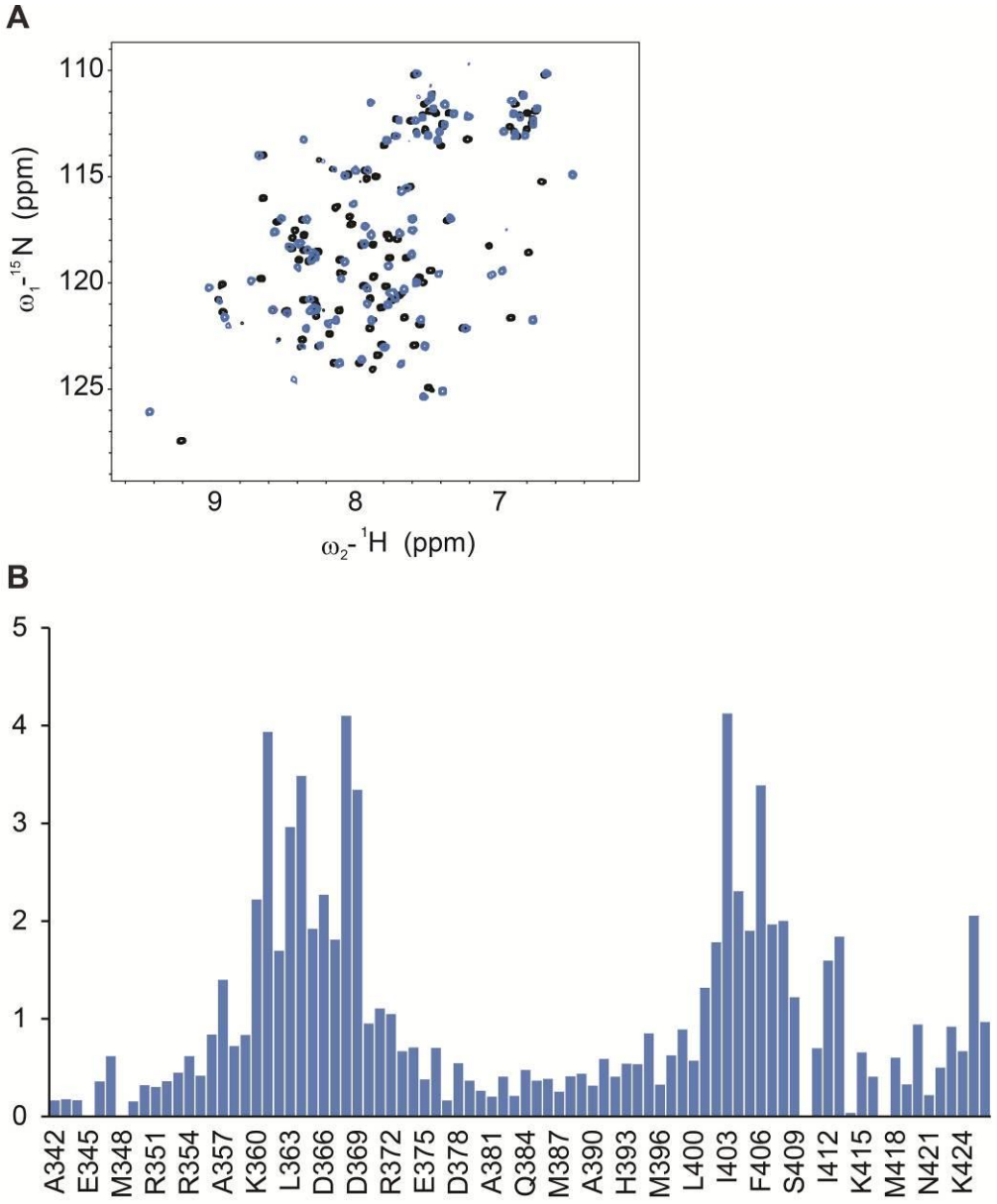


**Figure 53: Cellular partner induced changes in the backbone of IBD**

**(A)** Comparison of the 2D $^{15}N/^{1}H$ HSQC spectra of the free (black) and $IWS1_{352–548}$-bound (blue) IBD. The blue spectrum in was obtained from the complex of $^{15}N$-labelled IBD with unlabeled $IWS1_{352–548}$. **(B)** The column graph summarizes the minimal 3D chemical shifts of backbone resonances ($^{15}N$, $^{13}C'$ and $^{1}H^{N}$) divided by standard deviation of all minimal 3D shifts. The resulting values quantify degree of localized IBD backbone perturbations upon interaction with $IWS1_{352–548}$.

As expected, the interaction with $IWS1_{352–548}$ affected IBD residues comparable to those affected by other partners we investigated previously. Specifically, significant (over 1.5× standard deviation) changes were localized to residues K360–D369, K402–V408, I412 and M413 (Figure 53B). These residues mostly cover two interhelical loops that are located

close to each other on the surface of the IBD and were identified as interaction interface for $MLL_{140–160}$, $PogZ_{1117–1410}$ and $JPO2_{1–130}$ (Figure 54).

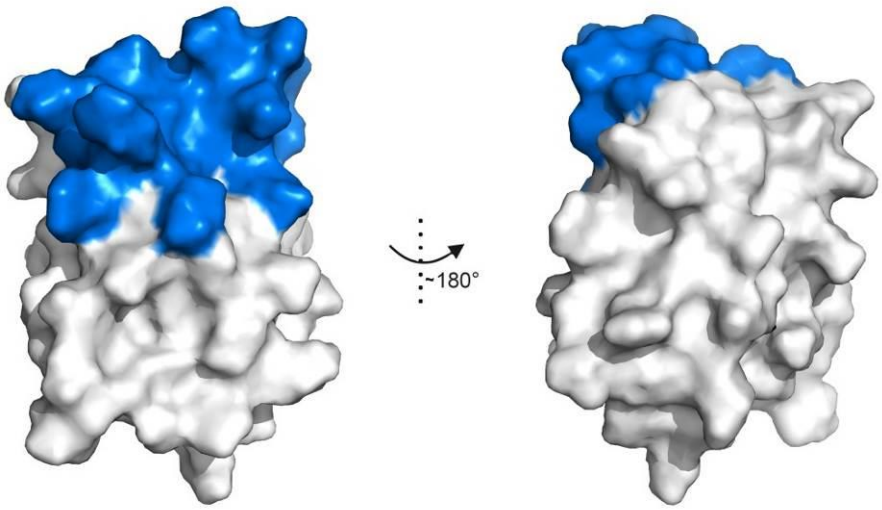


**Figure 54: Structural overview of IBD interaction interface for $IWS1_{352–548}$**

Specific changes induced in the IBD backbone resonances by interaction with $IWS1_{352–548}$ (blue). Residues with substantial chemical shifts (over 1.5× standard deviation) are highlighted on the IBD structure surface representation.

Our results demonstrated that MLL1, JPO2, PogZ and IWS1 interact with an overlapping interaction interface on IBD. As published previously, interactions of HIV IN JPO2, PogZ and MLL1 with IBD are mutually exclusive [22, 56, 137]. Therefore, the binding of IWS1 to IBD should also be mutually exclusive to the other partners. To test this hypothesis, we performed competition titration experiments of HIV IN catalytic core domain (CCD), $JPO2_{1–130}$, $PogZ_{1117–1410}$ and $MLL1_{123–160}$ against the LEDGF/p75–IWS1 complex monitored by AlphaScreen (Figure 55).

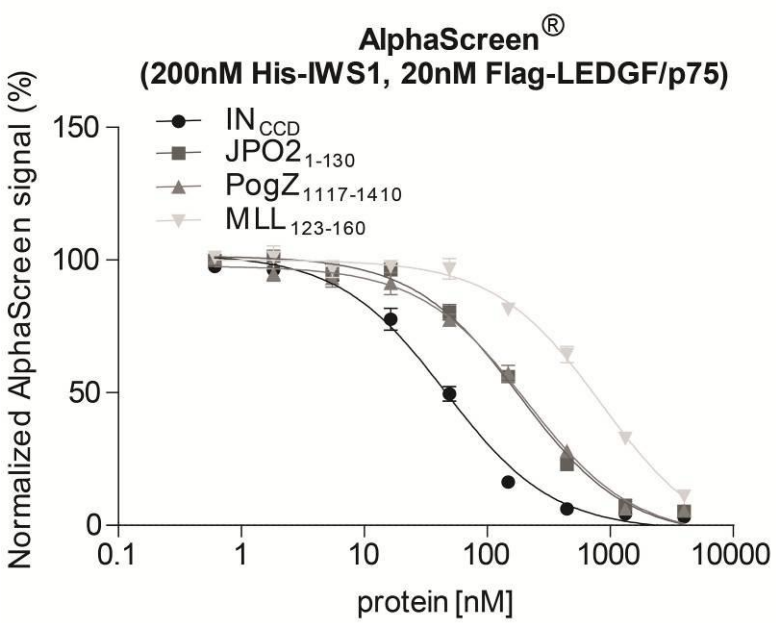


**Figure 55: Competition assay of LEDGF/p75 interaction partners**

All tested protein fragments (HIV IN CCD, $JPO2_{1–130}$, $PogZ_{1117–1410}$ and $MLL1_{123–160}$) interfered with LEDGF/p75–IWS1 complex formation in vitro. The interaction between recombinant full-length Flag-LEDGF/p75 (20 nM) and His-IWS1 (200 nM) was monitored in an AlphaScreen assay. Increasing amounts of single untagged competing protein (either HIV IN CCD, $JPO2_{1–130}$, $PogZ_{1117–1410}$ or $MLL1_{123–160}$) were added. Error bars represent standard deviations calculated from two independent experiments performed in duplicate. AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

All the tested untagged proteins (HIV IN CCD, $JPO2_{1–130}$, $PogZ_{1117–1410}$ and $MLL1_{123–160}$) competed for LEDGF/p75 binding with IWS1 in a concentration-dependent manner.

Taken together, all the obtained results unambiguously confirm a novel interaction between IWS1 and the IBD of LEDGF/p75. This interaction was revealed based on the identification of the IBM and is mediated through this motif.

## 5.6 Comparison of HIV IN and cellular binding partners' interaction interfaces on IBD

As described above, IBD represents an attractive therapeutic target for inhibition of its interaction with either HIV IN or MLL1. We identified and characterized a common part of interface on IBD used by physiological partners as well as by the HIV IN. Thus, a more

detailed analysis of available structural data was needed to evaluate possible specific targeting of inhibition.

The overlapping IBM-binding site of IBD is formed mostly by the interhelical loops that bridge helices α1 with α2 and α4 with α5. Although the IBM was responsible for all physiological interactions with IBD we studied, there were minor differences among these interactions identified by minimal chemical shift perturbations mapping. Therefore we performed a detailed comparison of IBD residues interacting with HIV IN and cellular binding partners. IBD residues buried upon formation of the complex of IBD with the CCD of HIV IN (revealed in the crystal structure of the complex PDB code 2B4J; [37]) were compared with residues exhibiting significant chemical shift changes upon $JPO2_{1–130}$, $PogZ_{1117–1410}$ or $IWS1_{352–548}$ binding (Figure 56 and Table 2).

