technische universität
dortmund
fakultät für chemie
und chemische biologie

MAX PLANCK INSTITUTE
OF MOLECULAR PHYSIOLOGY

CHEMICAL GENOMICS CENTRE
OF THE MAX PLANCK SOCIETY

Targeting RNA-Protein Interactions with Peptides and Small Molecule Inhibitors

Dissertation

Submitted for the degree of Doctor of Natural Sciences (Dr. rer. nat)

to the

Faculty of Chemistry and Chemical Biology

Technical University Dortmund

By

Jen-Yao Chang

張仁垚

Dortmund, 2023

“知之為知之，不知為不知，是知也。”

- 孔子《論語・為政》

“I know that I know nothing.”

- Socrates saying, Plato, Apology 22d, translated by Harold North Fowler, 1966

This work was prepared from January 2019 to September 2023 in the group of Dr. Peter 't Hart of Chemical Genomics Centre of the Max Planck Society at the Max-Planck Institute of Molecular Physiology in Dortmund.

The research was funded by Government Scholarship for Overseas Study from the Ministry of Education of Taiwan, the Max-Planck Institute of Molecular Physiology, AstraZeneca, Merck KGaA, and Pfizer.

Dean: Prof. Dr. Stefan M. Kast

First Referee: Prof. Dr. Dr. h.c. Herbert Waldmann

Second Referee: Priv.-Doz. Dr. Lief Dehmelt

Acknowledgement

It is a long road for my doctoral research. I am fortunate to be supported, guided, and accompanied by many others.

First of all, I would like to thank Prof. Dr. Herbert Waldmann and Dr. Peter 't Hart for supervising me on this project. Their knowledge and experience are the most important for the realization of the project.

Secondly, I would like to thank Priv.-Doz. Dr. Leif Dehmelt for agreeing to be the second examiner for this thesis at very short notice.

To Cora Neugebauer (M. Sc.) and Anne Mues gen. Koers (B. Sc.), I am honored to have had you both work with me during your Master/Bachelor research. Things were a bit crazy during and after the pandemic, but we made it, and a publication is coming out of your result. I believe you both will have great success in the future.

To the members of Ag Hart and CGC, it is a pleasure to work with you all. Dr. Stefan Schmeing, Gulshan Amrahova and Dr. Joseph Openy, thank you for being good collaborators on several projects. Dr. Stefan Schmeing, Dr. Jessica Nowacki-Hansen, Dr. Rachel O'Dea, Dr. Lydia Borgelt, Kai Gallant, and Yang Liu, who were always bombarded by my naive biological questions, thank you all for your passion and always welcome me for a discussion. Dr. Pascal Hommen, Dr. Jimin Hwang, Dr. Joseph Openy, Dr. Adrian Krzyzanowski, and Dr. Sunit Pal, thank you for spending so much time helping me solve the technical problems with HPLC, especially the strange errors that randomly appeared during my operations.

Special thanks are given to Prof. Jim-Min Fang, the supervisor of my bachelor's and master's research, for helping me polish the document to apply for the Government Scholarship for Overseas Study. I would also like to thank the Ministry of Education of Taiwan for the scholarship, and Taipeh Vertretung in der Bundesrepublik Deutschland for their support. I would also like to thank Dr. Huei-Ru Wu for introducing me to Dortmund. I am also grateful to Christa Hornemann for her help with all the paperwork to pursue my Ph.D. in Germany. It is so nice to have your support in communicating with the Stadthaus.

I would like to thank Dr. Jan-Erik Hoffmann and the team at the Dortmund Protein Facility for their help in obtaining the purified proteins. I would also like to thank Dr. Raphael Gasper-Schönenbrücher for his help with protein crystallization and X-ray analysis of the protein crystal structure. I would like to thank Prof. Daniel Summerer for the capacity of their FACS and Damian Schiller for operating the instrument and fruitful discussions on data analysis.

To the party team, Alejandro, Stefan, Pascal, Jimin, Jessica, Rachel, Kai, Lydia, Mirko, Joseph, Kim, Georg, and Philipp and so many others, thank you all for creating so many phenomenal memories with me. Cheers!

I would also like to thank the friends I met in Germany, including but not limited to Huei-Ru, Walter, Kuan-Hsun, Eva, Reika, Simon, Yu-Hsuan, Jeff, Yu-An, Shih-En, and Sebastian. They always know where the best restaurants are and how to have a good time.

Last but not least, this whole journey would not be possible without my wife Tzu-Chen, my parents, and my close family members. I love you all.

Publications

Parts of this work are published in:

Jen-Yao Chang, Cora Neugebauer, Stefan Schmeing, Gulshan Amrahova and Peter 't Hart. "Macrocyclic peptides as inhibitors of WDR5–lncRNA interactions." *Chem. Comm.*, **2023**, *59*, 10656-10659
DOI: https://doi.org/10.1039/D3CC03221C

Jen-Yao Chang, Cora Neugebauer, Anne Mues gen. Koers and Peter 't Hart. "Small molecule WDR5 inhibitors down-regulate lncRNA expression." (Submitting)

Other Publications:

Schmeing, S., Amrahova, G., Bigler, K., **Chang, J.-Y.**, Openy, J., Pal, S., Posada, L., Gasper, R. and 't Hart P. "Rationally designed stapled peptides allosterically inhibit PTBP1 RNA-binding." *Chem. Sci.* **2023**, *14*, 8269–8278
DOI: https://doi.org/10.1039/D3SC00985H

Openy, J., Amrahova, G., **Chang, J.-Y.**, Noisier, A., ‘t Hart, P. "Solid-phase peptide modification via deaminative photochemical Csp3-Csp3 bond formation using Katritzky salts." *Chem. Eur. J.,* **2022***,* 28:e202201121.
DOI: https://doi.org/10.1002/chem.202201121

