Charles University in Prague

Faculty of Science

# Structural insights into LEDGF/p75 interactome

Ing. Václav Veverka, Ph.D.

HABILITATION THESIS

Prague 2021

## Table of Contents

## Abstract

Lens epithelium-derived growth factor, also known as LEDGF/p75 or PSIP1, is a key tethering factor for viral DNA integration into the host genome that also plays an important role in “mixed lineage” acute leukemia development. Due to its modular domain composition, LEDGF/p75 recognizes a specific covalent histone modification that marks chromatin regions of actively transcribed genes, and it is thus capable of targeting various cellular partners to these areas. The molecular mechanism of interaction between LEDGF/p75 and its physiological partners was not fully understood when we initiated this study. Using a combination of structural biology, biophysics, biochemistry and cell biology, we uncovered molecular features crucial for LEDGF/p75 interaction with its binding partners. In particular, we found that this interaction is achieved through a structurally conserved binding mode common to all of the LEDGF/p75 binding partners. This discovery led to the identification of previously unknown direct interactions between LEDGF/75 and other major transcriptional regulatory factors. This suggests that LEDGF/p75 participates in a much larger network of factors involved in transcription elongation than what was previously recognized. In addition, our data also revealed that binding between LEDGF/p75 and its interaction partners is strongly modulated by casein kinase 2-depenedent phosphorylation. We characterized in detail the mechanism of LEDGF/p75 dimerization that contributes to the regulation of LEDGF/p75 interactome. Our work contributed to the validation of LEDGF/p75 as a potential target for therapeutic intervention against “mixed lineage” leukemia, and we are currently actively participating in a drug discovery program. The results summarized in this thesis have raised a number of interesting questions that will shape our future research in LEDGF-related biology.

## Acknowledgements

Choosing a career in science was a fantastic decision, and I do enjoy every moment of it. It could not have been possible without inspiring mentors, wonderful colleagues and supportive friends. I am truly indebted to my PhD supervisor Richard Hrabal. I joined Richard the very moment he established his biomolecular NMR lab at UCT Prague. He introduced me to the NMR field, and I still heavily rely on the knowledge that I acquired during my PhD project. Trained as a chemist, I needed to embrace biochemistry. This brought me to IOCB Prague for the first time, where I learned to prepare and handle proteins in the Iva Pichová lab. This experience was instrumental in helping me to solve my first NMR structure and to earn my PhD. For my postdoctoral training, I joined Mark Carr at the University of Leicester to work on a collaborative project with the pharmaceutical industry. My goal was to validate NMR-based structural biology as a tool for drug discovery. This gave me a unique opportunity to directly contribute to the development of several drug candidates and to closely interact with Alistair Henry and Richard Taylor, the passionate scientists who led the research at UCB/Celltech. I then returned to IOCB Prague to establish a biomolecular NMR core facility and to pursue my independent research ideas. I am particularly grateful to Pavlína Řezáčová and Zdeněk Hostomský for their continuous support. I would like also to thank Jan Černý, who gave me the opportunity to join the Department of Cell Biology as well as allowing me to open new courses in Structural Biology for Charles University students. I have been lucky to find talented and hardworking postdocs, students and research scientists that have devoted their time and effort to the LEDGF-centric portfolio of projects: Káťa Čermáková, Vanda Lux, Rozálie Hexnerová, Eliška Koutná, Lisa-Maria Weinhold, Marcela Mádlíková, Magdalena Hořejší, Karolína Naušová, Markéta Hašplová, Majda Havlová, Milan Fábry, Pavel Srb, Lukáš Vrzal and Matúš Drexler. I would also like to thank my key collaborators Zeger Debyser (KU Leuven) and Courtney Hodges (Baylor College of Medicine) for their eagerness to work together and for inspiring me in new areas of biology. Special thanks belong to my family, friends and colleagues from all around the globe and Prague 6. Most of all, I want to thank to my supportive partner Káťa for her contagious enthusiasm and for bringing so much joy in my life.

# 1. Introduction

Structural biology has become invaluable in gaining a mechanistic understanding of complex biological processes with the use of X-ray crystallography complemented by nuclear magnetic resonance spectroscopy (NMR) and single particle cryo-electron microscopy (cryo-EM). Over the last few decades, it has gradually moved from a position of providing purely descriptive perception to hypothesis driven research. This could not have been possible without a close interplay with cell biology which has allowed for a two-way transfer of knowledge that has provided a much broader insight into the biological function of the studied systems (**Figure 1**).

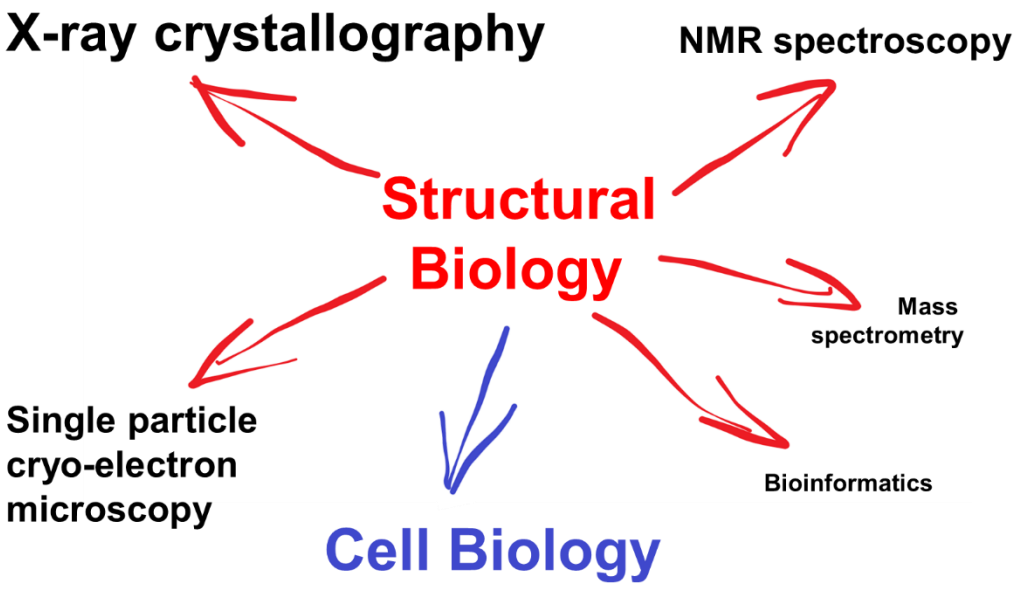


**Figure 1:** From a current perspective, structural biology is no longer limited to X-ray crystallography and NMR spectroscopy, but benefits from increasingly more important single particle cryo-electron microscopy and is often supported by mass spectrometry and computational techniques. An integration of structural and cell biology provides a strong basis for hypothesis driven research.

This thesis is a collection of four scientific papers, focused on biological roles of an epigenetic reader lens epithelium-derived growth factor/p75 (LEDGF/p75), that vastly benefited from a coordinated utilization of structural and cell biology. The following introductory chapters provide information on **LEDGF/p75 background** in the context of **epigenome** that is often maintained by **transient biomolecular interactions**. Attention is also paid to **NMR spectroscopy** as a versatile tool for studying of such dynamic interactions.

## 1.1. Epigenome

Genomic DNA is stored in the nucleus wrapped around the histone octamers in the form of nucleosomes that further associate with DNA-binding factors and RNA molecules into a higher

order structure, known as chromatin. N-terminal histone tails project out of the relatively compact nucleosomal core, and are subjected to various covalent posttranslational modifications that, in combination with DNA modifications, define the epigenetic code. The dynamic modifications of this code, described as epigenome, shape the flow of information from the genome to the proteome and allow eukaryotic cells to maintain their phenotype or respond to various environmental cues as well as disease states. Overall, the epigenome affects local chromatin structure and dynamics, and consequently determines the accessibility of specific DNA loci. In addition, epigenetic modifications represent a binding scaffold for transcriptional activators or repressors that modulate gene expression. DNA can by covalently modified by methylation, hydroxymethylation, formylation or carboxylation, while histone tail modifications include methylation, acetylation, phosphorylation, sumoylation, ubiquitylation and other less populated marks (Dawson and Kouzarides, 2012). Enzymes that modify specific histone residues or nucleotides are called “writers”, enzymes with the capacity to remove these marks are called “erasers”, and factors that bind chromatin modifications and therefore are able to sense the chromatin state of a given locus are generally referred to as “readers” (**Figure 2**). The distribution of epigenetic modifications marks the local chromatin state and efficiently regulate transcription, ensuring the desired gene expression pattern for any cell. The combination of various marks defines over fifty different chromatin states in humans (Ernst and Kellis, 2010).