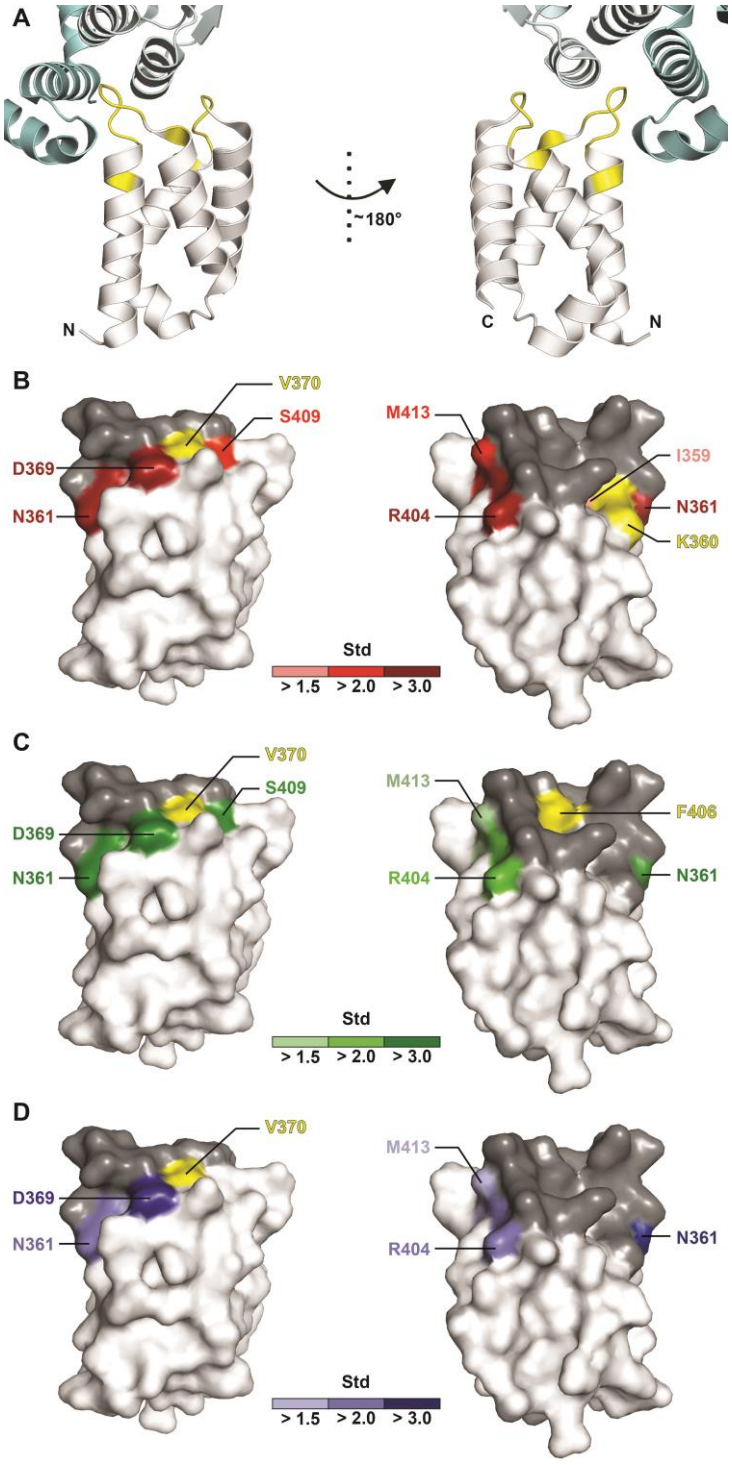


**Figure 56: IBD interfaces for HIV IN, JPO2$_{1–130}$, PogZ$_{1117–1410}$ and IWS$_{352–548}$**

Comparison of IBD interaction interfaces as determined by X-ray crystallography (A) and NMR (B, C, D). The interaction interface shared by HIV IN and the particular cellular binding partner highlighted on the IBD surface in dark grey. The color gradients reflect the relative degree of chemical shift perturbations on protein solvent-accessible surface as a multiple of standard deviation of the corresponding dataset. Residues interacting with HIV IN that were not detected to interact with the particular cellular binding partner are colored yellow (B, C, and D). **(A)** Crystal structure showing the interaction of IBD with HIV IN CCD (PDB code 2B4J; [37]). Proteins are shown as cartoons with dimeric HIV IN CCD colored in light blue, IBD in white and the interacting IBD aa residues in yellow. **(B)** Residues interacting with JPO2$_{1–130}$ but not with HIV IN are highlighted in gradients of red. **(C)** Residues interacting with PogZ$_{1117–1410}$ but not with HIV IN are highlighted in gradients of green. **(D)** Residues interacting with IWS$_{352–548}$ but not with HIV IN are highlighted in gradients of blue.

**Table 2: Comparison of IBD interactions with HIV IN, $JPO2_{1–130}$, $PogZ_{1117–1410}$, $MLL_{1–160}$ and $IWS_{352–548}$**

| IBD residues | | surface area buried (BSA) upon IBD-HIV IN complex formation | | minimal backbone shift values over standard deviation | | | |
|---|---|---|---|---|---|---|---|
| | | BSA[a] [Å$^2$] | BSA[a] [%] | upon Jpo2 binding | upon PogZ binding | Upon IWS1 binding | Upon MLL1 binding |
| 359 | Ile | 0 | 0 | > 1.5 | < 1.5 | < 1.5 | > 2.0 |
| 360 | Lys | 39.9 | 37.9 | > 1.5 | > 2.0 | > 2.0 | > 2.0 |
| 361 | Asn | 0 | 0 | > 3.5 | > 3.5 | > 3.5 | > 2.0 |
| 362 | Ser | 0 | 0 | > 1.5 | > 1.5 | > 1.5 | > 1.5 |
| 363 | Leu | 3.8 | 16.7 | > 2.5 | > 2.5 | > 2.5 | > 2.0 |
| 364 | Lys | 76.2 | 69.8 | > 3.0 | > 3.0 | > 3.0 | > 2.0 |
| 365 | Ile | 156.5 | 96.3 | > 1.5 | > 1.5 | > 1.5 | > 2.5 |
| 366 | Asp | 116.7 | 89.1 | > 2.5 | > 2.0 | > 2.0 | < 1.5 |
| 367 | Asn | 26.3 | 23.2 | > 1.5 | > 1.5 | > 1.5 | > 2.5 |
| 368 | Leu | 41.2 | 57.1 | > 4.0 | > 4.0 | > 4.0 | > 3.0 |
| 369 | Asp | 0 | 0 | > 3.5 | > 3.5 | > 3.0 | > 1.5 |
| 370 | Val | 5.3 | 13.5 | < 1.5 | < 1.5 | < 1.5 | < 1.5 |
| 402 | Lys | 50.6 | 32.8 | > 1.5 | > 1.5 | > 1.5 | > 1.5 |
| 403 | Ile | 5.9 | 49.7 | > 3.5 | > 3.5 | > 4.0 | > 4.0 |
| 404 | Arg | 0 | 0 | > 2.5 | > 2.0 | > 2.0 | > 2.0 |
| 405 | Arg | 14.1 | 8.9 | > 2.0 | > 2.0 | > 1.5 | > 2.5 |
| 406 | Phe | 34.7 | 99.1 | > 1.5 | < 1.5 | > 3.0 | > 2.5 |
| 407 | Lys | 15.3 | 9.9 | > 1.5 | > 1.5 | > 1.5 | > 2.5 |
| 408 | Val | 75.0 | 75.2 | > 3.0 | > 3.5 | > 2.0 | > 3.5 |
| 409 | Ser | 0 | 0 | > 1.5 | > 2.0 | < 1.5 | > 1.5 |
| 412 | Ile | 0 | 0 | > 1.5 | > 1.5 | > 1.5 | > 1.5 |
| 413 | Met | 0 | 0 | > 2.0 | < 1.5 | > 1.5 | > 1.5 |

[a] The solvent-accessible surface area (BSA) of individual IBD residues buried upon IBD–HIV IN complex formation is expressed in Å$^2$ and as a portion of the residue total solvent accessible area (in %). Average values for two IBD chains in crystal structure 2B4J are listed. BSA was calculated using PDBePISA tool [144].
Minimal backbone shift values lower than 1.5 × standard deviation are considered insignificant.

Table 2 summarizes calculated surface areas of individual IBD residues buried upon IBD–HIV IN interaction and relative degrees of their backbone chemical shift perturbations upon interaction with individual physiological binding partners. Both Table 2 and Figure 56 reveal variability among the IBD interfaces for individual interaction partners. However, a major part of the interface is common for all the interaction partners as revealed by our detailed NMR interface mapping experiments.

Combining the previously published crystal structure of IBD–MLL1–menin (PDB code: 3U88; [96]) with our structural information provided further detail in comparison of IBD–HIV IN and IBD–MLL1 interactions (Figure 57). Through the consensus motif MLL1 recognizes both the basic surface created by positively charged residues of the α4 helix together with the inter-helical loop connecting α4-α5 (Figure 57A) and the hydrophobic pockets created by inter-helical loops connecting IBD alpha helices α1-α2 and α4-α5 (aa residues I359, K360, L363, L368, T399, K402, I403, F406, K407 and V408; Figure 57B). The hydrophobic pockets are occupied by the MLL1 F148 and F151 residues, which represent phenylalanines conserved in the IBM. The basic surface supports the interaction by forming electrostatic contacts with the IBM containing conserved acidic residues represented by E144 and E146 of MLL1 (Figure 57A).