Table of Content

Abbreviations

2DCS	Two-dimensional combinatorial screening
3D	Three-dimensional
5mC	5-methyl cytosine
Ac	Acetyl
Ac_2O	Acetic anhydride
ACN	Acetonitrile
Alloc	Allyloxycarbonyl
AML	Acute myeloid leukemia
aq	Aqueous
Boc	tert-Butyloxycarbonyl
Cbz	Carbobenzoxy
CLIP	Cross-linking and immunoprecipitation
fCLIP	formaldehyde CLIP
UV-CLIP	ultraviolet CLIP
COSY	homonuclear correlation spectroscopy
CryoEM	Cryogenic electron microscopy
Dab	L-2,4-diamminobutyric acid
Dap	L-2,3-diamminopropionic acid
DCM	Dichloromethane
DMF	*N,N*-Dimethylformamide
DNA	Deoxyribonucleic acid
DMEM	Dulbecco's Modified Eagle Medium
EMSA	Electrophoretic mobility shift assay
ESCC	Esophageal squamous cell carcinoma
FITC	Fluorescein isothiocyanate
Fmoc	9-Fluorenylmethoxycarbonyl
Fmoc-PEG	8-(9-Fluorenylmethyloxycarbonyl-amino)-3,6-dioxaoctanoic acid
FP	Fluorescence polarization
HF	Hydrogen fluoride
HMBC	Heteronuclear Multiple Bond Correlation
HOTTIP	**HO**XA **t**ranscript at the distal **tip**
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
HTS	High throughput screening
HSQC	Heteronuclear Single Quantum Coherence
IgH	Immunoglobulin heavy chain

IgL	Immunoglobulin light chain
IPTG	isopropyl-beta-D-thiogalactopyranoside
K_D	Dissociation constant
K_I	Inhibition constant
ITC	Isothermal titration calorimetry
LC	Liquid chromatography
MS	Mass spectrometry
m/z	Mass-to-charge ratio
MeOH	Methanol
MLL	Mixed lineage leukemia
MAX	MYC associated factor X
MYC	Myelocytomatosis viral oncogene homologue
NMR	Nuclear magnetic resonance
NRT	No reverse transcriptase control
NSL	Non-specific lethal
NTC	No template control
nt	Nucleotide
NuRD	Nucleosome remodeling and deacetylase
o-Ns	2-Nitrobenzensulfonyl
o-NsCl	2-Nitrobenzensulfonyl Chloride
PI	Propidium iodide
PPI	Protein-protein interactions
RBBP5	Retinoblastoma-binding protein 5
RIP	RNA immunoprecipitation
RNA	Ribonucleic acid
lncRNA	long non-coding RNA
lincRNA	long intergenic non-coding RNA
mRNA	messenger RNA
miRNA	microRNA
ncRNA	non-coding RNA
rRNA	ribosomal RNA
siRNA	short interfering RNA
tRNA	transfer RNA
RNAi	RNA interfering
RPI	RNA-protein interactions
SAR	Structure-activity relationship
SPPS	Solid phase peptide synthesis
TFA	Trifluoroacetic acid

TIPS	Triisopropylsilane
TSS	Transcription start site
WBM	WDR5-Binding Motif
WDR5	WD repeat-containing protein 5
WIN	WDR5-Interacting
WRAD	WDR5-RBBP5-ASHL2-DPY30
XCI	X-Chromosome Inactivation
Xist	X-inactive-specific transcript

Abstract

WD repeat domain 5 (WDR5) is a scaffold protein involved in protein-protein or RNA-protein complexes, and most of these complexes play an important role in various epigenetic modulation processes. In particular, some of the long non-coding RNAs (lncRNAs) rely on the formation of the lncRNA-WDR5 complex to exert their epigenetic modulation, such as the upregulation of the lncRNA itself. If an oncogenic lncRNA relies on this pathway to maintain its expression level, it is possible to inhibit its positive feedback loop and thus reduce the expression of the oncogenic lncRNA. Therefore, this thesis focuses on investigating the potential of targeting lncRNA-WDR5 interactions, followed by evaluating the downregulation of lncRNA *in cellulo*.

Targeting lncRNA-WDR5 interactions could be achieved by designing an inhibitor that targets the same binding pocket on WDR5. Several lncRNA are reported to recognize WDR5-binding motif (WBM) site, as a result, a study of structure-activity relationship of peptide-based inhibitors derivatize from protein-WDR5 interactions to target WBM site are performed. Further optimizations are performed by tailor-design macrocycle structure to facilitate peptides adopt the binding conformation and indeed the binding affinity is improved. The ability of macrocycle to disrupt lncRNA-protein interaction is verified by competitive *in vitro* RNA immunoprecipitation (*iv*-RIP).

Cellular experiments are performed to determine whether targeting lncRNA-WDR5 interactions leads to downregulation of the lncRNA itself. Several strategies show their ability to enhance cell uptake of macrocycles. Finally, two molecules show that targeting lncRNA-WDR5 complexes can lead to reduction of the lncRNA itself and that different lncRNA have different sensitivity to the treatment.

Zusammenfassung

Die WD-Repeat-Domäne 5 (WDR5) ist ein Gerüstprotein, das an Protein-Protein- oder RNA-Protein-Komplexen beteiligt ist. Die meisten dieser Komplexe spielen eine wichtige Rolle in verschiedenen epigenetischen Modulationsprozessen. Insbesondere einige der langen nicht-kodierenden RNAs (lncRNAs) sind auf die Bildung des lncRNA-WDR5-Komplexes angewiesen, um ihre epigenetische Modulation auszuüben, wie z.B. das Hochregulieren der lncRNA selbst. Wenn eine onkogene lncRNA auf diesen Weg angewiesen ist, um ihr Expressionsniveau aufrechtzuerhalten, ist es möglich, die positive Rückkopplungsschleife zu hemmen und so die Expression der onkogenen lncRNA zu reduzieren. Daher konzentriert sich diese Arbeit auf die Untersuchung des Potenzials einer gezielten Beeinflussung der lncRNA-WDR5-Interaktionen, gefolgt von einer Evaluierung der Herunterregulierung der lncRNA in der Zelle.

Eine gezielte Beeinflussung der lncRNA-WDR5-Interaktionen könnte durch die Entwicklung eines Inhibitors erreicht werden, der auf die gleiche Bindungstasche in WDR5 abzielt. Es wurde berichtet, dass mehrere lncRNAs das WDR5-Bindemotiv (WBM) erkennen. Daher wurde eine Studie zur Struktur-Wirkungs-Beziehung von Peptid-basierten Inhibitoren durchgeführt, die von Protein-WDR5-Interaktionen abgeleitet sind und auf das WBM abzielen. Weitere Optimierungen werden durch eine maßgeschneiderte Makrozyklusstruktur erreicht, die es den Peptiden erleichtert, die Bindungskonformation einzunehmen, wodurch die Bindungsaffinität verbessert wird. Die Fähigkeit des Makrozyklus, die lncRNA-Protein-Interaktion zu unterbrechen, wird durch kompetitive *in vitro* RNA-Immunpräzipitation (*iv*-RIP) getestet.