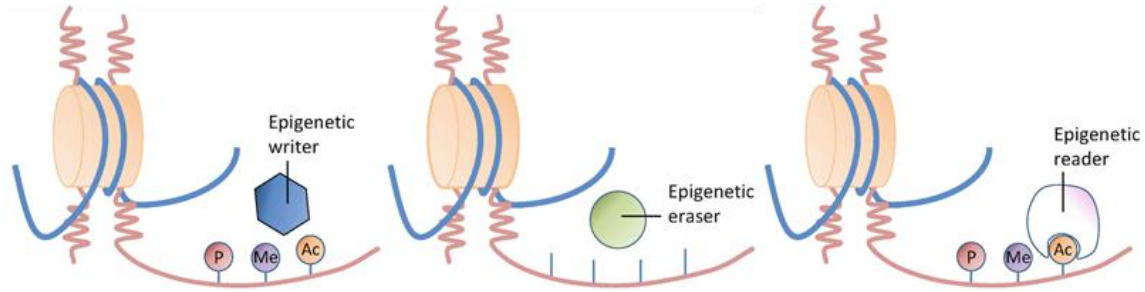


**Figure 2:** Epigenetic writes, erasers and readers - epigenetic modifiers that are responsible for installing (DNA/arginine/lysine methyltransferases – DMT/RMT/KMTs, histone acetyltransferases – HAT/KATs, etc.), removing (histone deacetylases -– HDACs or lysine demethylases – KDMs), and recognizing (Tudor, chromodomain – CHD, bromodomain – BRD, plant homeodomain – PHD, etc.) posttranslational modifications. Adapted from (Duncan and Campbell, 2018).

Histone lysines can be methylated at multiple positions, and the deposition of these marks leads to diverse transcriptional outcomes (Bannister et al., 2002; Greer and Shi, 2012). The complexity of this type of modification is increased by a number of methyl groups attached to lysine side-chains that can be mono- (me1), di- (me2) or trimethylated (me3), which provides an additional layer of information that determines the final epigenetic outcome. Active chromatin is

marked by lysine methylation of histone at positions 4, 26 and 79, while the repressed chromatin states are associated with methylations at positions 9, 27 and 20 (Greer and Shi, 2012).

One of these modifications, trimethylation of lysine at position 36 on histone 3 (H3K36me3), is an abundant chromatin mark, specifically enriched at gene bodies of actively transcribed genes as well as at centromeric regions. The association of the H3K36me3 mark with actively transcribed genes is conserved from yeast to human, which underlines its importance for correct cellular function (Wagner and Carpenter, 2012). The actual regulatory activities include transcription elongation, cryptic start sites prevention and pre-mRNA splicing/processing (Kim et al., 2011; Kolasinska-Zwierz et al., 2009; Krogan et al., 2003; Neri et al., 2017). An additional role for H3K36me3 has been linked with the recruitment of DNA repair machinery to mismatch regions (Li et al., 2013). As a consequence, there is an enhanced DNA damage protection activity found in H3K36me3-marked actively transcribed genes.

H3K36 mono- and dimethylation marks in human cells are deposited by eight distinct histone methyltransferases; however, the trimethylation mark can only be deposited by Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain-containing protein 2 (SETD2) (Wagner and Carpenter, 2012). The methylation mark can be removed either enzymatically by Jumonji domain-containing histone demethylase protein 1 and 3 (JHDM1 and JHDM3) family members or by histone turnover that is maintained by ATP-dependent chromatin remodeling complexes and histone chaperones (Hyun et al., 2017). In addition to the “writers” and “erasers” of this mark, H3K36 trimethylation is recognized by a range of the “reader” proteins harboring a domain from the Royal superfamily (Maurer-Stroh et al., 2003), such as transcription regulators or factors involved in DNA damage response.

## 1.2. Lens epithelium-derived growth factor (LEDGF)

One of the epigenetic readers of H3K36 methylation marks is encoded by the PC4- and SFRS1-interacting protein 1 (*PSIP1*) gene that leads to expression of two alternatively spliced proteins, p52 and p75 (**Figure 3**), also known as lens epithelium-derived growth factor (LEDGF/p52 and LEDGF/p75). Both proteins belong to the hepatoma-derived growth factor (HDGF)-related protein family of nuclear proteins that are distinguished by the presence of the N-terminal Pro-Trp-Trp-Pro (PWWP)-domain responsible for actual mark recognition (residues 1-91). The longer p75 variant carries an additional C-terminal protein interaction module, known as an integrase-binding domain (IBD), that is responsible for the ability to tether lentiviral integrase to actively transcribed genes (Ciuffi et al., 2005; Maertens et al., 2003). The detailed insight into the tethering process that relies on direct interaction between the LEDGF/p75 IDB and HIV-1 integrase catalytic core

was revealed by X-ray crystallography (**Figure 4**) (Cherepanov et al., 2005). This protein-protein interaction was subsequently exploited in several drug-discovery campaigns that yielded a novel class of potential anti-HIV therapeutic compounds targeting non-catalytic sites (Christ et al., 2010; Jurado et al., 2013). The small molecule inhibitors block the IBD interaction site on the surface of the HIV-1 integrase dimer.

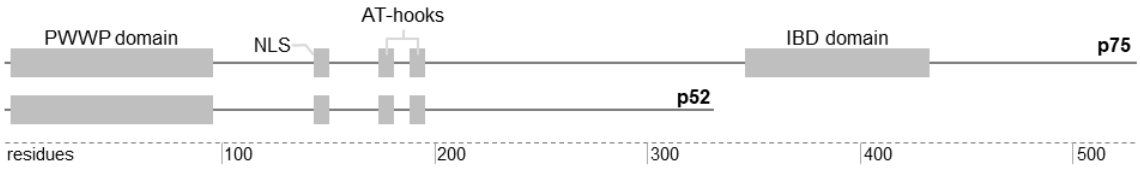


**Figure 3:** Domain architecture of the LEDGF splice variants p75 and p52. Both variants share the N-terminal part (residues 1-333) and comprise the chromatin reader Pro-Trp-Trp-Pro (PWWP) domain, nuclear localization signal sequence (NLS) and a pair of AT-hooks, canonical non-specific DNA binding motifs. The longer variant contains an additional protein binding scaffold domain, known as integrase binding domain (IBD, residues 245-431). The remaining parts of the protein are predicted to be intrinsically disordered.

The role of LEDGF/p75 in the integration of viral genetic information into the host genome, relying on the interplay between the PWWP and IBD domain interactions, was further extended to transcription co-activation (Ge et al., 1998), regulation of developmental genes (Sutherland et al., 2006), acute leukemia development (dependent on MLL1 fusion proteins) (Méreau et al., 2013), DNA homologous recombination-mediated repair (Daugaard et al., 2012) and overcoming the nucleosomal barrier to transcription in differentiated cells that are lacking the FACT (Facilitates Chromatin Transcription) complex (LeRoy et al., 2019). In particular, the link to Mixed-Lineage leukemia (MLL) revived interest in LEGDF/p75 as a drug target. It was shown that the IBD requires MENIN, an additional accessory partner that stabilizes MLL1, and facilitates this interaction with LEDGF/p75 (Huang et al., 2012) (**Figure 4**).

Initially, the chromatin recognition by the LEDGF PWWP domain was linked exclusively to the H3K36me3 mark (Pradeepa et al., 2012; Sankaran et al., 2016; Van Nuland et al., 2013), but later studies revealed its association with H3K36me2 marks (LeRoy et al., 2019; Okuda et al., 2014; Zhu et al., 2016). The structure for the PWWP domain was obtained using NMR spectroscopy (Eidahl et al., 2013; Pradeepa et al., 2014) (**Figure 5**). These studies also revealed that the affinity towards the methylated H3 histone N-terminal tail is relatively weak (mM) and requires additional substantial contacts with DNA in order to sustain the complex with the nucleosome. The high-resolution structural data for PWWP domain-bearing proteins in the context of methylated nucleosome is to date limited to a single cryo-electron microscopy (cryo-EM) structure of the LEDGF/p75 PWWP domain bound to H3K36me3 nucleosome that revealed the cooperative nature

of extensive interactions of the PWWP domain with the histone tail and DNA (Wang et al., 2019) (**Figure 5**). Despite the fact that the samples included a full-length LEDGF/p75 variant and the interaction to nucleosome was further stabilized by chemical cross-linking, only the PWWP domain bound to nucleosome was resolved. The IBD as well as the central and C-terminal intrinsically disordered regions remained undetected in the cryo-EM maps. The study also revealed that the nucleosome can symmetrically bind two molecules of LEDGF/p75 at the exit channel of the histone H3 N-terminal tail.