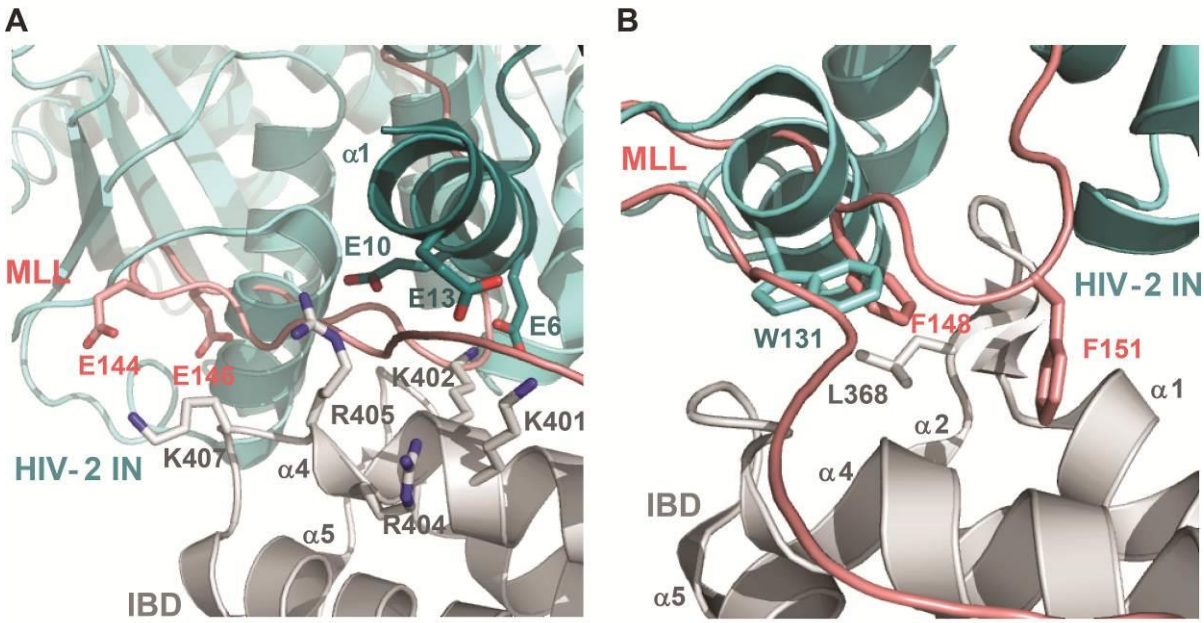


**Figure 57: Comparison of IBD–MLL1 and IBD–HIV-2 IN interfaces**

**(A)** The basic patch on the IBD recognized by HIV-2 IN E10 and MLL1 E144. **(B)** The hydrophobic pocket on IBD recognized by HIV-2 IN W131 and MLL1 F148.
Structural information were combined for the IBD–MLL1 and IBD–HIV-2 IN complexes (PDB codes: 2MSR [137] and 3F9K [39] respectively).

From all the available LEDGF/p75–HIV IN structures, only one contains both the N-terminal domain (NTD) and the catalytic core domain (CCD) of HIV integrase in complex with IBD (PDB code: 3F9K; [39]). The structural comparison clearly shows an overlap between the MLL1 and HIV-2 IN interfaces on the IBD (Figure 57). One of the hydrophobic pockets on the IBD surface is occupied by W131 of HIV-2 IN. The basic surface of

LEDGF/75 is recognized by the NTD of HIV-2 IN. The acidic residues of the α1 helix of the NTD containing E6, E10, and E13 face positively charged residues of the IBD (K401, K402, R404, and R405) (Figure 57A). It was reported previously, that LEDGF/p75 residues K401, R404 and R405 play a crucial role in HIV infectivity [39]. Interestingly, their interaction counterparts (E10 and E13 of HIV-2 IN NTD) are conserved among lentiviral integrases (Figure 58).

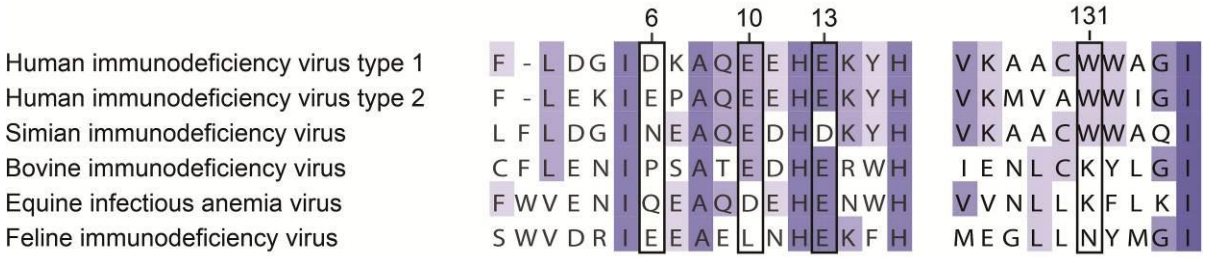


**Figure 58: Partial amino-acid sequence alignment of lentiviral integrases**

Residues of NTD and CCD responsible for interaction with LEDGF/p75 IBD are depicted. Conserved residues are colored according to percentage of identity. Amino-acid residues corresponding to HIV-2 IN E6, E10, E13 and W131 are boxed.

Based on our structural data, we completed a comprehensive mutation screening of full-length LEDGF/p75 to identify mutants defective for interaction with MLL1 (for details see Čermáková *et al.* [137]). These mutations expanded those identified by Bartholomeeusen *et al.* [22] as defective for interaction with JPO2 or PogZ. Among other aa residues, we targeted L368, which directly engages F148 of MLL1 or W131 of HIV integrases (Figure 59). Next, we tested these mutations for interactions with IWS1, $PogZ_{1117–1410}$ or HIV IN. As expected, the LEDGF/p75 L368A mutation reduced the interaction with all binding partners (Figure 59).

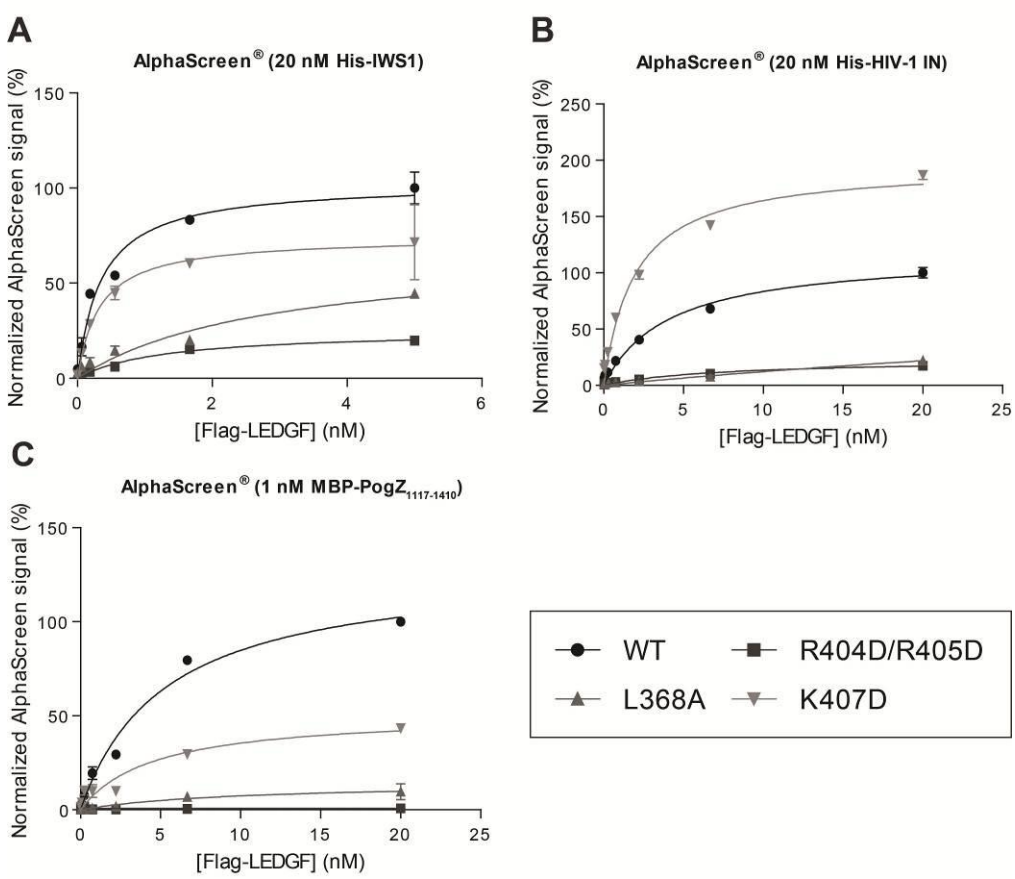


**Figure 59: Effects of LEDGF/p75 mutations on its interactions**

Interactions of full-length LEDGF/p75 with either IWS1, JPO2, $PogZ_{1117–1410}$ or HIV IN monitored by AlphaScreen. Three different LEDGF/p75 mutants versions defective for binding to MLL1 (R404D/R405D, L368A or K407D) were tested for interaction with one other binding partner along with wild-type (WT) LEDGF/p75. **(A)** Interactions between recombinant full-length WT His-IWS1 and LEDGF/p75 protein constructs. **(B)** Interactions between recombinant full-length His-HIV IN and LEDGF/p75 constructs. **(C)** Interactions between recombinant MBP-$PogZ_{1117–1410}$ and LEDGF/p75 constructs.
Error bars represent standard deviations calculated from three independent experiments, each performed in duplicate. AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

A comparable effect was obtained with R404D/R405D double mutant. Both R404 and R405 bridge the interaction with E10 of HIV-2 IN as well as electrostatic contacts with MLL1. Unlike these residues, LEDGF/p75 K407 is involved in interaction with negatively charged side-chains of MLL1 but does not engage in any contacts with HIV IN (Figure 59). Therefore, the K407D mutation decreased interaction affinities of LEDGF/p75 to IWS1 and $PogZ_{1117–1410}$ but increased its binding to HIV IN (Figure 59).

In conclusion, the mutation analyses corroborated our structural data and confirmed the overlapping interaction interface on IBD for all its interaction partners. This universal

overlapping surface includes one of the hydrophobic pockets and the basic patch necessary for electrostatic contacts.

However, the other hydrophobic pocket critical for anchoring the second phenylalanine of the IBM (represented by F151 of MLL1 in Figure 57) is not in contact with any HIV IN aa residue side-chains. Moreover, there is an additional interface identified between MLL1 and IBD maintained by MLL1 aa residues F129 and F133 (for details see Chapter 3.2.3 and Figures 8 and 19) [96]. These MLL1 residues contact I397, Y420, K424 and F427 of the IBD. This represents a unique feature of MLL1 among all the interaction partners of LEDGF/p75 IBD (Figure 60).