In Zellexperimenten wird untersucht, ob die gezielte Beeinflussung von lncRNA-WDR5-Interaktionen zu einer Herunterregulierung der lncRNA selbst führt. Mehrere Strategien zeigen ihre Fähigkeit, die Aufnahme von Makrozyklen in Zellen zu verbessern. Schließlich zeigen zwei Moleküle, dass das Targeting von lncRNA-WDR5-Komplexen zu einer Reduktion der lncRNA selbst führen kann und dass verschiedene lncRNAs unterschiedlich empfindlich auf die Behandlung reagieren.

Chapter 1. General Introduction

1.1 Central Dogma of Molecular Biology and Epigenetics

In 1965, Nobel Laureate James Watson described the central dogma of molecular biology as having three major parts, starting with the deoxyribonucleic acid (DNA), which records all the information on the gene and inherits the information during division and replication, like a blueprint for cells. Through the process of transcription, DNA can transfer its information to ribonucleic acid (RNA), which serves as an interface for information processing, e.g., alternative splicing, allowing for faster and more flexible responses to stimuli or changing conditions. Through the process of translation, RNA can transfer the processed information into protein, which carries out the effect of the gene and interacts with the environment. Despite the existence of various exceptions, such as ribosomal RNA (rRNA) and transfer RNA (tRNA), this sequential decoding process explains the central dogma of molecular biology. Although rRNA and tRNA do not carry the code themselves, their role in the decoding process for transcription and translation does not contradict with the general description of the central dogma.

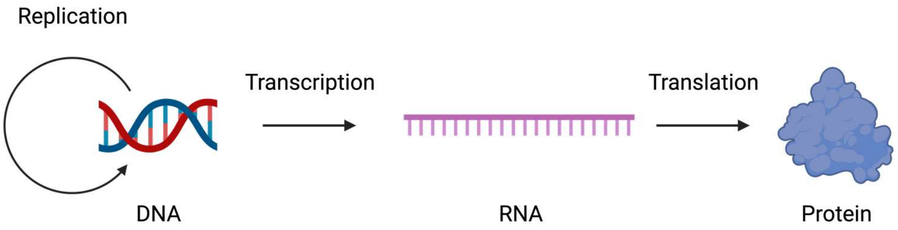


Figure 1. James Watson's illustration of the central dogma of molecular biology. Created with BioRender.com

However, the model of Watson's central dogma model cannot fully explain the difference between genotype and phenotype. For example, the number of genes encoded in the human genome is similar to that of Drosophila, but the diversity of human phenotypes at the cellular level is much greater. In addition, the same genomic information is shared by all somatic cells, but more than 220 different cell types have been reported to have unique functional and protein expression patterns[1]. Furthermore, a "g-value paradox" arises from the fact that the number of coding genes in *Homo sapiens* (approximately 3 x 10^{31} somatic cells, 31000 protein-coding genes) and *Caenorhabditis elegans* (approximately 1 x 10^{3} cells) are very similar (approximately 19000 protein-coding genes)[2,3] even though they are organisms of very different complexity.

Therefore, a new field focused on discovering the mechanism of modulation between genotype and phenotype began to emerge. In the 1990s[4], scientists began to call this field "epigenetics", a term derived from the word "epigenotype", which was introduced by Conrad Hal Waddington in 1942 to describe the concept that chromatin could interact with the environment and change the phenotype as a response[5]. The precise definition of epigenetics was provided at a meeting at Cold Spring Harbor in 2008, where it was stated that "An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence."[6].

1.2 The Molecular Basis of Epigenetics

A human diploid cell contains about six billion nucleotides, which in its linear form would be about 2 m long[7]. To fit into the compact environment of a somatic cells, which have an average radius of 20 µm[8], DNA is wrapped around histones to form the basic repeating unit called a nucleosome, which can be further condensed into a chromosome. Each nucleosome consists of 147 bp of DNA and a histone octamer composed of H2A, H2B, H3, and H4[9–11]. When the nucleosomes are packed in the condensed formation, it is called heterochromatin, and it is transcriptionally inactive due to lack of accessibility to the transcription machinery and flexibility to unwind for transcription to occur (Figure 2)[12]. Therefore, the nucleosome can be unwound into a loose conformation, euchromatin, which is transcriptionally active because the conditions are ready for transcription to occur[13]. The transition between heterochromatin and euchromatin is called chromatin remodeling[13]. This is typically controlled by changing the strength of the interaction between DNA and the histone complex, specifically by changing the strength of electrostatic interactions or hydrophobic interactions by modifying the chemical structure of their interacting surface, e.g., post-synthetic methylation on DNA or acetylation, methylation, and phosphorylation on histones[14].

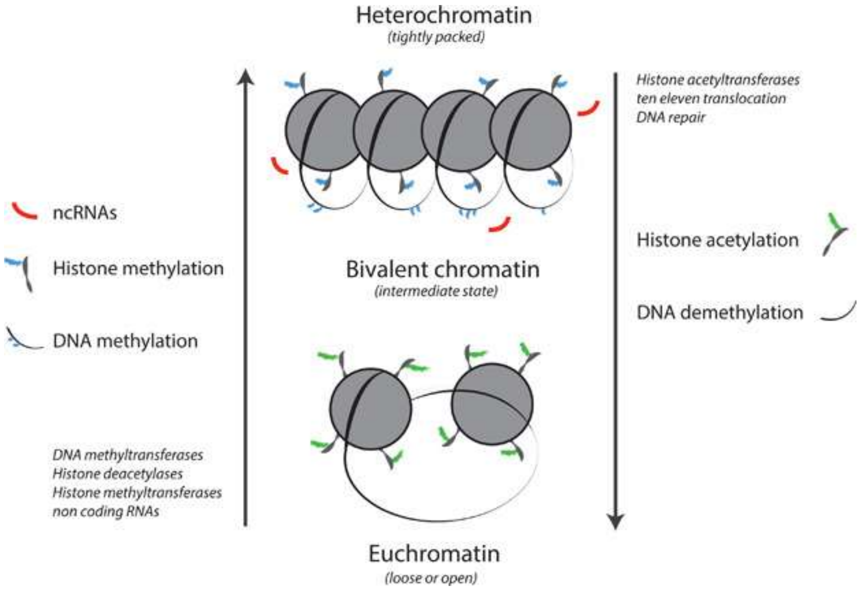


Figure 2. Illustration of nucleosome packing and control of chromatin remodeling[15]. Figure is reprinted with permission (6.7) from (Brander et al., 2017).