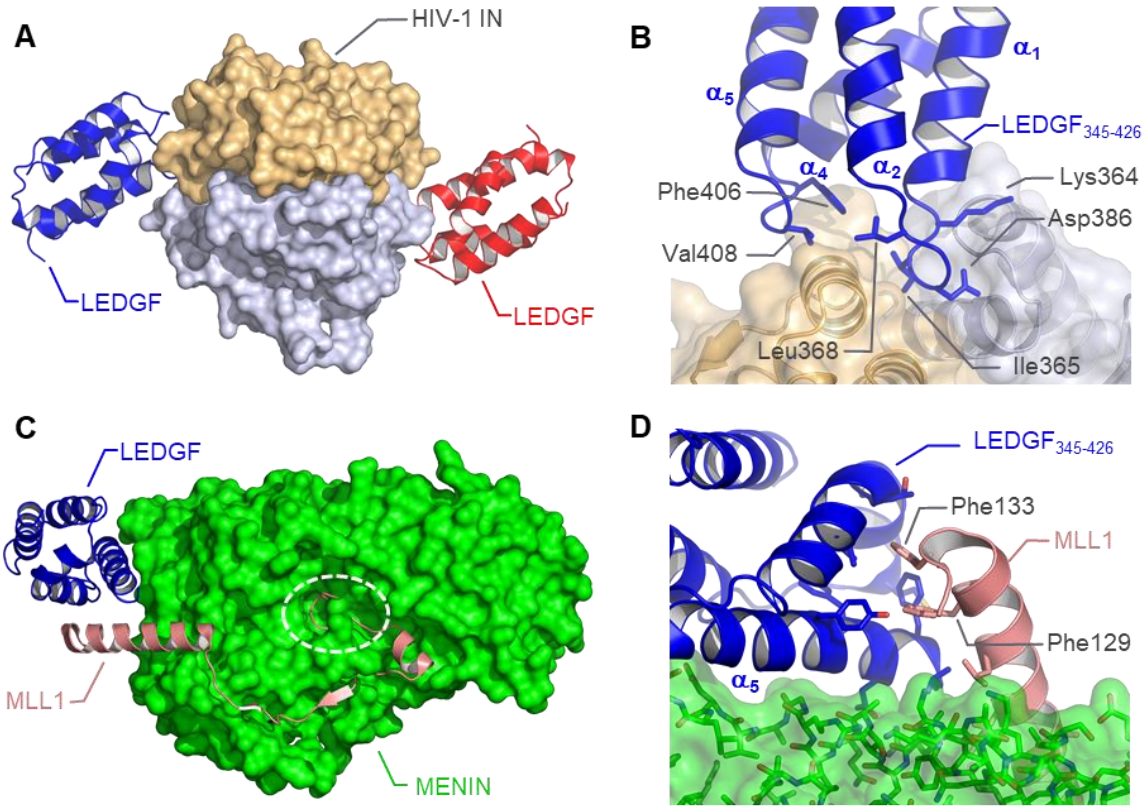


**Figure 4:** Structural basis for pathological roles of LEDGF/p75. (**A**) X-ray structure of dimeric catalytic core of HIV-1 integrase (highlighted in gold and gray) bound to two molecules of the LEDGF/p75 IBD domains (ribbon representation highlighted in blue and red) (Cherepanov et al., 2005; PDB: 2B4J). (**B**) The detailed insight in the interaction between the dimeric interface of HIV-1 integrase and IBD loops interconnecting helices α1- α2 and α4- α5. The IBD residues key for the interaction are shown as sticks. (**C**) X-ray structure of the ternary complex formed between the LEDGF/p75 IBD domain (ribbon representation highlighted in blue), MLL1 (residues 1-133 in pink) and MENIN (surface representation highlighted in green) (Huang et al., 2012; PDB: 3U88). The deep cleft on the MENIN surface that is utilized for interactions with the intrinsically-disordered region of MLL1 is highlighted by a dashed circle. This site was successfully targeted by small molecules that disrupt the protein-protein interaction (Christ et al., 2010; Jurado et al., 2013). (**D**) The detailed view of the interface between the LEDGF/p75 IBD domain (blue) and a structured MLL1 helix (pink). A pair of MLL1 bulky aromatic residues (Phe129 and Phe133) is essential for the integrity of the complex.

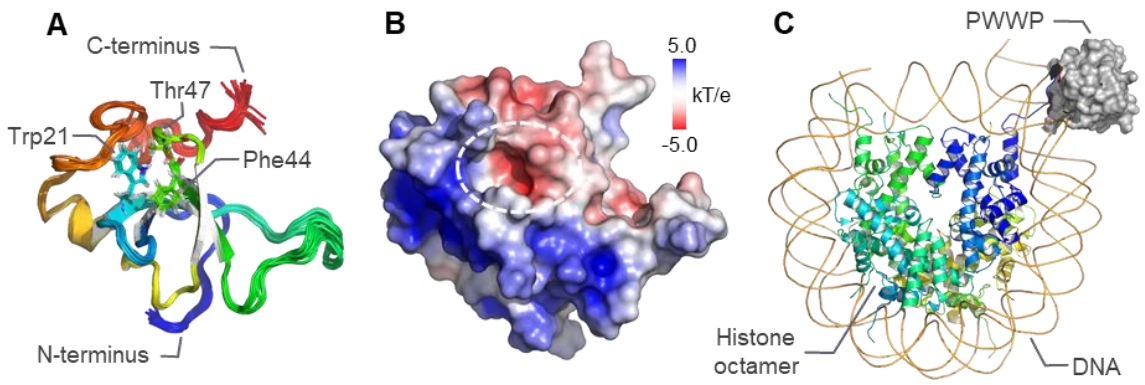


**Figure 5:** The molecular mechanism of chromatin recognition by the PWWP domain. (**A**) A set of converged structures obtained for the PWWP domain using NMR spectroscopy (Eidahl et al., 2013; PDB:2M16). The residues lining the cleft forming the site that recognizes the methylated lysine 36 chain on histone 3 are shown as sticks. (**B**) The PWWP surface charge distribution. There are extensive positively charged areas close to the methylation mark recognition cleft (highlighted by a dashed circle) that bind to the negatively charged nucleosomal DNA. (**C**) Cryo-electron microscopy structure of the LEDGF/p75 PWWP domain bound to the H3 exit channel of the nucleosome core particle. The histone octamer is represented as a cartoon, with H3 histones highlighted in green and blue; the DNA backbone is shown as an orange ribbon and PWWP as a grey surface (Wang et al., 2019; PDB:6S01).

## 1.3. Transient interactions

One of the prerequisites of life is the tight regulation of cellular processes. The regulatory networks heavily rely on biomolecular interactions that exhibit a broad range of binding affinities (Nooren and Thornton, 2003; Perkins et al., 2010) (**Figure 6**). The particular benefit of weak transient interactions, with $K_D$ in a micromolar range, is their relatively faster dissociation that allows for a rapid response to various cellular cues. As a consequence, these interactions are often short-lived. Transiently interacting biomolecules utilize smaller interfaces than permanently interacting ones, and their residue composition is similar to non-interacting surfaces, with a small enrichment of neutral polar groups. In addition, they are compact and are comprised of a central fully buried core, surrounded by peripheral regions that have properties that are generally not divergent from a non-interacting surface (Perkins et al., 2010). The formation of transient complexes is often accompanied by a conformational re-arrangement of interacting molecules, which in special cases includes disorder-to-order transition upon binding (Janin et al., 2008). Conformational stabilization of otherwise intrinsically disordered regions by transient interactions is entropically unfavorable, and therefore results in lower affinity complexes (Singh et al., 2007). An important class of molecular segments that form transient complexes are short linear motifs (SLiMs) (Ren et al., 2008), conserved amino acid sequences usually found within protein intrinsically disordered regions that interact with globular domains. They are frequently involved

in cell signaling as well as in the regulation of gene transcription. The eukaryotic genome accessibility is important for DNA transcription, replication and repair. These processes rely on transient interactions between DNA, histones and various chromatin-associated factors, including enzymes responsible for epigenetic marks maintenance, ATP-dependent chromatin remodelers, transcriptional activators or repressors and nuclear hormone receptors (Cermakova and Hodges, 2018).

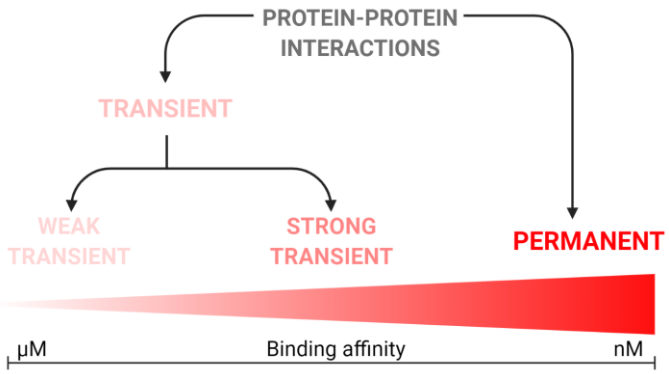


**Figure 6:** Classification of protein-protein interactions. Permanent interactions (low nanomolar range of binding affinities) are strong and often irreversible. Transient interactions are characterized by a limited lifetime. The strong transient category represents interactions that are triggered by effectors molecules or conformational change, while weak transient complexes are more dynamic and their lifetime is limited to seconds. Adapted from (Perkins et al., 2010).