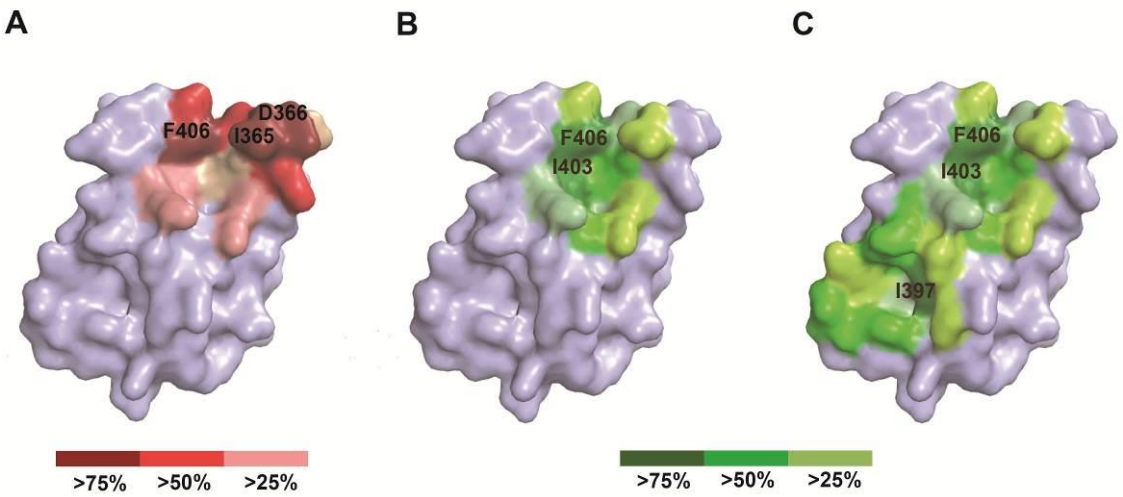


**Figure 60. Most affected surface area of IBD upon interaction with MLL1 or HIV IN**

The color gradients reflect the most affected aa residues of IBD upon particular interaction as a function of buried surface area. **(A)** Interaction with HIV-1 integrase. **(B)** Menin-independent interface for $MLL_{140–160}$. (**C**) Combination of previously published IBD–MLL1 interface with our structural data for $MLL_{140–160}$ interaction (PDB codes: 3U88 [96] and 2MSR). Solvent-accessible surface areas (BSAs) were calculated using PDBePISA tool [144].

Thus, MLL1 contacts an overlapping part IBD common to all interaction partners as well as an additional one. Considering this unique character of IBD–MLL1 interaction, we tested a set of LEDGF/p75 mutant constructs defective for interaction with HIV IN as defined by Cherepanov *et al.* [38] for their ability to interact with $MLL1_{1–160}$ (Table 3).

**Table 3: Effects of LEDGF/p75 mutations on interaction with $MLL1_{1–160}$**

LEDGF/p75 mutants defective for binding to HIV IN were tested for interaction with wild-type (WT) $MLL1_{1–160}$ as measured by AlphaScreen (0.3 nM Flag-LEDGF, 1 nM $MLL_{1–160}$-GST). AlphaScreen experiments were performed by Kateřina Čermáková at KU Leuven and are used with her consent.

| Flag-LEDGF/p75 | Normalized AlphaScreen signal (%) |
|---|---|
| WT | 100.0 |
| K360A | 84.9 |
| I365A | 88.2 |
| D366N | 105.7 |
| V370A | 94.6 |
| F406A | 82.6 |
| V408A | 74.8 |

Interestingly, none of the LEDGF/p75 mutant constructs defective for interaction with HIV IN dramatically affected the interaction with $MLL1_{1–160}$. This confirmed the specific character of the IBD–MLL1 interaction involving a distinct set of IBD residues.

# 6. DISCUSSION

LEDGF/p75 is a highly conserved nuclear protein involved in a number of important physiological and pathophysiological processes. As a transcriptional co-activator, it has been implicated in various types of stress responses by increasing the expression of stress-associated genes and small heat shock proteins [8, 11-14]. More recently, it has been highlighted as a marker of transcriptionally active chromatin, suggesting a more general function [2]. LEDGF/p75 associates with active chromatin through its PWWP domain which recognizes H3K36me3 [2, 32]. The crucial role of LEDGF/p75 has been supported by perinatal mortality observed in mice with knockout of its coding gene *PSIP1* [16].

LEDGF/p75 is implicated in the pathogeneses of various types of cancer and also plays a prominent role in the HIV life-cycle [32, 33, 40-42]. The role of LEDGF/p75 in HIV integration has been studied extensively, and high-resolution structural data describing its interaction with HIV IN are available [37]. This information enabled the development of inhibitors that disrupt the interaction between LEDGF/p75 and HIV IN and dramatically reduce the viability of the virus (reviewed by Crist *et al.* [145]). In contrast, the cellular role of the HIV IN binding domain (IBD) of LEDGF/p75 is not well understood. Perhaps the most interesting question regarding the IBD as a potential therapeutic target is whether it would be possible to target one of its many protein-protein interactions while not interfering with the others. Moreover, gaining deeper insight into the interactions of IBD with cellular binding partners may contribute to the understanding of complex biological processes and facilitate future efforts in the development of efficient chemotherapies.

Such chemotherapies could be used for treatment of MLL fusion–mediated transformation by selectively blocking the LEDGF/p75–MLL1 interaction while not interfering with the MLL1–menin complex formation. We have shown that this would be possible by identifying a novel menin-independent interaction interface between LEDGF/p75 IBD and MLL1 [137]. This represents an important extension of the previously reported menin-dependent interaction interface capable of maintaining the MLL1–LEDGF/p75 interaction *in vitro*. Our structural characterization by NMR revealed details of the interaction between LEDGF/p75 IBD and MLL1-derived peptide $MLL_{140-160}$ (PDB code 2MSR). In this MLL1 peptide, the aromatic side-chain ring of F148 is in contact with IBD aa residues L363, L368, I403, F406, K407, and V408. Another phenylalanine F151 resides in a deeper hydrophobic pocket contacting IBD aa residues I359, K360, L363, T399, K402, and I403.

Moreover, critical role of MLL1 glutamic acids E144 and E146 was revealed by mutation analysis. Contacts of these residues with the IBD were not detected in our NMR experiments. This is probably due to deprotonated character of their carboxyl groups in solution. Carboxylate anions lack protons that would allow $^{1}H–^{1}H$ NOE contacts detection.

The LEDGF/p75–MLL–menin complex was suggested as a new therapeutic target in MLL fusion–driven acute leukemias by several independent studies [21, 95, 146]. Small molecule inhibitors targeting the interaction between MLL1 and menin have been identified [147, 148]. However, it is not clear how disrupting this interaction would affect the role of menin as a tumor suppressor protein. Targeting the menin-independent interface between MLL1 and LEDGF/p75 could also be effective without blocking the interaction of MLL1 with menin. Utilizing our structural characterization of this interface, we tested inhibition by cyclic peptide CP65 previously identified to target the site of menin-independent MLL1 interface on IBD [124, 149]. This peptide inhibitor proved to specifically block the LEDGF/p75–MLL1 interaction. When expressed in cells, it reduced MLL fusion–dependent leukemic transformation in mice primary murine leukemic blasts by 66%. This effect could be cancelled by overexpression of WT LEDGF/p75 but not by mutant MLL1-binding defective LEDGF/p75. Overall, these results proved that the IBD–MLL1 interface represents a potential therapeutic target in MLL fusion–driven acute leukemias. Our results were later independently corroborated by Murai *et al.* [150]. In this study, overexpression of $MLL_{110–160}$ or HIV IN significantly inhibited colony formation of MLL-AF9-transformed bone marrow cells.

Next, we aimed to characterize interactions of IBD with two physiological partners of LEDGF/p75: JPO2 and PogZ. We successfully cloned, expressed and purified the domains of JPO2 and PogZ involved in binding of the LEDGF/p75 IBD. Biochemical and biophysical characterization of the optimized protein constructs revealed that they both interact with IBD through intrinsically unstructured regions. Monitoring the shifts in the positions of backbone NMR signals ($^{15}N$, $^{13}C'$ and $^{1}H^{N}$) of isotopically labelled $JPO2_{1-130}$ and $PogZ_{1117-1410}$ led to identification of amino acid residues involved in interactions with IBD. Utilizing these data together with the detailed structural information of IBD–MLL1 interaction, we were able to reveal a consensus IBD-binding motif of these physiological binding partners. This [E/D]-X-E-X-F-X-G-F motif was called IBM (IBD-binding motif) and validated by comprehensive mutation analysis using full-length proteins when possible.

In JPO2, the IBM is present in two copies. Previous studies identified aa residue stretches sufficient for interaction with IBD called specific interaction domains (SIDs) [22,

23, 56]. There were two SIDs identified in JPO2 by two independent studies (aa residues 62–94 or 77–98) [23, 56]. We discovered that one copy of IBM lies within the previously proposed SID, spanning residues E80–F87. The other IBM is located between residues D22 and F29 and had not been identified previously. Presence of two IBM copies explained the intriguing results of our NMR and ITC experiments. In ITC, the resulting stoichiometry suggested two molecules of IBD interacting with one molecule of $JPO2_{1\text{-}130}$. In the NMR minimal backbone shift analysis, $JPO2_{1\text{-}130}$ residues V27, G28 and F29 were significantly affected by the interaction with IBD.

In PogZ the consensus motif was identified at the C-terminus spanning residues E1393–F1400 and verified by comprehensive mutation analysis monitored by AlphaScreen. This position is not in agreement with the previously proposed SID between residues 1139–1248 [22]. This discrepancy can be explained by different experimental approach using several PogZ constructs in yeast two-hybrid assay, which does not provide high resolution information on amino acid level. We have structurally characterized and validated the IBM by solution structure of the IBD–$PogZ_{1389–1404}$ complex (PDB code: 2N3A). In this complex, the IBM sequence exhibited protein-protein contacts analogous to those found in the IBD–$MLL_{140\text{-}160}$ complex. The two critical phenylalanines (F1397 and F1400) anchored the unstructured $PogZ_{1389–1404}$ peptide in the same hydrophobic pockets on the surface of IBD. Critical role of the two conserved glutamic acids (E1393 and E1395 in PogZ) was again revealed by mutation analysis.

After validating the IBM, we set out to investigate whether this motif could be responsible for interactions of other known partners of IBD. We have found IBMs in the two remaining established binding partners of IBD: MLL2 and ASK [3, 24]. In the case of ASK, two copies of IMB are present at the very C-terminus, which is consistent with previously published approximation [24]. However, there are substitutions of glutamic acid residues by serines in both IBM copies. These serines would have to be phosphorylated to substitute for the acidic residues of the consensus, which is consistent with the previously described mechanism of LEDGF/p75–CDC7–ASK interaction regulation [24]. This would represent a unique conditional interaction feature among the physiological interaction partners of the IBD. However, there are phosphorylation sites in close vicinity to IBMs in other IBD physiological interaction partners (www.phosphosite.org [151]). Theoretically, phosphorylation in the close vicinity of the conserved acidic residues of the IBM could increase the electrostatic compatibility to the opposite basic patch on IBD surface. Interestingly, these sites are proposed to be phosphorylated by Casein Kinase 2 (CK2), based

on its target sequence specificity. Although this assumption remains purely speculative, CK2 is the closest homolog of CDC7, a confirmed phosphorylation agent of the C-terminal part of ASK [24, 104]. Moreover, both these kinases are inhibited by analogous inhibitors, indicating structural similarity [101, 152]. However, effects of any post-translational modifications of IBM-containing proteins remain to be elucidated.