Modulation by post-translational modifications (PTM) on histones and the corresponding influence on chromatin remodeling is very well studied[16]. Briefly, after PTM on histones loosen the nucleosome, an ATP-dependent chromatin-remodeling complexes would be recruited and initiate the remodeling process of the chromatin instructed by the PTM markers[16]. Gene activation is typically accompanied by chromatin remodeling into euchromatin through histone acetylation and specific histone methylation (e.g., H3K4me3)[17]. On the other hand, gene silencing is typically achieved by packaging the nucleosome into more condensed heterochromatin through histone deacetylation or histone methylation at H3K27me3 (Figure 3)[18,19].

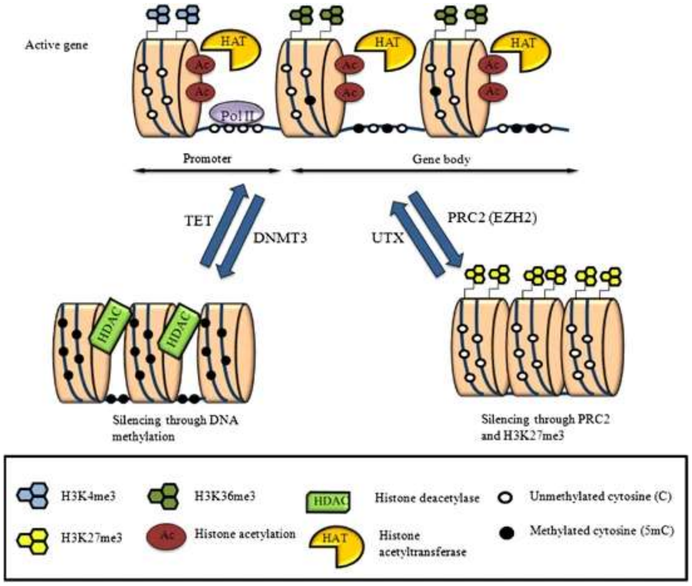


Figure 3. Common events for chromatin remodeling between euchromatin and heterochromatin[18]. Figure is reprinted with permission (6.7) from (Dimopoulos et al., 2014).

In addition to modulation by histone acetylation/methylation, post-synthetic modification of DNA also modulates gene expression. Among the four DNA bases, cytosine is the most abundant methylated DNA observed in mammalian cells and approximately 2-7% of cytosine is observed as 5-methylcytosine (5mC)[20,21]. Methylation of CpG dinucleotides typically results in regional silencing and the differential methylation pattern has been shown to play an important role in stem cell development and differentiation[22–24].

Epigenetic modulators can be categorized into three main functions: writers, erasers, and readers (Figure 4)[25]. Writers are able to mark sequence (DNA) or location (histone) through chemical modifications, including methylation on histone or DNA, or acetylation on histone. In contrast, erasers, named because they have the opposite effect of writers, are able to neutralize the effect of markers by reversing the post-translational/synthetic modification made by writers. Readers are the components that read the state of the marker and thus carry out the indication of the label.

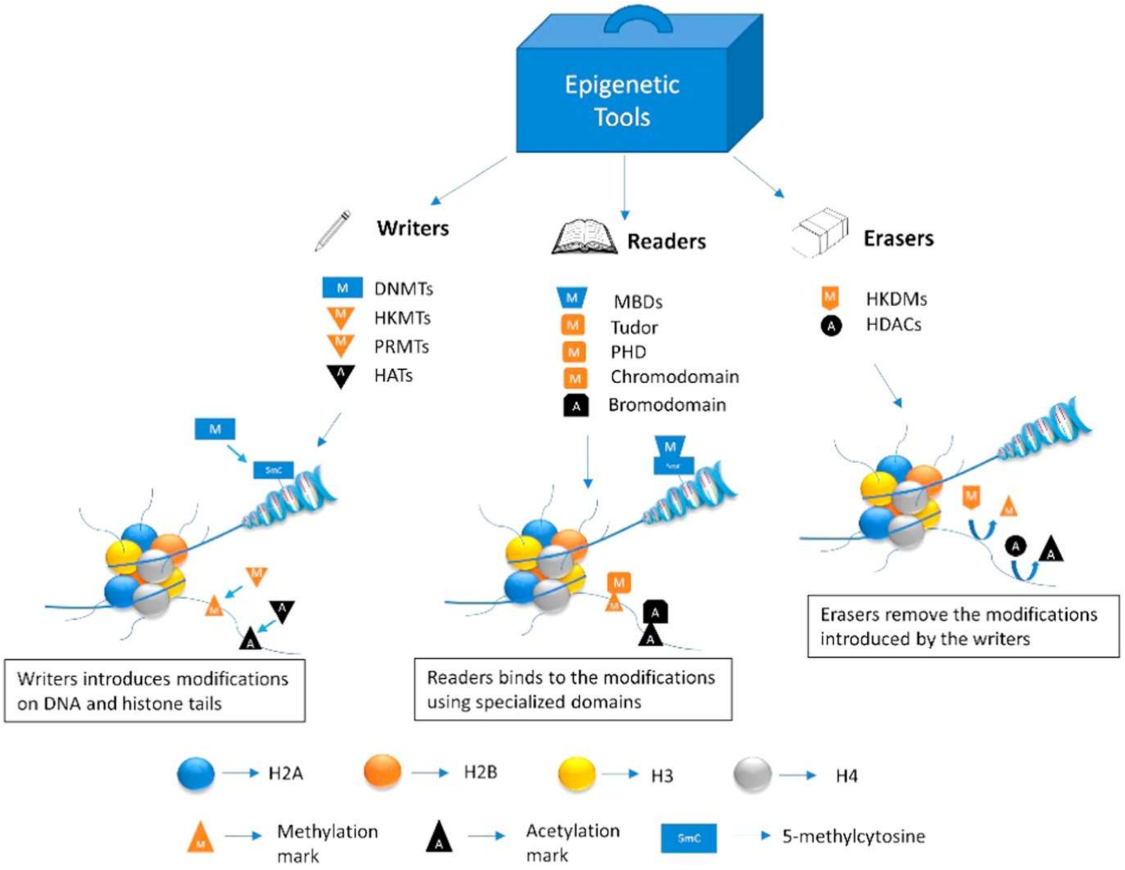


Figure 4. Tools in epigenetics, categorized by function[25]. Figure is reprinted with permission (6.7) from (Biswas et al., 2018).