Capturing weak protein-protein interactions is more challenging than studying stable complexes. However, the current portfolio of techniques offers a broad range of approaches to study all types of physical as well as functional associations of biomolecules (**Figure 7**). One of the most widely used methods for detection of *de novo* interactions is a yeast two-hybrid screening (Fields and Song, 1989). The principle is based on fusing proteins of interests with either an N-terminal DNA binding or the activation domain of a yeast transcriptional activator Gal4 that is reconstituted into a functional form, leading to transcription of a reported gene upon binary interaction of tested proteins. The technique can be adapted for a genome-wide screen using a cDNA library for the construction of activation domain fusions. Despite a relatively high rate of false positives, this technique can be extended to enhance accuracy by using a three-hybrid arrangement, which is instrumental for deciphering complex protein interaction networks on a domain level (Rawłuszko-Wieczorek et al., 2018). Another robust screening method for detection of protein-protein interactions is tandem affinity purification coupled to mass spectrometry detection (Collins and Choudhary, 2008). It is based on the double tagging of the protein of interest on its chromosomal locus, followed by a two-step purification and analysis. Its capability

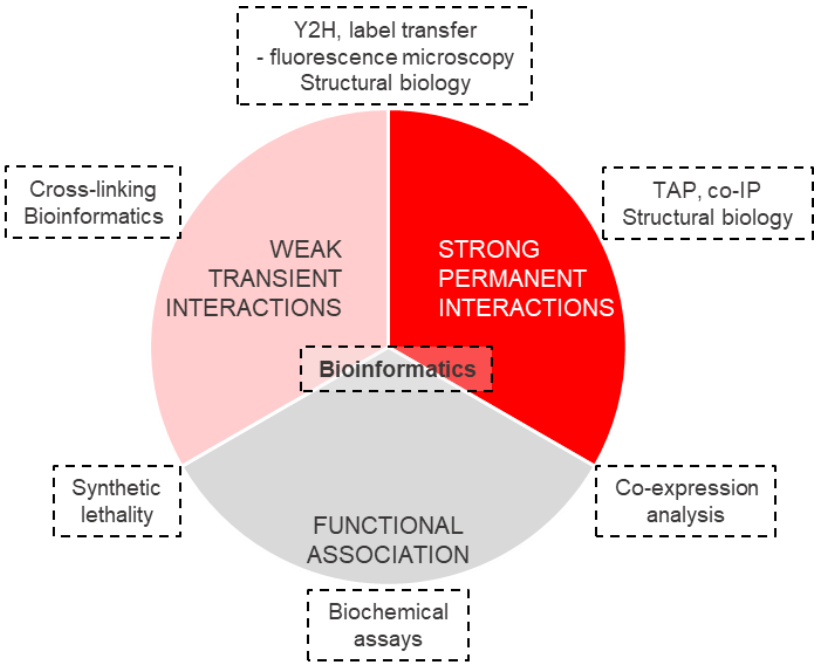


**Figure 7:** Methods for studying protein-protein interactions. Bioinformatics relies on algorithms that exploit genomic and evolutionary information. Y2H (yeast2hybrid), TAP (tandem affinity purification. Adapted from (Perkins et al., 2010).

to capture transiently established complexes was gradually increased by the introduction of chemical crosslinking that allows the stabilization of complexes *in vitro* or in cells (Worthington et al., 2006). Particular interactions are then studied *in vivo* using fluorescence microcopy techniques, such as fluorescence resonance energy transfers (FRET), bioluminescence resonance energy transfer (BRET) or bimolecular fluorescence complementation (BiFC). *In vitro* interaction characterization methods range from affinity purification, co-immunoprecipitation, and microarray-based analysis to more complex structural biology techniques. Experimental approaches are complemented by computational methods for *in silico* prediction of biomolecular interactions, which are based on sequence or structural comparisons, chromosome proximity or gene expression patterns (Rao et al., 2014). Nevertheless, only structural biology (**Figure 8**) can currently provide atomic-resolution information for a particular interaction.

X-ray crystallography and to some extent NMR spectroscopy were, over several decades, considered to be core structural biology methods (**Figure 1**). However, the recent leaps in cryo-electron miscopy instrumentation and data processing dramatically changed this stereotype (Kühlbrandt, 2014). Mass spectrometry coupled with chemical crosslinking, small angle X-ray or neutron scattering and computational methods then complement the core methods. In particular, the latest implementation of machine learning in the computational determination of protein folds from amino acid sequence will increase the impact of *in silico* methods (14$^{th}$ Critical Assessment of Structural Prediction competition) and stimulate further development of

integrative methodology based on sparse experimental data (Seffernick and Lindert, 2020). The overview of structural biology methods, together with the summary of provided structural information, is listed in

**Table *1***. The limited lifetime of transient interactions is a major obstacle in their detailed characterization by X-ray crystallography and cryo-electron microscopy. They either resist successful crystallization, the key pre-requisite for X-ray analysis, or remain unresolved in experimental cryo-electron microscopy maps (Hanske et al., 2018). NMR spectroscopy therefore offers a unique opportunity to study transient and dynamic assemblies, and its capability was fully exploited in the work described in the Results and Discussion chapter below.

***Table 1:*** *Structural biology methods. Adapted from (Ziegler et al., 2021).*

| Method | Information |
|---|---|
| **X-ray crystallography** | Captures atomic resolution detail of stable biomolecular conformations |
| | Provides structures that can be fitted into experimental SAXS/molecular docking/low-resolution cryo-EM data |
| **NMR spectroscopy** | Captures atomic detail of small biomolecules |
| | Provides data for flexible biomolecules |
| | Allows for rapid Identification of interacting regions |
| | Provides details for conformational dynamics |
| | Provides structures can be fitted into experimental SAXS/molecular docking/low-resolution cryo-EM data |
| **SAXS/SANS** | Provides overall protein complex shape that can be fitted with atomic resolution structures |
| **Cryo-electron microscopy (single-particle)** | Captures high-resolution stable conformations that can be fitted into SAXS/molecular docking models |
| | Captures lower-resolution flexible protein conformations |
| | Provides overall shape for isolated molecules or their assemblies that can be fitted with atomic resolution structures |
| **Computational modeling** | Delivers detailed atomic resolution subunit predictions which can be fitted into SAXS/molecular docking models of biomolecular complexes |
| **Mass spectrometry (chemical crosslinking)** | Captures tightly and weakly interacting partners and identifies specific surface exposed residues that are in proximity to each other |

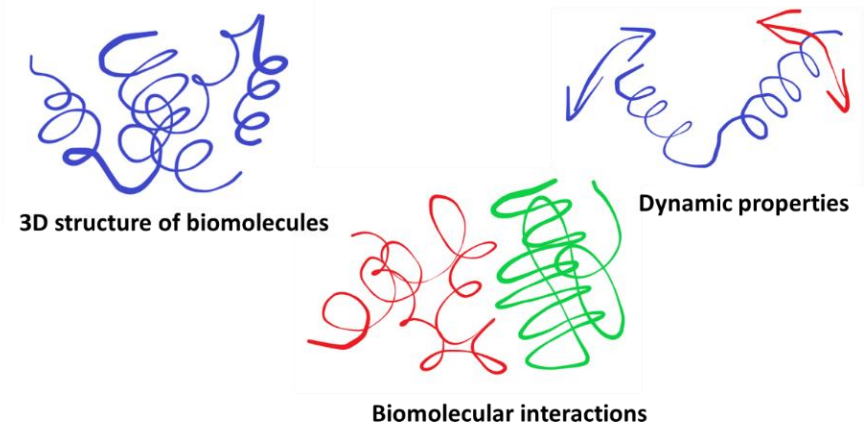


**Figure 8:** Structure biology tools can provide atomic-resolution information about the structure of biomolecules or their complexes. Despite the fact that the data are represented as static 'pictures', the data include information about dynamic properties, such as global or local molecular motions.

## 1.4. NMR spectroscopy from a biologist's perspective

The common goal of structural biology is to obtain any form of structural information that can help to test hypotheses or to answer research questions. A comprehensive description of the studied system is not always required. NMR spectroscopy is an excellent tool to deliver specific information with atomic resolution without the need to obtain full structural coordinates (Barrett et al., 2013).