Next to the known LEDGF/p75 interaction partners, several other human proteins were found to contain an IBM, namely: IWS1, Med1, eiF4H, ABCA1 and RP1. However, the IBMs of eiF4H, ABCA1 and RP1 are situated in structured regions according to disorder predictions and domain composition. These motifs are therefore unlikely to be available for interaction with LEDGF/p75 IBD. This is also supported by extranuclear localization and physiological roles of these proteins. Thus, only IWS1 and Med1 were predicted as potential IBD-interaction candidates. Based on its interesting physiological function and disorder predictions, we selected IWS1 for further investigation.

IWS1 forms a stable complex with SPT6 and the CTD of the large subunit of RNA polymerase II [153]. In addition, IWS1 has been shown to recruit HYPB/Setd2, the only histone methyltransferase known to create H3K36me3 marks [141]. Since the LEDGF/p75 PWWP specifically recognizes these histone marks, one could speculate on a feedback loop where LEDGF/p75 binding to H3K36me3 marks would be followed by the recruitment of the Setd2-IWS1(-Spt6) complex and increased H3K36 methylation. Of note, the H3K36me3 marks have been linked to RNA splicing. The distribution of H3K36me3 depends on splicing [154, 155] and splicing is regulated by H3K36me3 marks [142, 156]. Genes with multiple introns have a higher amount of H3K36me3 which is proportional to the transcription rate [155, 157]. Interestingly, LEDGF has also been linked to alternative splicing [142]. Moreover, LEDGF/p75 interacts with transcription factor NOVA1 through its PWWP domain [158]. As such, the LEDGF/p75-Setd2-IWS1(-Spt6)-complex might represent a mediator in the bidirectional communication between H3K36 trimethylation and pre-mRNA splicing [155].

We have confirmed the interaction of IWS1 with LEDGF/p75 by various independent methods including NMR, ITC, AlphaScreen and pull-down assays. The intrinsically disordered character of IWS1 was experimentally established by comprehensive biophysical characterization. Thus, in all the studied IBD physiological interaction partners, the IBM is present as a short linear motif (SLiM) in an intrinsically disordered part. Analogously to SLiMs, intrinsically disordered proteins (IDPs) are also typically responsible for mediating transient protein-protein interactions (reviewed by Van Roey *et al.* [159]). Although many

IDPs undergo binding-induced folding [160], some stay significantly disordered upon interaction, resulting in heterogeneous "fuzzy" complexes [161]. These proteins are prevalent in higher eukaryotes where they carry out specific biological functions albeit lacking stable tertiary structure (reviewed by Dunker *et al.* [162]). Among the most frequent processes involving IDPs are those connected with transcription regulation [162, 163]. The SLiMs allow for high level of plasticity and complexity in these regulations. Numerous pathogens have evolved the ability to mimic SLiMs and usurp the cellular mechanisms they regulate (reviewed by Via *et al.* [164]).

We have shown that the IBD of LEDGF/p75 serves as a hub for physiological interactions with intrinsically disordered consensus motif present in several physiological binding partners of LEDGF/p75. In the case of JPO2, ASK and PogZ, two of these motives occur in close vicinity (Figure 61) as a result of either dimerization (PogZ) or multiplicity in single protein sequence (JPO2 and ASK).

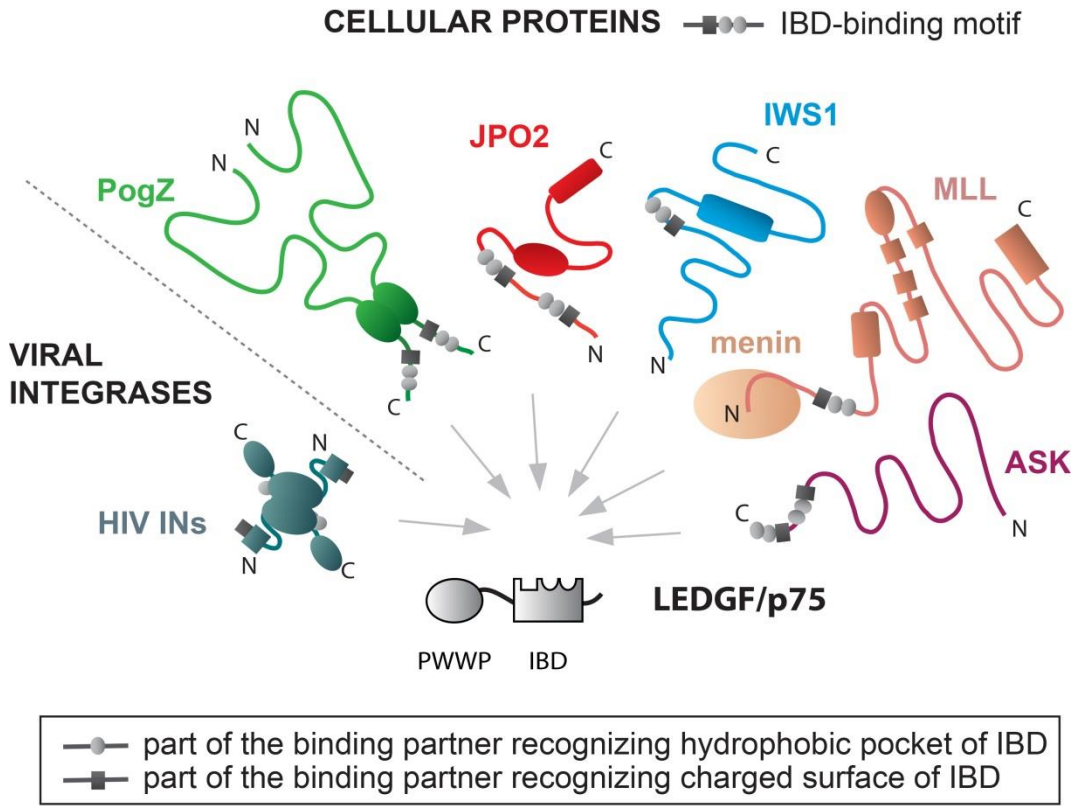


**Figure 61: Schematic representation of LEDGF/p75 IBD interactions.**

Common specific complementary regions in the IBD interactions are schematically depicted as oval and rectangular shapes. In cellular interaction partners of LEDGF/p75 IBD, these regions are part of the IBM. Known structured domains in protein sequences are indicated as larger geometric shapes. C-termini and N-termini of proteins are indicated by letters C and N.

Presence of two IBMs in close vicinity could explain the discrepancy in results describing oligomeric state of LEDGF/p75 in different studies as discussed recently [165]. While Cherepanov *et al.* reported LEDGF/p75 to be strictly monomeric *in vitro* [38], a recent study revealed DNA strands bridging to be mediated by dimers of LEDGF/p75 exclusively [165]. Our findings suggest, that dimerization of LEDGF/p75 could be mediated through its interaction with dimeric binding partners (e.g. PogZ) or binding partners containing two interaction motives (JPO2 or ASK). In contrast, IWS1 or MLL1 dimers have not been described so far. However, some MLL1 fusions were found to induce dimerization of the fusion protein [166]. In the case of HIV IN, a stable tetramer of IN has been shown to associate with two molecules of LEDGF/p75 [30]. Altogether, it would be tempting to speculate that two or more LEDGF/p75 molecules form a functional unit. However, this speculation lacks experimental evidence.

As there are multiple physiological interaction partners of LEDGF/p75 competing for a common interaction interface, the question of regulation arises. Utilization of a single domain for multiple competing protein-protein interactions is not unique among transcription regulators. One such example would be the kinase-inducible domain interacting (KIX) domain which also interacts with a host of intrinsically disordered proteins (reviewed by Thakur *et al.* [167]). In the case of KIX domain, the regulation of its interaction promiscuity is carried out by post-translational modifications and allosteric effects. Such information are lacking in the case of LEDGF/p75 IBD. Although phosphorylation sites in close vicinity of the IBM acidic residues were detected [151], there is no direct evidence that these are involved in regulation of interactions with the IBD.

Along with the regulation of interactions with LEDGF/p75 IBD, physiological roles of the resulting protein complexes are also not clear. It is safe to say that LEDGF/p75 can act as a chromatin tethering factor for all its binding partners. However, it was shown that JPO2 and MLL1 can interact directly with chromatin [57, 58, 78], while IWS1 and PogZ are associated with it through interactions with other proteins [69, 153]. Therefore, it is unclear whether chromatin tethering is the primary function of LEDGF/p75 in these complexes. Physiological roles of IBD interaction partners do not hint at any universal process, although there are certain common features among them. While both PogZ and ASK are implicated in cell-cycle regulation [69, 168], MLL1 and JPO2 are involved in transcription control [58, 83]. As mentioned above, there is a functional link between two of the IBD interaction partners. Along with LEDGF/p75 IBD, menin can also interact with ASK and completely repress CDC7–ASK-induced cell proliferation [112]. Interestingly, both LEDGF/p75 and ASK have

been implicated in DNA damage response [20, 169] and their transcription is regulated by Sp1 promoter elements [15, 170]. Thus far, no dominant defining function of LEDGF/p75 has been proposed and its physiological interactions are considered a way of assuring the required localization in the chromatin context.

Out-competing all the physiological interaction partners of LEDGF/p75, the HIV IN has evolved an alternative way to bind IBD domain and achieve efficient integration into sites of transcriptionally active chromatin. The affinity of HIV IN for IBD is significantly higher than any of the physiological interactors. We have determined affinities of $JPO2_{1–130}$ ($K_d$ = 1.6 μM), $PogZ_{1117–1410}$ ($K_d$ = 1.9 μM), $MLL1_{140–160}$ ($K_d$ = 86.3 μM) and $IWS1_{352–548}$ ($K_d$ = 6.7 μM) to IBD using ITC. The $K_d$ of LEDGF/p75–HIV IN interaction was reported earlier to be 10.9 nM [171]. Although our results might be influenced by using truncated protein constructs, they are in accordance with quantification of LEDGF/p75 complexes in HeLa cells reported by van Nuland *et al.* [3]. Notably, LEDGF/p75–JPO2 complex was found to be 18-fold more abundant than LEDGF/p75–MLL1/MLL2 complexes [3]. In contrast to the cellular partners, the IBD interaction interface of HIV IN is structured and prearranged, which explains the high affinity.

As shown before, HIV IN, JPO2, PogZ, IWS1 and MLL1 compete for interaction with LEDGF/p75. Our results revealed that IBD interaction interfaces for $JPO2_{1-130}$, $IWS1_{352–548}$, $MLL1_{140–160}$ and $PogZ_{1117-1410}$ are almost identical and overlapping with the interface for HIV IN. This universal overlapping surface includes one of the hydrophobic pockets and the basic patch necessary for electrostatic contacts (Figure 61). However, there is an additional interface between MLL1 and IBD maintained by MLL1 aa residues F129 and F133. So far this is unique among all IBD interaction partners and possibility of such additional contacts in other cellular interactors of LEDGF/p75 remains to be tested.