1.3 RNA in Molecular Biology

RNA and DNA are very similar in their structure and purine/pyrimidine bases, which is reflected in their ability to carry the information from DNA for translation into protein. However, the additional hydroxyl group at C2 also makes RNA more hydrophilic, resulting in a more flexible linear structure[26]. Unlike DNA, which is known for its double-stranded helical structure, RNA is typically single-stranded and folded by intramolecular base pairing[27]. This ability to maintain a three-dimensional structure is critical for RNA to possess the catalytic activity in several biochemical reactions, e.g. for the ribosome to catalyze the translation process (Figure 5)[28].

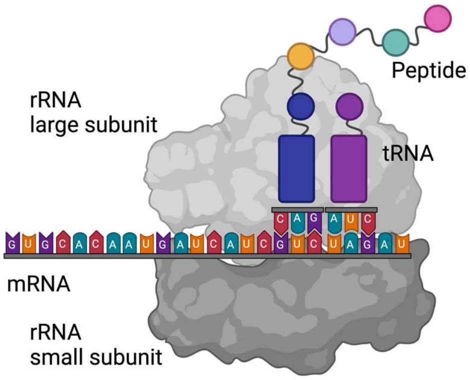


Figure 5. rRNA, tRNA and mRNA in the process of translation. Created with BioRender.com

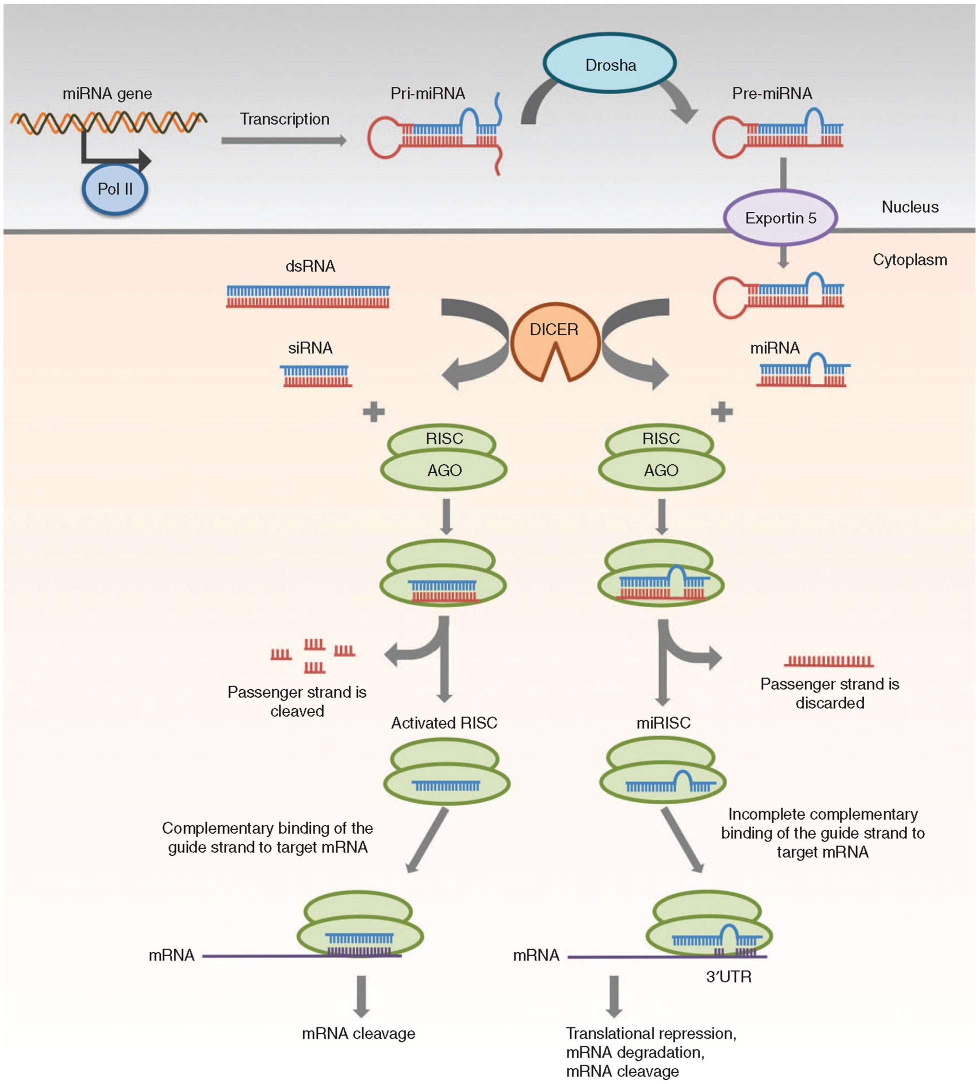


Figure 6. Typical functions of miRNA and siRNA in RNAi[29]. Figure is reprinted with permission (6.7) from (Lam et al., 2015).

In the central dogma of molecular biology, three major types of RNAs are commonly mentioned: messenger RNA (mRNA), which carries the genetic information for translation; ribosomal RNA (rRNA), which can be assembled into the ribosome and is the essential component in the translation process; and transfer RNA (tRNA), which carries the corresponding amino acid according to its codon and provides it during protein synthesis (Figure 5).

1.3.1 Synthesis and Functions of non-coding RNAs

Those RNAs that do not contain the information for translation into protein are categorized as non-coding RNA (ncRNA). With the exception of mRNA, all other RNAs belong to this category. In fact, their synthesis is different in eukaryotes. The rRNA 45S, which matures into 28S, 18S and 5.8S, is typically synthesized by RNA polymerase I, whereas rRNA 5S, tRNA and other small RNA (sRNA, < 200 nt) are synthesized by RNA polymerase III[30,31]. miRNAs typically originate from RNA polymerase II synthesis, either by direct synthesis or as a product of the spliced region during mRNA maturation[32]. 40% of miRNA can be located within introns or exons of other genes and experience the same modulation of the host gene[33].