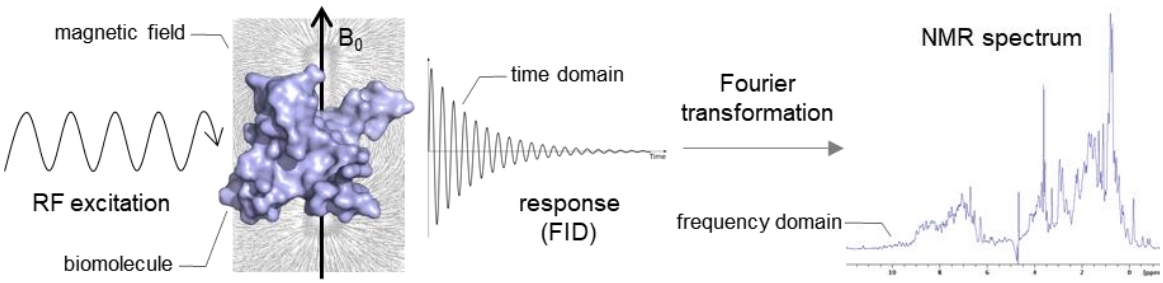


**Figure 9:** The NMR active nuclei ($^{1}H$, $^{13}C$, $^{15}N$ and $^{31}P$ in biomolecules) placed in the homogeneous magnetic field can selectively 'absorb' and 'dissipate' radiation of a particular frequency. The measured signal can be transformed in the NMR spectrum. The frequencies reflect the type of the nucleus as well as its covalent/spatial context, and can be used for structural characterization of the molecules.

The principle of NMR spectroscopy is based on the interaction of radiofrequency radiation with NMR-active nuclei from studied molecules that are placed in a homogeneous magnetic field (**Figure 9**). The excitation (also known as resonance) frequencies of nuclei are affected by their environment and can be detected. Decades of instrumentation and methodology development made NMR accessible for routine use in the structural characterization of biomolecular systems.

The biggest leap came with a realization, inspired by the X-ray crystallography field, that there is no need for a deep understanding of the physical principles of the experiments in order to use them. However, a full appreciation of experimental requirements and limitations is desirable.

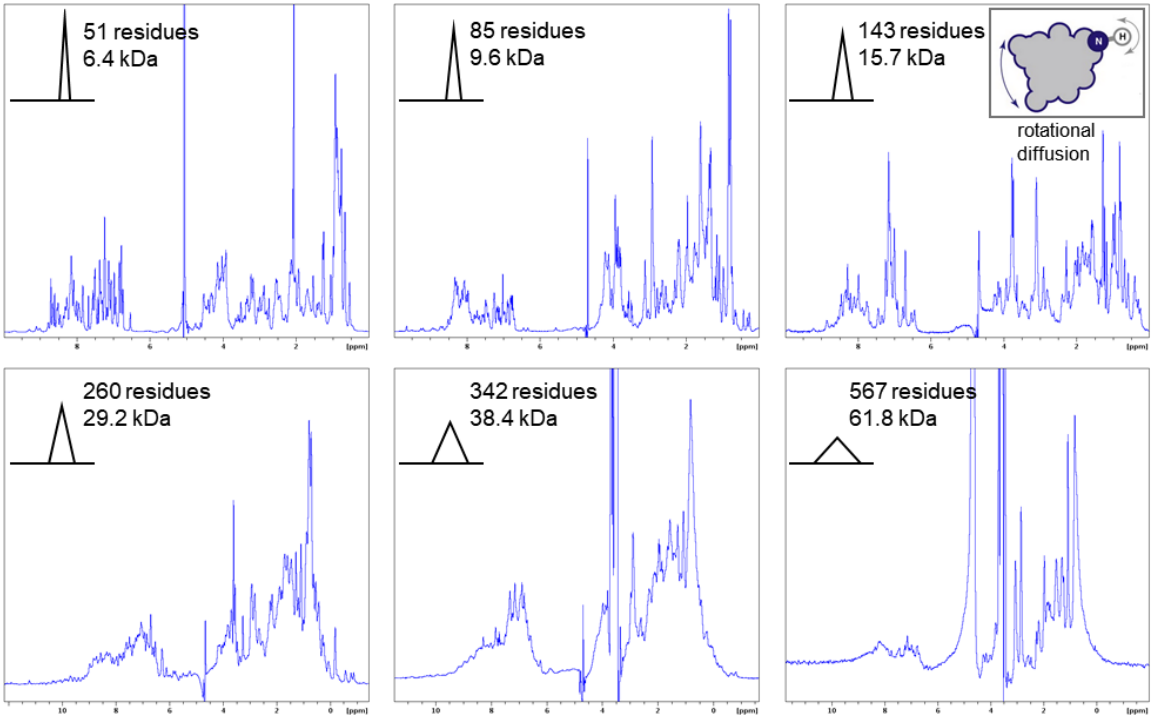


**Figure 10:** Examples of 1D NMR spectra of proteins (human insulin, ubiquitin, carbonic anhydrase II, scFV fragment of an anti-interleukin-1β monoclonal antibody, Fab fragment of a monoclonal antibody, full-length monoclonal antibody) (Addis et al., 2014; Krizkova et al., 2014; Pecina et al., 2018; Siva et al., 2016; Wilkinson et al., 2009). The protein size causes a slower rotational diffusion ('tumbling') of molecules that increases signal linewidth. The complexity of spectra is further increased by a limited covalent diversity of amino acids.

The major drawback for the application of NMR technology in biology is the limit on the size of studied systems. The global molecular motion rates, such as rotational diffusion, become gradually lower as the size of the molecules increases (**Figure 10**), which is manifested in increasing line-broadening of NMR signals. The widely accepted molecular weight limits for NMR (Yu, 1999) do not take into account additional properties, such as the non-spherical hydrodynamic shape of molecules, susceptibility to aggregation, or unfavorable conformational heterogeneity, that can prevent NMR analysis of much smaller systems. Full structural characterization is accessible to molecules up to 35 k Da, with molecules over 25 kDa requiring either fractional or uniform deuteration of non-labile hydrogens (Sattler and Fesik, 1996). In specific cases, even bigger 'well-behaving' protein-protein complexes, e.g. single-chain antibody fragments bound to globular antigens (~ 45 kDa), are amenable for structural characterization (Addis et al., 2014; Wilkinson et

al., 2009). Larger systems can still yield good quality NMR data that can provide information on binding interfaces or local dynamics of studied biomolecules. However, they require a combination of deuteration and selective labelling (Sprangers and Kay, 2007).