Our results characterizing the IBD interactions with various partners are important in development of inhibitors targeting protein-protein interactions with either HIV IN or MLL1 (reviewed by Christ *et al.* [145] and Cierpicki *et al.* [172]). As discussed before, large part of interaction interface on IBD is common for all the interaction partners. Whether it would be possible to achieve specific inhibition remains speculative. In the case of MLL1, there is an additional unique interface engaged in its interaction with IBD. Such feature could allow for inhibition specificity. However, should the inhibitor target the common interaction interface, it would disturb other physiological interactions of LEDGF/p75 IBD. This was observed in the case of cyclic peptide inhibitor CP65, which disrupted interactions of LEDGF/p75 IBD with JPO2, MLL1 and HIV IN [124]. Interestingly, overexpression of this peptide in HeLaP4

cells did not result in any obvious toxicity. Nonetheless, potent inhibitors of LEDGF/p75–HIV IN interaction were selected to target the HIV IN instead of the genetically more stable LEDGF/p75 IBD (reviewed by Christ *et al.* [145]). However, this approach would be problematic in the case of intrinsically disordered interaction partners of IBD such as MLL1.

# 7. CONCLUSIONS

The general focus of this thesis is on physiological interactions of LEDGF/p75 IBD. We aimed to address the particular main goals as listed in the Aims of the thesis Chapter and obtained the following results:

- We prepared expression vectors encoding protein fragments for recombinant expression in *Escherichia coli*. We designed and cloned numerous expression vectors encoding different regions of these proteins. Expressed protein fragment boundaries were chosen based on numerous *in silico* predictions and on previous studies. Our IBD construct is truncated at the C-terminus in comparison to the fragment used in previous studies [37, 38]. Inclusion of N425 and M426 at the C-terminus dramatically improved protein expression yields.
- We optimized the expression and purification protocols for all protein fragments used in this thesis. Generally, we utilized affinity tags and SEC to reach high protein yield and purity. We expressed and purified following protein fragments: IBD (LEDFG/p75$_{345–426}$), JPO2$_{1–130}$, PogZ$_{1117–1410}$ and MLL1$_{1–160}$.
- We characterized the purified protein fragments using broad range of methods including CD spectroscopy, NMR, analytical SEC, LC-MS and MALDI-TOF. Several lines of evidence led us to conclude that JPO2$_{1-130}$ is mostly intrinsically disordered while PogZ$_{1117–1410}$ forms well-folded dimers with short disordered C-terminal region.
- We successfully reconstituted the IBD–JPO2$_{1-130}$, IBD–PogZ$_{1117–1410}$ and IBD–MLL1$_{1–160}$ complexes. We determined optimal buffer conditions in which the complexes were stable at room temperature during NMR experiments. Affinities and stoichiometries of IBD complexes with the respective binding partners were determined by ITC.
- Structural changes taking place upon interaction with binding partners were confirmed and mapped at an amino acid level. Identification of interaction interfaces in MLL1, JPO2 and PogZ revealed a consensus IBD-binding motif (IBM) responsible for interactions with IBD. This motif is located in intrinsically disordered regions of these interaction partners. We validated the role and importance of the IBM by a comprehensive mutation analysis.
- Based on the knowledge of the IBM, we identified and validated IWS1 as a novel LEDGF/p75 interaction partner. We confirmed the interaction of IWS1 with

LEDGF/p75 by various independent methods including NMR, ITC, AlphaScreen and pull-down assays. The intrinsically disordered character of IWS1 was experimentally established by a comprehensive biophysical characterization.

- We utilized NMR spectroscopy to solve solution structures of IBD–$MLL_{140–160}$ and IBD–$PogZ_{1389–1404}$ complexes. We revealed structural details of IBD–IBM interaction. Critical roles of two phenylalanine residues and acidic residues in the IBM were confirmed by mutation analyses.

We defined a major common LEDGF/p75 IBD interaction interface for JPO2, PogZ, MLL1, IWS1 and HIV IN, which is most likely also utilized by other known LEDGF interaction partners (ASK and MLL2). This represents a challenge for the development of selective inhibitors as they might disrupt off-target physiological interactions. However, there are additional unique interaction features in the case of HIV IN and MLL1, which could allow for inhibition specificity.

# 8. REFENCES


1. Singh, D.P., et al., *Lens epithelium-derived growth factor: effects on growth and survival of lens epithelial cells, keratinocytes, and fibroblasts.* Biochem Biophys Res Commun, 2000. **267**(1): p. 373-81.
2. Eidahl, J.O., et al., *Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes.* Nucleic Acids Res, 2013. **41**(6): p. 3924-36.
3. van Nuland, R., et al., *Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes.* Mol Cell Biol, 2013. **33**(10): p. 2067-77.
4. Hendrix, J., et al., *The transcriptional co-activator LEDGF/p75 displays a dynamic scan-and-lock mechanism for chromatin tethering.* Nucleic Acids Res, 2011. **39**(4): p. 1310-25.
5. Tsutsui, K.M., et al., *Nuclear protein LEDGF/p75 recognizes supercoiled DNA by a novel DNA-binding domain.* Nucleic Acids Res, 2011. **39**(12): p. 5067-81.
6. Maertens, G., et al., *Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75.* J Biol Chem, 2004. **279**(32): p. 33421-9.
7. Turlure, F., et al., *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, 2006. **34**(5): p. 1653-65.
8. Ge, H., Y. Si, and R.G. Roeder, *Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation.* EMBO J, 1998. **17**(22): p. 6723-9.
9. Desfarges, S., et al., *LEDGF/p75 TATA-less promoter is driven by the transcription factor Sp1.* J Mol Biol, 2011. **414**(2): p. 177-93.
10. Cherepanov, P., et al., *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, 2004. **279**(47): p. 48883-92.
11. Singh, D.P., et al., *Lens epithelium-derived growth factor: increased resistance to thermal and oxidative stresses.* Invest Ophthalmol Vis Sci, 1999. **40**(7): p. 1444-51.
12. Cohen, B., et al., *Transcriptional regulation of vascular endothelial growth factor C by oxidative and thermal stress is mediated by lens epithelium-derived growth factor/p75.* Neoplasia, 2009. **11**(9): p. 921-33.
13. Singh, D.P., et al., *LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes.* Biochem Biophys Res Commun, 2001. **283**(4): p. 943-55.
14. Fatma, N., et al., *Transcriptional regulation of the antioxidant protein 2 gene, a thiol-specific antioxidant, by lens epithelium-derived growth factor to protect cells from oxidative stress.* J Biol Chem, 2001. **276**(52): p. 48899-907.

15. Singh, D.P., et al., *Transcriptional protein Sp1 regulates LEDGF transcription by directly interacting with its cis-elements in GC-rich region of TATA-less gene promoter.* PLoS One, 2012. **7**(5): p. e37012.
16. Sutherland, H.G., et al., *Disruption of Ledgf/Psip1 results in perinatal mortality and homeotic skeletal transformations.* Mol Cell Biol, 2006. **26**(19): p. 7201-10.
17. Pradeepa, M.M., et al., *Psip1/Ledgf p75 restrains Hox gene expression by recruiting both trithorax and polycomb group proteins.* Nucleic Acids Res, 2014. **42**(14): p. 9021-32.
18. Mallo, M., D.M. Wellik, and J. Deschamps, *Hox genes and regional patterning of the vertebrate body plan.* Dev Biol, 2010. **344**(1): p. 7-15.
19. Alharbi, R.A., et al., *The role of HOX genes in normal hematopoiesis and acute leukemia.* Leukemia, 2013. **27**(5): p. 1000-8.
20. Daugaard, M., et al., *LEDGF (p75) promotes DNA-end resection and homologous recombination.* Nat Struct Mol Biol, 2012. **19**(8): p. 803-10.
21. Yokoyama, A. and M.L. Cleary, *Menin critically links MLL proteins with LEDGF on cancer-associated target genes.* Cancer Cell, 2008. **14**(1): p. 36-46.
22. Bartholomeeusen, K., et al., *Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ.* J Biol Chem, 2009. **284**(17): p. 11467-77.
23. Maertens, G.N., P. Cherepanov, and A. Engelman, *Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin.* J Cell Sci, 2006. **119**(Pt 12): p. 2563-71.
24. Hughes, S., et al., *Transcriptional co-activator LEDGF interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity.* J Biol Chem, 2010. **285**(1): p. 541-54.
25. De Rijck, J., et al., *High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region.* Nucleic Acids Res, 2010. **38**(18): p. 6135-47.
26. Maertens, G., et al., *LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells.* J Biol Chem, 2003. **278**(35): p. 33528-39.
27. Gijsbers, R., et al., *Role of the PWWP domain of lens epithelium-derived growth factor (LEDGF)/p75 cofactor in lentiviral integration targeting.* J Biol Chem, 2011. **286**(48): p. 41812-25.
28. Ciuffi, A., et al., *A role for LEDGF/p75 in targeting HIV DNA integration.* Nat Med, 2005. **11**(12): p. 1287-9.
29. Llano, M., et al., *Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase.* J Biol Chem, 2004. **279**(53): p. 55570-7.
30. Cherepanov, P., et al., *HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells.* J Biol Chem, 2003. **278**(1): p. 372-81.
31. Cherepanov, P., *LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro.* Nucleic Acids Res, 2007. **35**(1): p. 113-24.

32. Llano, M., et al., *Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75.* J Mol Biol, 2006. **360**(4): p. 760-73.

33. Shun, M.C., et al., *LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration.* Genes Dev, 2007. **21**(14): p. 1767-78.

34. Schrijvers, R., et al., *LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGINs.* PLoS Pathog, 2012. **8**(3): p. e1002558.

35. Llano, M., et al., *An essential role for LEDGF/p75 in HIV integration.* Science, 2006. **314**(5798): p. 461-4.

36. Christ, F., et al., *Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication.* Nat Chem Biol, 2010. **6**(6): p. 442-8.

37. Cherepanov, P., et al., *Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17308-13.

38. Cherepanov, P., et al., *Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75.* Nat Struct Mol Biol, 2005. **12**(6): p. 526-32.

39. Hare, S., et al., *A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75.* PLoS Pathog, 2009. **5**(1): p. e1000259.

40. Sapoznik, S., et al., *Gonadotropin-regulated lymphangiogenesis in ovarian cancer is mediated by LEDGF-induced expression of VEGF-C.* Cancer Res, 2009. **69**(24): p. 9306-14.

41. Basu, A., et al., *Expression of the stress response oncoprotein LEDGF/p75 in human cancer: a study of 21 tumor types.* PLoS One, 2012. **7**(1): p. e30132.

42. Xu, X., et al., *Human MCS5A1 candidate breast cancer susceptibility gene FBXO10 is induced by cellular stress and correlated with lens epithelium-derived growth factor (LEDGF).* Mol Carcinog, 2012.