In addition to the well-known function of rRNA and tRNA in translation, the other ncRNAs are beginning to attract attention for their role in post-transcriptional modification. For example, microRNA (miRNA) is an endogenous single-stranded RNA consisting of approximately 19 to 25 nucleotides in sequence[34]. In contrast, short interfering RNA (siRNA) is typically exogenous, but has greater target specificity compared to miRNA in RNAi[35]. To carry out RNAi, both miRNA and siRNA can recognize and recruit the complementary RNA, typically mRNA, to the RNA-induced silencing complex (RISC) and lead to translational repression either by directly degrading the mRNA or by preventing the binding of translation initiation factors (Figure 6)[36].

1.3.2 Long non-coding RNA

Initially, an arbitrary definition of lncRNA was that an ncRNA contains more than 200 nt in sequence[2]. Despite the reported existence of lncRNAs such as *H19* and X-inactive-specific transcript (*Xist*), they were considered curiosities rather than a general component of RNA biology[2,37]. With the advent of next-generation sequencing, transcriptomics is blossoming, and during data analysis, scientists realized that the importance and expression levels of lncRNAs that are spliced and polyadenylated have been underestimated. In fact, the data show that mammalian cells transcribe much more lncRNA than the total amount of protein-coding genes, and the difference is about an order of magnitude or more[37,38]. In 2023, a new definition is proposed that a non-coding RNA has more than 500 nt and is generally synthesized by RNA Pol II to replace the conventional definition of long non-coding RNA (lncRNA)[2]. This new definition considered that many lncRNAs are spliced and polyadenylated, giving them mRNA-like properties. Although several types of ncRNAs are larger than 200 nt, they have other unique functions and do not behave like the current understanding of lncRNA. Therefore, the discussion of lncRNA in this dissertation focuses on those lncRNAs that also fit the updated definition.

Compared to mRNA, lncRNA has a more cell-type specific expression pattern and controls the fate of several tissue stem cells (Figure 7)[39]. In addition, one of the specific types of lncRNA, long intergenic non-coding RNA (lincRNA), shows a tendency to be co-expressed in close proximity to protein-coding genes[40]. As lncRNAs have been identified as essential components in epigenetic regulation and pluripotency of stem cells, the mechanism of action behind the modulation has been extensively studied[41]. The typical function of lncRNA can be divided into six major classes (Figure 8): 1. To guide a chromatin modifier to the target site to facilitate modification at the site; 2. A scaffold molecule to promote complex formation at the target site; 3. A decoy to distract the protein from its binding site; 4. A sponge to lure the miRNA away from its target; 5. A precursor for miRNAs; 6. A scaffold to assist in chromatin looping and modulate long-range gene regulation.

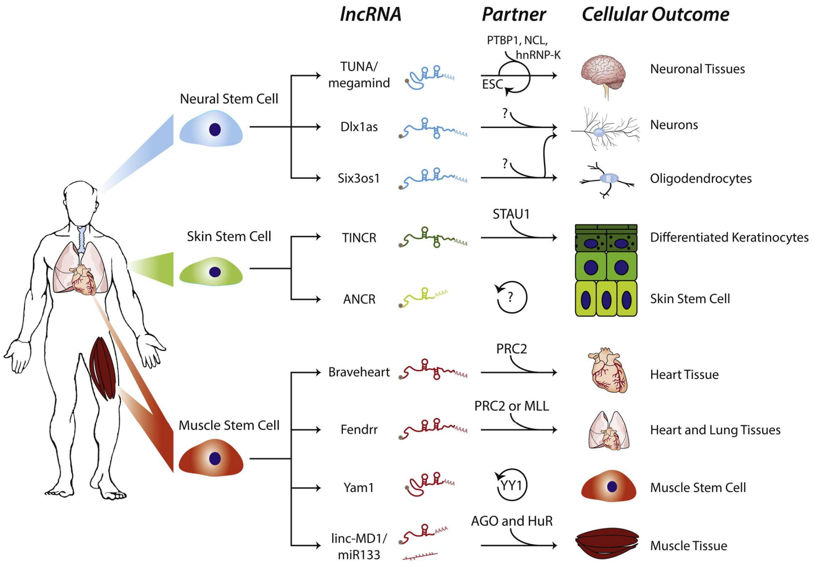


Figure 7. Several somatic stem cells are regulated by lncRNA[39]. Figure is reprinted with permission (6.7) from (Flynn and Chang, 2014).

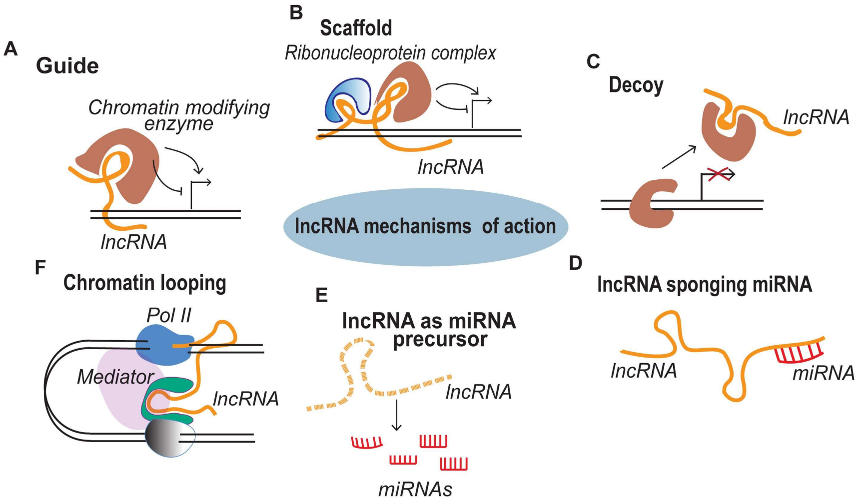


Figure 8. The typical functions of lncRNA[42]. Figure is reprinted with permission (6.7) from (Sweta et al., 2019).