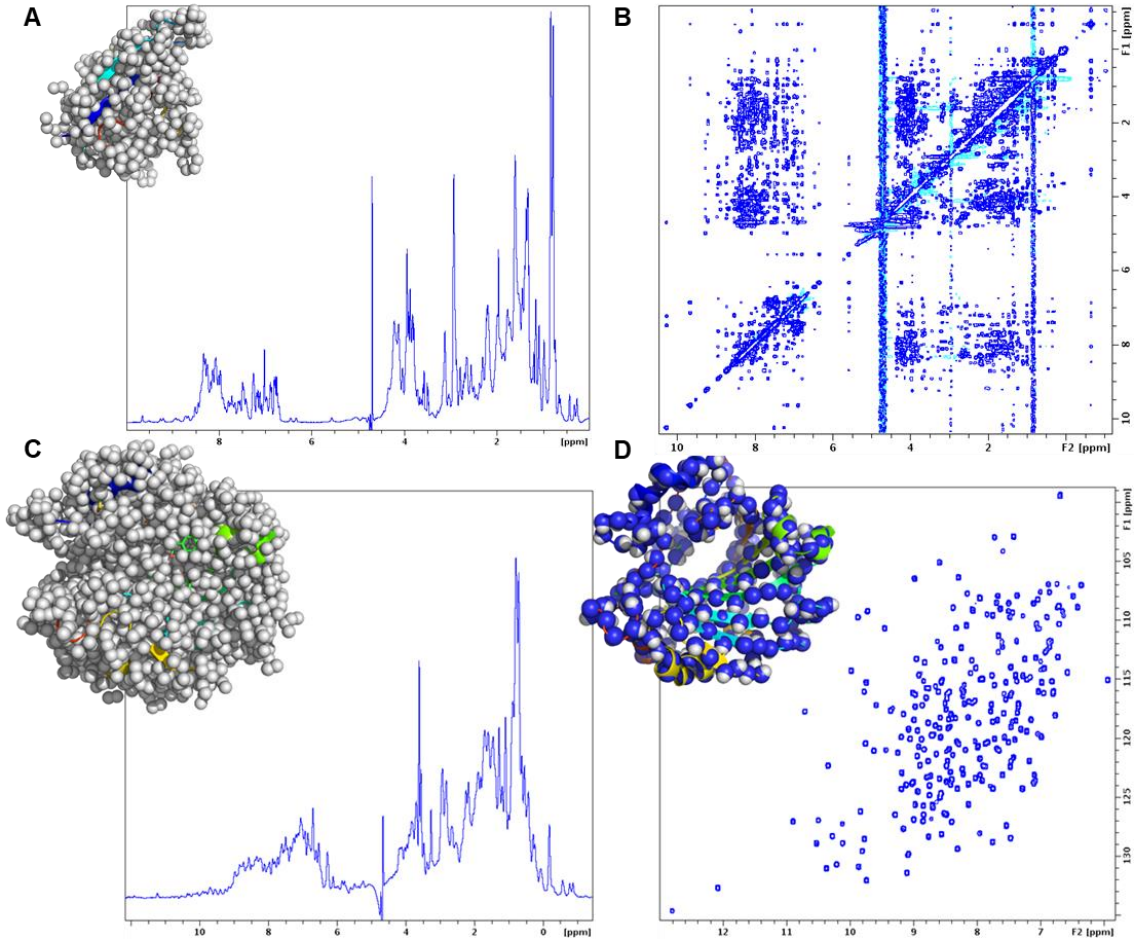


**Figure 11:** Increasing size of biomolecules leads to a higher complexity of NMR spectra. (**A**) One-dimensional $^{1}$H NMR spectrum of 10 kDa human ubiquitin. Some hydrogen signals are well resolved, but most of them are in overlapped regions. The hydrogens that yield NMR signals in the spectrum are shown as white spheres in the structural model. (**B**) Two-dimensional $^{1}$H/$^{1}$H (homonuclear) NMR spectrum of 10 kDa human ubiquitin correlating spatially proximal hydrogens through nuclear Overhauser effect. The increased dimensionality increased the interpretability of the spectrum. However, certain regions are still heavily overlapped. (**C**) One-dimensional $^{1}$H NMR spectrum of 16 kDa human carbonic anhydrase II. (**D**) Two-dimensional $^{15}$N/$^{1}$H (heteronuclear) NMR spectrum of human carbonic anhydrase II correlating directly attached amide nitrogen with hydrogens through covalent bond. The introduction of stable isotopes into proteins during their expression is an elegant way to increase resolution of the spectra by the correlation of specific groups of nuclei. This spectrum allows for selective observation of protein backbone and sidechain amide groups. The amide groups are represented in the structure as blue (nitrogens) and white (hydrogens) spheres, respectively.

In addition, both proteins and nucleic acids are assembled from a limited number of building block types. Larger molecules have highly repetitive covalent structures which intensify the complexity of the NMR spectra due to severe signal overlaps (**Figure 11**). The overlaps are resolved

by an increased dimensionality of the spectra and by utilizing the other NMR-active nuclei alongside hydrogens (Cavanagh et al., 2007). The increased dimensionality requires disproportionally longer acquisition times due to the necessity of obtaining the required resolution in the additional dimension. As an illustration, a typical one-dimensional spectrum of a 100 μM 10 kDa protein can be acquired using a standard instrument suitable for biomolecular NMR in under a minute, two-dimensional spectra in tens of minutes to several hours, and three- or more dimensions require days of experimental time. The acquisition schemes based on non-uniform sampling of additional dimensions, complemented by suitable reconstruction algorithms, helped to alleviate this issue (Delaglio et al., 2017) and made NMR data from diluted samples more accessible. Unfortunately, the most abundant carbon and nitrogen isotopes ($^{12}C$ and $^{14}N$) are not NMR-active. Therefore, the biomolecules have to be enriched for the active stable isotopes ($^{13}C$ and $^{15}N$) either during expression in host cells (proteins and RNA) or synthetically (DNA). The introduction of NMR-active heteronuclei then open an access to various correlation spectra, which are indispensable for fundamental steps in NMR analysis of biomolecules, such as signal assignment (Cavanagh et al., 2007).

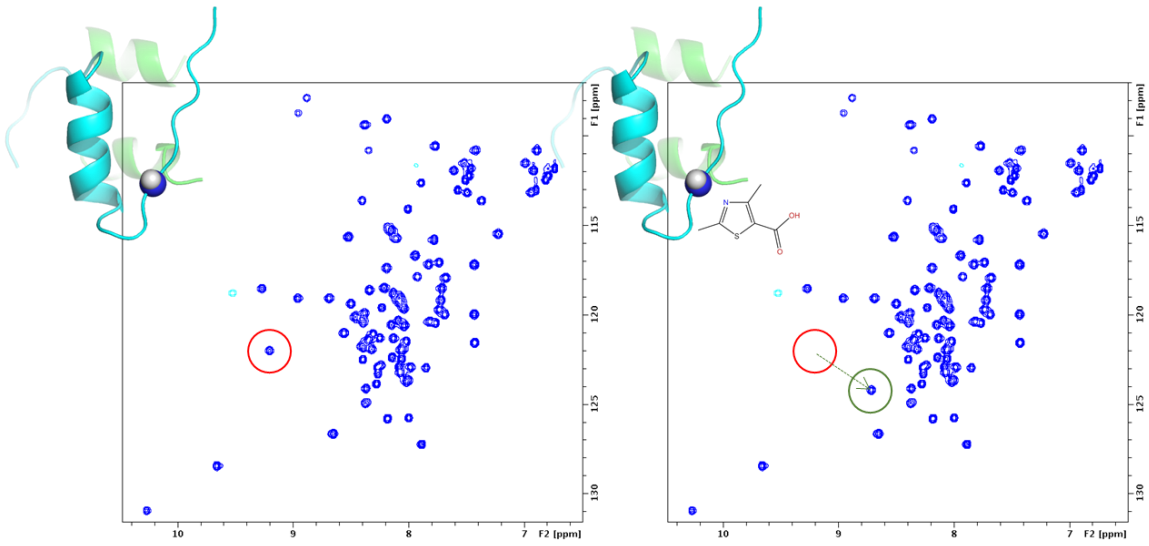


**Figure 12:** A hypothetical example of an NMR titration experiment using two-dimensional $^{15}N/^{1}H$ heteronuclear correlation spectrum. Each signal represents one amide group from a protein backbone. The binding of an unlabeled molecule, such as a low molecular weight inhibitor, to a specific site on the surface of the labelled protein leads to changes of the signals close to the site of interaction. This allows for direct identification of the site or determination of binding affinity (Williamson, 2013). Examples using this approach are provided throughout the results below.

NMR provides a unique opportunity for observation of any NMR-active nucleus. Each signal in the spectra is in fact an observation of a particular atom (one-dimensional spectrum) or group (multi-dimensional correlation spectra), unlike reflections in the crystal lattice. The position of a signal in the spectra is defined by the atom/group's surrounding and changes within this

environment, such as the binding of another molecule, lead to signal displacement from its original position (**Figure 12**). NMR-detected titrations of isotopically enriched biomolecules by unlabeled partner molecules are one of the most widely used biological applications of NMR (Williamson, 2013). The signals from the labelled molecule are visualized in well-resolved, typically two dimensional hetero-correlation spectra. The unlabeled partner remains invisible, but its effects are monitored as selective perturbations of NMR signals of the labelled molecule.

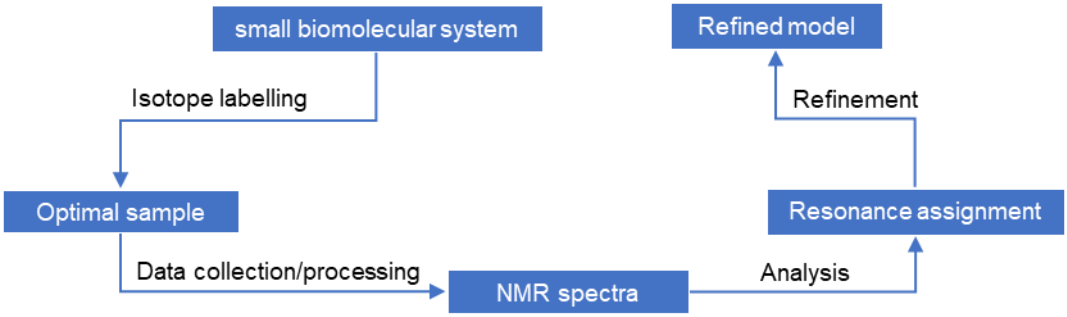


**Figure 13:** The path to the three-dimensional structure of biomolecules using NMR spectroscopy.

Despite the size limitations described above, NMR can be used for the full structural determination of biomolecules or their complexes. The relatively low sensitivity in comparison with other spectroscopies is caused by the utilization of low-energy radiofrequency signal. However, the main benefits of this method are the capability to capture dynamic molecules, and to tolerate some degree of heterogeneity. A flowchart illustrating key steps in obtaining structures by NMR is shown in **Figure 13**. The $^{13}C$ and $^{15}N$ isotope labelling is relatively inexpensive and straightforward in bacteria with the use of minimal media, but is several folds more expensive in eukaryotic or cell free expression systems. The optimal sample of a concentration > 100μM should be stable at room temperature for at least the several days that are required for dataset acquisition, as processes such as gradual aggregation or proteolysis negatively affect the data quality. Routine data collection can be accomplished within two weeks using the current instrumentation, depending on the size and concentration of the studied biomolecules. The data analysis includes a series of iterative steps: NMR resonance assignment, the generation of restraints for structural calculation, and the actual structural refinement and validation. Although there is a broad consensus on the choice of experiments, the actual implementation and experimental setup are highly variable. Our optimal pipeline includes (i) data processing using TOPSPIN (Bruker), (ii) analysis in NMRFAM-SPARKY (Lee et al., 2015), (iii) structural refinement combined with distance restraints assignment in CYANA (Herrmann et al., 2002), (iv) a final molecular dynamics simulation (approximately 10 ns) in explicit water using YASARA (Krieger et

al., 2002) and (v) validation (Protein Structure Validation Server). There are numerous efforts to accelerate the data analysis step through automation (Evangelidis et al., 2018; Guerry et al., 2015; Lee et al., 2019). However, a wider applicability of these algorithms is prevented by the size of molecules or the ubiquitous incompleteness of datasets.