43. Hussey, D.J., et al., *Fusion of the NUP98 gene with the LEDGF/p52 gene defines a recurrent acute myeloid leukemia translocation.* BMC Genet, 2001. **2**: p. 20.

44. Daugaard, M., et al., *Lens epithelium-derived growth factor is an Hsp70-2 regulated guardian of lysosomal stability in human cancer.* Cancer Res, 2007. **67**(6): p. 2559-67.

45. Huang, T.S., et al., *LEDGF/p75 has increased expression in blasts from chemotherapy-resistant human acute myelogenic leukemia patients and protects leukemia cells from apoptosis in vitro.* Mol Cancer, 2007. **6**: p. 31.

46. Wu, X., et al., *Caspase cleavage of the nuclear autoantigen LEDGF/p75 abrogates its pro-survival function: implications for autoimmunity in atopic disorders.* Cell Death Differ, 2002. **9**(9): p. 915-25.

47. Brown-Bryan, T.A., et al., *Alternative splicing and caspase-mediated cleavage generate antagonistic variants of the stress oncoprotein LEDGF/p75.* Mol Cancer Res, 2008. **6**(8): p. 1293-307.

48. Bueno, M.T., et al., *SUMOylation of the lens epithelium-derived growth factor/p75 attenuates its transcriptional activity on the heat shock protein 27 promoter.* J Mol Biol, 2010. **399**(2): p. 221-39.

49. Miyara, M., et al., *Clinical phenotypes of patients with anti-DFS70/LEDGF antibodies in a routine ANA referral cohort.* Clin Dev Immunol, 2013. **2013**: p. 703759.

50. Ochs, R.L., et al., *Autoantibodies to DFS 70 kd/transcription coactivator p75 in atopic dermatitis and other conditions.* J Allergy Clin Immunol, 2000. **105**(6 Pt 1): p. 1211-20.

51. Mahler, M., et al., *Anti-DFS70/LEDGF antibodies are more prevalent in healthy individuals compared to patients with systemic autoimmune rheumatic diseases.* J Rheumatol, 2012. **39**(11): p. 2104-10.

52. Ganapathy, V. and C.A. Casiano, *Autoimmunity to the nuclear autoantigen DFS70 (LEDGF): what exactly are the autoantibodies trying to tell us?* Arthritis Rheum, 2004. **50**(3): p. 684-8.

53. Shinohara, T., D.P. Singh, and N. Fatma, *LEDGF, a survival factor, activates stress-related genes.* Prog Retin Eye Res, 2002. **21**(3): p. 341-58.

54. Huang, A., et al., *Identification of a novel c-Myc protein interactor, JPO2, with transforming activity in medulloblastoma cells.* Cancer Res, 2005. **65**(13): p. 5607-19.

55. Tian, Y., et al., *CDCA7L promotes hepatocellular carcinoma progression by regulating the cell cycle.* Int J Oncol, 2013. **43**(6): p. 2082-90.

56. Bartholomeeusen, K., et al., *Differential interaction of HIV-1 integrase and JPO2 with the C terminus of LEDGF/p75.* J Mol Biol, 2007. **372**(2): p. 407-21.

57. Chen, K., et al., *R1, a novel repressor of the human monoamine oxidase A.* J Biol Chem, 2005. **280**(12): p. 11552-9.

58. Chen, K., et al., *Transcription factor E2F-associated phosphoprotein (EAPP), RAM2/CDCA7L/JPO2 (R1), and simian virus 40 promoter factor 1 (Sp1) cooperatively regulate glucocorticoid activation of monoamine oxidase B.* Mol Pharmacol, 2011. **79**(2): p. 308-17.

59. Ou, X.M., K. Chen, and J.C. Shih, *Glucocorticoid and androgen activation of monoamine oxidase A is regulated differently by R1 and Sp1.* J Biol Chem, 2006. **281**(30): p. 21512-25.

60. Ou, X.M., K. Chen, and J.C. Shih, *Monoamine oxidase A and repressor R1 are involved in apoptotic signaling pathway.* Proc Natl Acad Sci U S A, 2006. **103**(29): p. 10923-8.

61. Hendrix, J., et al., *Dynamics of the ternary complex formed by c-Myc interactor JPO2, transcriptional co-activator LEDGF/p75, and chromatin.* J Biol Chem, 2014. **289**(18): p. 12494-506.

62. Seki, N., et al., *Characterization of cDNA clones in size-fractionated cDNA libraries from human brain.* DNA Res, 1997. **4**(5): p. 345-9.

63. Gunther, M., M. Laithier, and O. Brison, *A set of proteins interacting with transcription factor Sp1 identified in a two-hybrid screening.* Mol Cell Biochem, 2000. **210**(1-2): p. 131-42.

64. Dyda, F., et al., *Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981-6.

65. Vermeulen, M., et al., *Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers.* Cell, 2010. **142**(6): p. 967-80.

66. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

67. Probst, A.V. and G. Almouzni, *Heterochromatin establishment in the context of genome-wide epigenetic reprogramming.* Trends Genet, 2011. **27**(5): p. 177-85.

68. Rosnoblet, C., et al., *Analysis of the human HP1 interactome reveals novel binding partners.* Biochem Biophys Res Commun, 2011. **413**(2): p. 206-11.

69. Nozawa, R.S., et al., *Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation.* Nat Cell Biol, 2010. **12**(7): p. 719-27.

70. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-20.

71. Fischer, T., et al., *Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast.* Proc Natl Acad Sci U S A, 2009. **106**(22): p. 8998-9003.

72. Fukai, R., et al., *A case of autism spectrum disorder arising from a de novo missense mutation in POGZ.* J Hum Genet, 2015. **60**(5): p. 277-9.

73. Deciphering Developmental Disorders, S., *Large-scale discovery of novel genetic causes of developmental disorders.* Nature, 2015. **519**(7542): p. 223-8.

74. Zeleznik-Le, N.J., A.M. Harden, and J.D. Rowley, *11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10610-4.

75. Birke, M., et al., *The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation.* Nucleic Acids Res, 2002. **30**(4): p. 958-65.

76. Muntean, A.G., et al., *The PHD fingers of MLL block MLL fusion protein-mediated transformation.* Blood, 2008. **112**(12): p. 4690-3.

77. Ernst, P., et al., *MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein.* Mol Cell Biol, 2001. **21**(7): p. 2249-58.

78. Milne, T.A., et al., *MLL targets SET domain methyltransferase activity to Hox gene promoters.* Mol Cell, 2002. **10**(5): p. 1107-17.

79. Yokoyama, A., et al., *Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties.* Blood, 2002. **100**(10): p. 3710-8.

80. Hsieh, J.J., et al., *Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization.* Mol Cell Biol, 2003. **23**(1): p. 186-94.

81. Xia, Z.B., et al., *MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8342-7.

82. Bernstein, B.E., et al., *Genomic maps and comparative analysis of histone modifications in human and mouse.* Cell, 2005. **120**(2): p. 169-81.

83. McMahon, K.A., et al., *Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal.* Cell Stem Cell, 2007. **1**(3): p. 338-45.

84. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. **13**(8): p. 713-9.

85. Yu, B.D., et al., *MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10632-6.

86. Krivtsov, A.V. and S.A. Armstrong, *MLL translocations, histone modifications and leukaemia stem-cell development.* Nat Rev Cancer, 2007. **7**(11): p. 823-33.

87. Domer, P.H., et al., *Molecular analysis of 13 cases of MLL/11q23 secondary acute leukemia and identification of topoisomerase II consensus-binding sequences near the chromosomal breakpoint of a secondary leukemia with the t(4;11).* Leukemia, 1995. **9**(8): p. 1305-12.

88. Meyer, C., et al., *The MLL recombinome of acute leukemias in 2013.* Leukemia, 2013. **27**(11): p. 2165-76.

89. Krivtsov, A.V., et al., *H3K79 methylation profiles define murine and human MLL-AF4 leukemias.* Cancer Cell, 2008. **14**(5): p. 355-68.

90. Richardson, C. and M. Jasin, *Frequent chromosomal translocations induced by DNA double-strand breaks.* Nature, 2000. **405**(6787): p. 697-700.

91. Giri, D., et al., *Multiple endocrine neoplasia syndromes 1 and 2: manifestations and management in childhood and adolescence.* Arch Dis Child, 2015.

92. Bassett, J.H., et al., *Characterization of mutations in patients with multiple endocrine neoplasia type 1.* Am J Hum Genet, 1998. **62**(2): p. 232-44.

93. Milne, T.A., et al., *Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 749-54.

94. Yokoyama, A., et al., *The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis.* Cell, 2005. **123**(2): p. 207-18.

95. Mereau, H., et al., *Impairing MLL-fusion gene-mediated transformation by dissecting critical interactions with the lens epithelium-derived growth factor (LEDGF/p75).* Leukemia, 2013. **27**(6): p. 1245-53.

96. Huang, J., et al., *The same pocket in menin binds both MLL and JUND but has opposite effects on transcription.* Nature, 2012. **482**(7386): p. 542-6.

97. He, S., et al., *High-affinity small-molecule inhibitors of the menin-mixed lineage leukemia (MLL) interaction closely mimic a natural protein-protein interaction.* J Med Chem, 2014. **57**(4): p. 1543-56.

98. Dreijerink, K.M., et al., *Menin links estrogen receptor activation to histone H3K4 trimethylation.* Cancer Res, 2006. **66**(9): p. 4929-35.

99. Johnston, L.H. and A.P. Thomas, *The isolation of new DNA synthesis mutants in the yeast Saccharomyces cerevisiae.* Mol Gen Genet, 1982. **186**(3): p. 439-44.

100. Masai, H. and K. Arai, *Dbf4 motifs: conserved motifs in activation subunits for Cdc7 kinases essential for S-phase.* Biochem Biophys Res Commun, 2000. **275**(1): p. 228-32.

101. Hughes, S., et al., *Crystal structure of human CDC7 kinase in complex with its activator DBF4.* Nat Struct Mol Biol, 2012. **19**(11): p. 1101-7.

102. Jones, D.R., et al., *The Dbf4 motif C zinc finger promotes DNA replication and mediates resistance to genotoxic stress.* Cell Cycle, 2010. **9**(10): p. 2018-26.

103. Labib, K., *How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells?* Genes Dev, 2010. **24**(12): p. 1208-19.

104. Matthews, L.A. and A. Guarne, *Dbf4: the whole is greater than the sum of its parts.* Cell Cycle, 2013. **12**(8): p. 1180-8.

105. Kim, J.M., et al., *Cdc7 kinase mediates Claspin phosphorylation in DNA replication checkpoint.* Oncogene, 2008. **27**(24): p. 3475-82.

106. Rainey, M.D., et al., *Cdc7-dependent and -independent phosphorylation of Claspin in the induction of the DNA replication checkpoint.* Cell Cycle, 2013. **12**(10): p. 1560-8.