One of the most studied lncRNAs is X-inactive-specific transcript (*Xist*)[43,44]. It is very important in the process of X chromosome inactivation (XCI), which is an essential process for females to compensate the dosage from the second X chromosome[44]. When *Xist* was defected from the X chromosome, the growth of female embryos was retarded[45,46]. Several elements in the sequence of *Xist* have been identified and characterized that allow it to perform multiple actions and control XCI (Figure 9)[47].

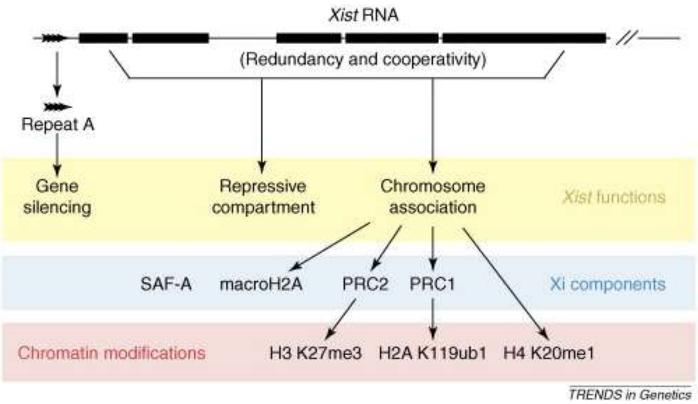


Figure 9. *Xist* RNA and the functions of its components. The arrows indicate the current view of the molecular basis of X chromosome inactivation[47]. Figure is reprinted with permission (6.7) from (Wutz, 2007).

1.4 lncRNA-Protein Interactions

As more and more evidence demonstrates the critical role of lncRNA in epigenetics, especially its role in regulating stem cell development, studying how lncRNA modulates epigenetics becomes very important[48]. In most cases, the functions of lncRNA require the assistance of proteins. Therefore, studies of lncRNA-protein interactions are very important and still a very challenging field. Several experiments demonstrate their ability to study RNA-protein interactions. To characterize the specific RNA-protein interactions, ultraviolet CLIP (UV-CLIP) and formaldehyde CLIP (fCLIP) are the most commonly used methods that allow the identification and characterization of the co-precipitated binding partners[49,50]. Each method has its own strengths and limitations. The UV-activated radical has a very short lifetime, so the cross-linking reaction can only occur in very close proximity. In addition, UV cannot induce protein-protein cross-linking, allowing a more selective cross-linking reaction for analysis. However, these features also result in low cross-linking efficiency and the preference of uridine for cross-linking. On the other hand, fCLIP allows more dynamics between the cross-linking substrates and does not rely on the intrinsic efficiency of UV-activated radical formation. It has a strong preference for lysine residues and also allows protein-protein cross-linking, making it difficult to distinguish whether the RNA is interacting directly with the protein or with the protein complex[51]. Numerous protocols for CLIP methods have been reported for specific research interests (Figure 10)[51]. RNA immunoprecipitation (RIP) can be used to analyze the RNAs that interact with the protein of interest.

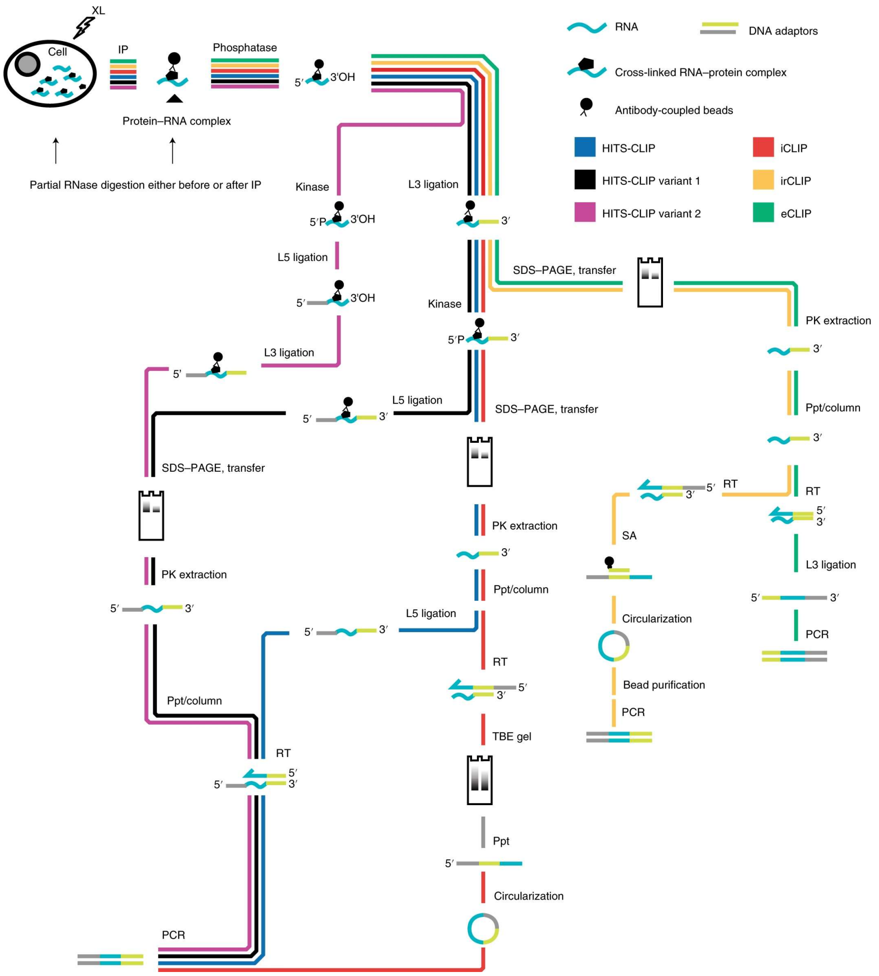


Figure 10. Maps of current strategy in RNA-centric methods for studying protein-RNA interactions[51]. Figure is reprinted with permission (6.7) from (Ramanathan et al., 2019).

1.4.1 Limitations in the Study of lncRNA-Protein Interactions

Although several protein-centered methods have been presented, most of them are designed to identify the binding partners rather than to characterize the interactions. Even existing analysis methods have their limitations and are not directly applicable to lncRNA-protein interactions. The main difficulties lie in the preparation of high-quality lncRNA material with the correctly folded structure in its native conformation. When the lncRNA is transcribed *in vitro*, it may not fold into the native conformation, leading to difficulties in reproducing biophysical experiments[52]. In addition, the lncRNA may lose its structure when the sequence is truncated, limiting the ability to identify the binding elements of the sequence simply by screening fragments of the lncRNA.