Intrinsic disorder is a structural feature of a surprisingly high percentage of especially eukaryotic proteomes (Uversky, 2019). Over 60% of eukaryotic proteins include long, intrinsically disordered protein regions that are not amenable to structural characterization by X-ray crystallography or cryo-electron microscopy. NMR spectroscopy is the only high-resolution technique that can provide information on conformational properties of these regions, such as revealing conformational propensities within molecular ensembles or monitoring changes upon interaction with other molecules (Prestel et al., 2018). It is a surprisingly common scenario to have the studied molecules include both well-ordered and semi-flexible or fully flexible segments, which particularly benefits from the versatility of NMR spectroscopy (**Figure 14**).

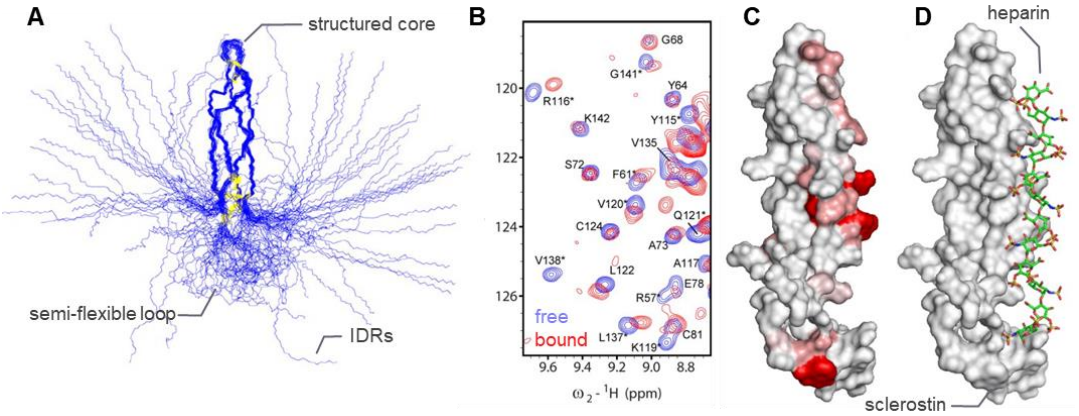


**Figure 14:** NMR can tolerate conformational heterogeneity and provide information on ordered and disordered regions simultaneously. (**A**) Set of converged structures obtained by NMR for sclerostin, negative regulator of Wnt signaling in bone. The 150 residues long protein consists of a structured core, stabilized by a cystine knot, semi-flexible loop and intrinsically disordered termini. The effect of heparin binding to sclerostin was monitored using two-dimensional $^{15}N/^{1}H$ heteronuclear NMR spectra (**B**) and the most affected residues were highlighted on the surface of the central part of sclerostin (**C**). The NMR data for the binding were utilized for construction of the model for the complex of sclerostin with heparin (**D**). Adapted from (Veverka et al., 2009).

## 2. Results and Discussion

The results and discussion include four papers published between 2014-2020 that are related to LEDGF/p75 biology. The papers that follow span an impressive range of disciplines, combining protein biochemistry and biophysics with cell and structural biology. They were published jointly with collaborators from KU Leuven and Baylor College of Medicine. As a corresponding author, my role consisted of helping to conceive the ideas behind the research, to supervise and analyze the experimental designs, to discuss key aspects of the manuscript, and to help write and revise the final manuscript.

# 4. References


Addis, P.W., Hall, C.J., Bruton, S., Veverka, V., Wilkinson, I.C., Muskett, F.W., Renshaw, P.S., Prosser, C.E., Carrington, B., Lawson, A.D.*, et al.* (2014). Conformational heterogeneity in antibody-protein antigen recognition: implications for high affinity protein complex formation. J Biol Chem *289*, 7200-7210.

Bannister, A.J., Schneider, R., and Kouzarides, T. (2002). Histone Methylation: Dynamic or Static? Cell *109*, 801-806.

Barrett, P.J., Chen, J., Cho, M.K., Kim, J.H., Lu, Z., Mathew, S., Peng, D., Song, Y., Van Horn, W.D., Zhuang, T.*, et al.* (2013). The quiet renaissance of protein nuclear magnetic resonance. Biochemistry *52*, 1303-1320.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M., and Skelton, N.J. (2007). Protein Nmr Spectroscopy Principles and Practice Second Edition Preface. Protein Nmr Spectroscopy: Principles and Practice, 2nd Edition, V-Vi.

Cermakova, K., and Hodges, H.C. (2018). Next-Generation Drugs and Probes for Chromatin Biology: From Targeted Protein Degradation to Phase Separation. Molecules *23*.

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

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

Christ, F., Voet, A., Marchand, A., Nicolet, S., Desimmie, B.A., Marchand, D., Bardiot, D., Van der Veken, N.J., Van Remoortel, B., Strelkov, S.V.*, et al.* (2010). Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nature chemical biology *6*, 442-448.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nature Medicine *11*, 1287-1289.

Collins, M.O., and Choudhary, J.S. (2008). Mapping multiprotein complexes by affinity purification and mass spectrometry. Current Opinion in Biotechnology *19*, 324-330.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sørensen, C.S., Petersen, N.H.T., Sorensen, P.H.B., Lukas, C., Bartek, J.*, et al.* (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nature Structural and Molecular Biology *19*, 803-810.

Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. Cell *150*, 12-27.

Delaglio, F., Walker, G.S., Farley, K.A., Sharma, R., Hoch, J.C., Arbogast, L.W., Brinson, R.G., and Marino, J.P. (2017). Non-Uniform Sampling for All: More NMR Spectral Quality, Less Measurement Time. Am Pharm Rev *20*, 339681.

Duncan, K.W., and Campbell, J.E. (2018). Epigenetic Modulators. In Cancer II, M.J. Waring, ed. (Cham: Springer International Publishing), pp. 227-227.

Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S.*, et al.* (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Res *41*, 3924-3936.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol *28*, 817-825.

Evangelidis, T., Nerli, S., Nováček, J., Brereton, A.E., Karplus, P.A., Dotas, R.R., Venditti, V., Sgourakis, N.G., and Tripsianes, K. (2018). Automated NMR resonance assignments and structure determination using a minimal set of 4D spectra. Nat Commun *9*, 384.

Fields, S., and Song, O.-k. (1989). A novel genetic system to detect protein–protein interactions. Nature *340*, 245-246.

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

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nature Reviews Genetics *13*, 343-357.

Guerry, P., Duong, V.D., and Herrmann, T. (2015). CASD-NMR 2: robust and accurate unsupervised analysis of raw NOESY spectra and protein structure determination with UNIO. Journal of Biomolecular NMR *62*, 473-480.

Hanske, J., Sadian, Y., and Müller, C.W. (2018). The cryo-EM resolution revolution and transcription complexes. Current opinion in structural biology *52*, 8-15.

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

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

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Experimental & Molecular Medicine *49*, e324-e324.

Janin, J., Bahadur, R.P., and Chakrabarti, P. (2008). Protein–protein interaction and quaternary structure. Q Rev Biophys *41*, 133-180.

Jurado, K.A., Wang, H., Slaughter, A., Feng, L., Kessl, J.J., Koh, Y., Wang, W., Ballandras-Colas, A., Patel, P.A., Fuchs, J.R.*, et al.* (2013). Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. Proceedings of the National Academy of Sciences *110*, 8690.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proceedings of the National Academy of Sciences *108*, 13564-13569.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nature genetics *41*, 376-381.

Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. Proteins *47*, 393-402.