107. Yamashita, N., et al., *Functional analyses of mouse ASK, an activation subunit for Cdc7 kinase, using conditional ASK knockout ES cells.* Genes Cells, 2005. **10**(6): p. 551-63.

108. Kim, J.M., et al., *Inactivation of Cdc7 kinase in mouse ES cells results in S-phase arrest and p53-dependent cell death.* EMBO J, 2002. **21**(9): p. 2168-79.

109. Oshiro, G., et al., *Cell cycle control of Cdc7p kinase activity through regulation of Dbf4p stability.* Mol Cell Biol, 1999. **19**(7): p. 4888-96.

110. Bonte, D., et al., *Cdc7-Dbf4 kinase overexpression in multiple cancers and tumor cell lines is correlated with p53 inactivation.* Neoplasia, 2008. **10**(9): p. 920-31.

111. Montagnoli, A., J. Moll, and F. Colotta, *Targeting cell division cycle 7 kinase: a new approach for cancer therapy.* Clin Cancer Res, 2010. **16**(18): p. 4503-8.

112. Schnepp, R.W., et al., *Functional interaction between tumor suppressor menin and activator of S-phase kinase.* Cancer Res, 2004. **64**(18): p. 6791-6.

113. Yuan, Y.W. and S.R. Wessler, *The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies.* Proc Natl Acad Sci U S A, 2011. **108**(19): p. 7884-9.

114. Li, X., et al., *Structural biology of retroviral DNA integration.* Virology, 2011. **411**(2): p. 194-205.

115. Fujiwara, T. and K. Mizuuchi, *Retroviral DNA integration: structure of an integration intermediate.* Cell, 1988. **54**(4): p. 497-504.

116. Krishnan, L. and A. Engelman, *Retroviral integrase proteins and HIV-1 DNA integration.* J Biol Chem, 2012. **287**(49): p. 40858-66.

117. Hazuda, D.J., et al., *Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells.* Science, 2000. **287**(5453): p. 646-50.

118. Steigbigel, R.T., et al., *Raltegravir with optimized background therapy for resistant HIV-1 infection.* N Engl J Med, 2008. **359**(4): p. 339-54.

119. Mesplede, T. and M.A. Wainberg, *Resistance against Integrase Strand Transfer Inhibitors and Relevance to HIV Persistence.* Viruses, 2015. **7**(7): p. 3703-18.

120. Vandekerckhove, L., et al., *Transient and stable knockdown of the integrase cofactor LEDGF/p75 reveals its role in the replication cycle of human immunodeficiency virus.* J Virol, 2006. **80**(4): p. 1886-96.

121. De Rijck, J., et al., *Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication.* J Virol, 2006. **80**(23): p. 11498-509.

122. Tsiang, M., et al., *New class of HIV-1 integrase (IN) inhibitors with a dual mode of action.* J Biol Chem, 2012. **287**(25): p. 21189-203.

123. Christ, F., et al., *Small-molecule inhibitors of the LEDGF/p75 binding site of integrase block HIV replication and modulate integrase multimerization.* Antimicrob Agents Chemother, 2012. **56**(8): p. 4365-74.

124. Desimmie, B.A., et al., *Phage display-directed discovery of LEDGF/p75 binding cyclic peptide inhibitors of HIV replication.* Mol Ther, 2012. **20**(11): p. 2064-75.

125. Stols, L., et al., *A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site.* Protein Expr Purif, 2002. **25**(1): p. 8-15.

126. Busso, D., B. Delagoutte-Busso, and D. Moras, *Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli.* Anal Biochem, 2005. **343**(2): p. 313-21.

127. Chiu, J., et al., *Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h.* Nucleic Acids Res, 2004. **32**(21): p. e174.

128. Mladkova, J., et al., *Phenotyping breast cancer cell lines EM-G3, HCC1937, MCF7 and MDA-MB-231 using 2-D electrophoresis and affinity chromatography for glutathione-binding proteins.* BMC Cancer, 2010. **10**: p. 449.

129. Selicharova, I., et al., *2-DE analysis of breast cancer cell lines 1833 and 4175 with distinct metastatic organ-specific potentials: comparison with parental cell line MDA-MB-231.* Oncol Rep, 2008. **19**(5): p. 1237-44.

130. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Anal Biochem, 2006. **357**(2): p. 289-98.

131. Kopec, J. and G. Schneider, *Comparison of fluorescence and light scattering based methods to assess formation and stability of protein-protein complexes.* J Struct Biol, 2011. **175**(2): p. 216-23.

132. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods Mol Biol, 2004. **278**: p. 353-78.

133. Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR, 2009. **44**(4): p. 213-23.

134. Herrmann, T., P. Guntert, and K. Wuthrich, *Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.* J Mol Biol, 2002. **319**(1): p. 209-27.

135. Doreleijers, J.F., et al., *CING: an integrated residue-based structure validation program suite.* J Biomol NMR, 2012. **54**(3): p. 267-83.

136. Richardson, J.M., et al., *Mechanism of Mos1 transposition: insights from structural analysis.* EMBO J, 2006. **25**(6): p. 1324-34.

137. Cermakova, K., et al., *Validation and structural characterization of the LEDGF/p75-MLL interface as a new target for the treatment of MLL-dependent leukemia.* Cancer Res, 2014. **74**(18): p. 5139-51.

138. Prilusky, J., et al., *FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded.* Bioinformatics, 2005. **21**(16): p. 3435-8.

139. Xue, B., et al., *PONDR-FIT: a meta-predictor of intrinsically disordered amino acids.* Biochim Biophys Acta, 2010. **1804**(4): p. 996-1010.

140. de Castro, E., et al., *ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W362-5.

141. Yoh, S.M., J.S. Lucas, and K.A. Jones, *The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation.* Genes Dev, 2008. **22**(24): p. 3422-34.

142. Pradeepa, M.M., et al., *Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing.* PLoS Genet, 2012. **8**(5): p. e1002717.

143. Kolasinska-Zwierz, P., et al., *Differential chromatin marking of introns and expressed exons by H3K36me3.* Nat Genet, 2009. **41**(3): p. 376-81.

144. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.

145. Christ, F. and Z. Debyser, *The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy.* Virology, 2013. **435**(1): p. 102-9.

146. Grembecka, J., et al., *Molecular basis of the mixed lineage leukemia-menin interaction: implications for targeting mixed lineage leukemias.* J Biol Chem, 2010. **285**(52): p. 40690-8.

147. Shi, A., et al., *Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia.* Blood, 2012. **120**(23): p. 4461-9.

148. Grembecka, J., et al., *Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia.* Nat Chem Biol, 2012. **8**(3): p. 277-84.

149. Cavalluzzo, C., et al., *Identification of small peptides inhibiting the integrase-LEDGF/p75 interaction through targeting the cellular co-factor.* J Pept Sci, 2013. **19**(10): p. 651-8.

150. Murai, M.J., et al., *The same site on the integrase-binding domain of lens epithelium-derived growth factor is a therapeutic target for MLL leukemia and HIV.* Blood, 2014. **124**(25): p. 3730-7.

151. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D261-70.

152. Koltun, E.S., et al., *Discovery of XL413, a potent and selective CDC7 inhibitor.* Bioorg Med Chem Lett, 2012. **22**(11): p. 3727-31.

153. Yoh, S.M., et al., *The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export.* Genes Dev, 2007. **21**(2): p. 160-74.

154. Kim, S., et al., *Pre-mRNA splicing is a determinant of histone H3K36 methylation.* Proc Natl Acad Sci U S A, 2011. **108**(33): p. 13564-9.

155. de Almeida, S.F., et al., *Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36.* Nat Struct Mol Biol, 2011. **18**(9): p. 977-83.

156. Luco, R.F., et al., *Regulation of alternative splicing by histone modifications.* Science, 2010. **327**(5968): p. 996-1000.

157. Huff, J.T., et al., *Reciprocal intronic and exonic histone modification regions in humans.* Nat Struct Mol Biol, 2010. **17**(12): p. 1495-9.

158. Morchikh, M., et al., *TOX4 and NOVA1 proteins are partners of the LEDGF PWWP domain and affect HIV-1 replication.* PLoS One, 2013. **8**(11): p. e81217.

159. Van Roey, K., et al., *Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation.* Chem Rev, 2014. **114**(13): p. 6733-78.

160. Wright, P.E. and H.J. Dyson, *Linking folding and binding.* Curr Opin Struct Biol, 2009. **19**(1): p. 31-8.

161. Tompa, P. and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions.* Trends Biochem Sci, 2008. **33**(1): p. 2-8.

162. Dunker, A.K., et al., *Intrinsically disordered proteins and multicellular organisms.* Semin Cell Dev Biol, 2014.

163. Fukuchi, S., et al., *Development of an accurate classification system of proteins into structured and unstructured regions that uncovers novel structural domains: its application to human transcription factors.* BMC Struct Biol, 2009. **9**: p. 26.

164. Via, A., et al., *How pathogens use linear motifs to perturb host cell networks.* Trends Biochem Sci, 2015. **40**(1): p. 36-48.

165. Vanderlinden, W., et al., *Structure, mechanics, and binding mode heterogeneity of LEDGF/p75-DNA nucleoprotein complexes revealed by scanning force microscopy.* Nanoscale, 2014.

166. So, C.W., et al., *Dimerization contributes to oncogenic activation of MLL chimeras in acute leukemias.* Cancer Cell, 2003. **4**(2): p. 99-110.

167. Thakur, J.K., A. Yadav, and G. Yadav, *Molecular recognition by the KIX domain and its role in gene regulation.* Nucleic Acids Res, 2014. **42**(4): p. 2112-25.

168. Kitamura, R., et al., *Molecular mechanism of activation of human Cdc7 kinase: bipartite interaction with Dbf4/activator of S phase kinase (ASK) activation subunit stimulates ATP binding and substrate recognition.* J Biol Chem, 2011. **286**(26): p. 23031-43.

169. Lee, A.Y., et al., *Dbf4 is direct downstream target of ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related (ATR) protein to regulate intra-S-phase checkpoint.* J Biol Chem, 2012. **287**(4): p. 2531-43.

170. Wu, X. and H. Lee, *Human Dbf4/ASK promoter is activated through the Sp1 and MluI cell-cycle box (MCB) transcription elements.* Oncogene, 2002. **21**(51): p. 7786-96.

171. Tsiang, M., et al., *Affinities between the binding partners of the HIV-1 integrase dimer-lens epithelium-derived growth factor (IN dimer-LEDGF) complex.* J Biol Chem, 2009. **284**(48): p. 33580-99.

172. Cierpicki, T. and J. Grembecka, *Challenges and opportunities in targeting the menin-MLL interaction.* Future Med Chem, 2014. **6**(4): p. 447-62.