Krizkova, K., Veverka, V., Maletinska, L., Hexnerova, R., Brzozowski, A.M., Jiracek, J., and Zakova, L. (2014). Structural and functional study of the GlnB22-insulin mutant responsible for maturity-onset diabetes of the young. PLoS One *9*, e112883.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C.*, et al.* (2003). Methylation of Histone H3 by Set2 in <em>Saccharomyces cerevisiae</em> Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and Cellular Biology *23*, 4207.

Kühlbrandt, W. (2014). The Resolution Revolution. Science *343*, 1443.

Lee, W., Bahrami, A., Dashti, H.T., Eghbalnia, H.R., Tonelli, M., Westler, W.M., and Markley, J.L. (2019). I-PINE web server: an integrative probabilistic NMR assignment system for proteins. Journal of Biomolecular NMR *73*, 213-222.

Lee, W., Tonelli, M., and Markley, J.L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics *31*, 1325-1327.

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

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.-M. (2013). The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutSα. Cell *153*, 590-600.

Lux, V., Brouns, T., Cermakova, K., Srb, P., Fabry, M., Madlikova, M., Horejsi, M., Kukacka, Z., Novak, P., Kugler, M.*, et al.* (2020). Molecular Mechanism of LEDGF/p75 Dimerization. Structure.

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

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci *28*, 69-74.

Mereau, H., De Rijck, J., Cermakova, K., Kutz, A., Juge, S., Demeulemeester, J., Gijsbers, R., Christ, F., Debyser, Z., and Schwaller, J. (2013). Impairing MLL-fusion gene-mediated transformation by dissecting critical interactions with the lens epithelium-derived growth factor (LEDGF/p75). Leukemia *27*, 1245-1253.

Méreau, H., De Rijck, J., Čermáková, K., Kutz, A., Juge, S., Demeulemeester, J., Gijsbers, R., Christ, F., Debyser, Z., and Schwaller, J. (2013). Impairing MLL-fusion gene-mediated transformation by dissecting critical interactions with the lens epithelium-derived growth factor (LEDGF/p75). Leukemia *27*, 1245-1253.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77.

Nooren, I.M.A., and Thornton, J.M. (2003). Structural Characterisation and Functional Significance of Transient Protein–Protein Interactions. Journal of Molecular Biology *325*, 991-1018.

Okuda, H., Kawaguchi, M., Kanai, A., Matsui, H., Kawamura, T., Inaba, T., Kitabayashi, I., and Yokoyama, A. (2014). MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters. Nucleic Acids Research *42*, 4241-4256.

Pecina, A., Brynda, J., Vrzal, L., Gnanasekaran, R., Horejsi, M., Eyrilmez, S.M., Rezac, J., Lepsik, M., Rezacova, P., Hobza, P.*, et al.* (2018). Ranking Power of the SQM/COSMO Scoring Function on Carbonic Anhydrase II-Inhibitor Complexes. Chemphyschem *19*, 873-879.

Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. Structure *18*, 1233-1243.

Pradeepa, M.M., Grimes, G.R., Taylor, G.C., Sutherland, H.G., and Bickmore, W.A. (2014). Psip1/Ledgf p75 restrains Hox gene expression by recruiting both trithorax and polycomb group proteins. Nucleic Acids Res *42*, 9021-9032.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genetics *8*.

Prestel, A., Bugge, K., Staby, L., Hendus-Altenburger, R., and Kragelund, B.B. (2018). Chapter Eight - Characterization of Dynamic IDP Complexes by NMR Spectroscopy. In Methods in Enzymology, E. Rhoades, ed. (Academic Press), pp. 193-226.

Rao, V.S., Srinivas, K., Sujini, G.N., and Kumar, G.N.S. (2014). Protein-Protein Interaction Detection: Methods and Analysis. International Journal of Proteomics *2014*, 147648.

Rawłuszko-Wieczorek, A.A., Knodel, F., Tamas, R., Dhayalan, A., and Jeltsch, A. (2018). Identification of protein lysine methylation readers with a yeast three-hybrid approach. Epigenetics & Chromatin *11*, 4.

Ren, S., Uversky, V.N., Chen, Z., Dunker, A.K., and Obradovic, Z. (2008). Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. BMC Genomics *9*, S26.

Sankaran, S.M., Wilkinson, A.W., Elias, J.E., and Gozani, O. (2016). A PWWP Domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated lys-36 of histone H3 and regulates NSD2 function at chromatin. Journal of Biological Chemistry *291*, 8465-8474.

Sattler, M., and Fesik, S.W. (1996). Use of deuterium labeling in NMR: overcoming a sizeable problem. Structure *4*, 1245-1249.

Seffernick, J.T., and Lindert, S. (2020). Hybrid methods for combined experimental and computational determination of protein structure. The Journal of Chemical Physics *153*, 240901.

Sharma, S., Cermakova, K., De Rijck, J., Demeulemeester, J., Fabry, M., El Ashkar, S., Van Belle, S., Lepsik, M., Tesina, P., Duchoslav, V.*, et al.* (2018). Affinity switching of the LEDGF/p75 IBD

interactome is governed by kinase-dependent phosphorylation. Proc Natl Acad Sci U S A *115*, E7053-E7062.

Singh, G.P., Ganapathi, M., and Dash, D. (2007). Role of intrinsic disorder in transient interactions of hub proteins. Proteins: Structure, Function, and Bioinformatics *66*, 761-765.

Siva, M., Svoboda, M., Veverka, V., Trempe, J.F., Hofmann, K., Kozisek, M., Hexnerova, R., Sedlak, F., Belza, J., Brynda, J.*, et al.* (2016). Human DNA-Damage-Inducible 2 Protein Is Structurally and Functionally Distinct from Its Yeast Ortholog. Sci Rep *6*, 30443.

Sprangers, R., and Kay, L.E. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature *445*, 618-622.

Sutherland, H.G., Newton, K., Brownstein, D.G., Holmes, M.C., Kress, C., Semple, C.A., and Bickmore, W.A. (2006). Disruption of Ledgf/Psip1 Results in Perinatal Mortality and HomeoticSkeletal Transformations. Molecular and Cellular Biology *26*, 7201-7210.

Tesina, P., Cermakova, K., Horejsi, M., Prochazkova, K., Fabry, M., Christ, F., Demeulemeester, J., Debyser, Z., De Rijck, J., Veverka, V.*, et al.* (2015). Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. . Nature communications *submitted*, third revision.

Uversky, V.N. (2019). Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. Frontiers in Physics *7*.

Van Belle, S., El Ashkar, S., Čermáková, K., Matthijssens, F., Goossens, S., Canella, A., Hodges, C.H., Christ, F., De Rijck, J., Van Vlierberghe, P.*, et al.* (2021). Unlike its Paralog LEDGF/p75, HRP-2 Is Dispensable for MLL-R Leukemogenesis but Important for Leukemic Cell Survival. Cells *10*.

Van Nuland, R., Van Schaik, F.M.A., Simonis, M., Van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.T.M., and Van Ingen, H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. Epigenetics and Chromatin *6*, 1-12.

Veverka, V., Henry, A.J., Slocombe, P.M., Ventom, A., Mulloy, B., Muskett, F.W., Muzylak, M., Greenslade, K., Moore, A., Zhang, L.*, et al.* (2009). Characterization of the structural features and interactions of sclerostin: molecular insight into a key regulator of Wnt-mediated bone formation. J Biol Chem *284*, 10890-10900.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nature Reviews Molecular Cell Biology *13*, 115-126.

Wang, H., Farnung, L., Dienemann, C., and Cramer, P. (2019). Structure of H3K36-methylated nucleosome–PWWP complex reveals multivalent cross-gyre binding. Nature Structural & Molecular Biology.

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

Wilkinson, I.C., Hall, C.J., Veverka, V., Shi, J.Y., Muskett, F.W., Stephens, P.E., Taylor, R.J., Henry, A.J., and Carr, M.D. (2009). High resolution NMR-based model for the structure of a scFv-IL-1beta complex: potential for NMR as a key tool in therapeutic antibody design and development. J Biol Chem *284*, 31928-31935.

Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. Progress in nuclear magnetic resonance spectroscopy *73*, 1-16.

Worthington, A.S., Rivera, H., Torpey, J.W., Alexander, M.D., and Burkart, M.D. (2006). Mechanism-Based Protein Cross-Linking Probes To Investigate Carrier Protein-Mediated Biosynthesis. ACS Chemical Biology *1*, 687-691.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

Yu, H. (1999). Extending the size limit of protein nuclear magnetic resonance. Proceedings of the National Academy of Sciences *96*, 332.

Zhu, L., Li, Q., Wong, S.H.K., Huang, M., Klein, B.J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C.*, et al.* (2016). ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia. Cancer Discovery *6*, 770-783.

Ziegler, S.J., Mallinson, S.J.B., St. John, P.C., and Bomble, Y.J. (2021). Advances in integrative structural biology: Towards understanding protein complexes in their cellular context. Computational and Structural Biotechnology Journal *19*, 214-225.