An RNA-protein interaction (RPI) is generally not considered a druggable target due to its large interaction surface[48,53]. To rationally design an inhibitor for an enzyme, which usually has a well-defined pocket for orienting the substrate into an orientation suitable for chemical reactions to occur, one can always start by mimicking its natural substrate. However, an RPI usually involves numerous residues both on the protein and RNA side, and mimicking the entire interaction surface with a small molecule is not feasible.

1.5 *HOTTIP*-WDR5 Interactions

HOXA transcript at the distal tip (*HOTTIP*) is a lincRNA discovered by Wang et al. in 2011[54]. As its name suggests, *HOTTIP* is located at the 5' end of the *HOXA* locus and was initially recognized for its role in the upregulation of the late *HOXA* gene cluster and the relationship between cell development. Similar to typical lincRNAs, *HOTTIP* possesses multiple modulatory functions in epigenetics. More importantly, most of these modulations are highly correlated with the modulation of cell pluripotency and development, the common features that are also required for cancer cells. For example, *HOTTIP* is required for gastric cancer to recruit the WDR5-MLL complex to the 5' end of the HOXA gene cluster[54], leading to the upregulation of HOXA13 and thus promoting the tumorigenic properties of the cells[55]. A similar dependence of HOXA13 on *HOTTIP* has also been reported in esophageal squamous cell carcinoma (ESCC). In addition to activation via recruitment of the WDR5-MLL complex, they demonstrate a parallel pathway for *HOTTIP* to enhance HOXA13 expression by competing with miR-30b for degradation of *HOXA13* mRNA[56]. Furthermore, many clinical reports demonstrate the strong correlation between dysregulation of *HOTTIP* expression and tumor progression, suggesting the oncogenicity of *HOTTIP*.

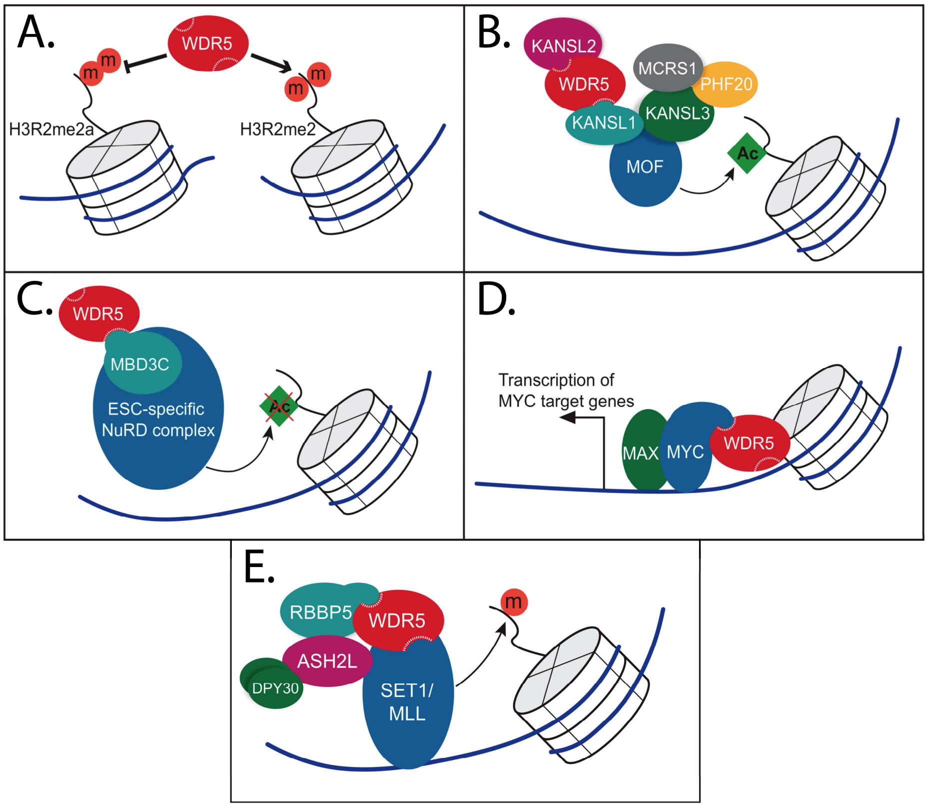


Figure 11. Functions of WDR5 in cells[57]. (A) WDR5 binds selectively to symmetric dimethylated Arg2 on histone H3 tails, whereas asymmetric dimethylation of Arg2 on histone H3 tails prevents recognition. (B) WDR5 is the scaffold protein in the NSL complex that performs histone acetylation. (C) WDR5 is the scaffold protein in an embryonic stem cell-specific form of the NuRD complex that performs the histone deacetylation. (D) WDR5 assists the MYC-MAX complex in chromatin recognition and transcriptional activation. (E) WDR5 is the scaffold protein of the WRAD complex to facilitate the SET1/MLL complex for histone H3K4 methylation. Figure is adapted from (Guarnaccia and Tansey, 2018) with permission (6.7).

WD repeat-containing protein 5 (WDR5) belongs to the WD repeat protein family and contains seven WD repeats to form an iconic β-propeller shape. WDR5 is involved in several epigenetic modulation complexes such as the WDR5-RBBP5-ASHL2-DPY30 (WRAD) complex, the non-specific lethal (NSL) complex, the nucleosome remodeling and deacetylase (NuRD) complex, and the MYC-MAX complex (Figure 11. B - D)[57–61]. In addition to its role as a scaffold protein, WDR5 is able to discriminate between asymmetric and symmetric dimethylation of Arg2 on the H3 tail (H3R2me2a/H3R2me2) and is therefore also considered to be an epigenetic reader (Figure 11. A) Interestingly, the selective recognition of H3R2me2 (a marker for euchromatin) while being unable to bind H3R2me2a (a marker for heterochromatin) reflects the nature of WDR5 as an epigenetic reader responsible for gene activation[57,62,63]. WDR5 contains two major binding sites, the WDR5-Interacting (WIN) site and the WDR5-Binding Motif (WBM) site, each of which has a specific binding partner while being located on the opposite side of the surface, allowing it to accommodate two proteins simultaneously in the large protein complex (Figure 12).

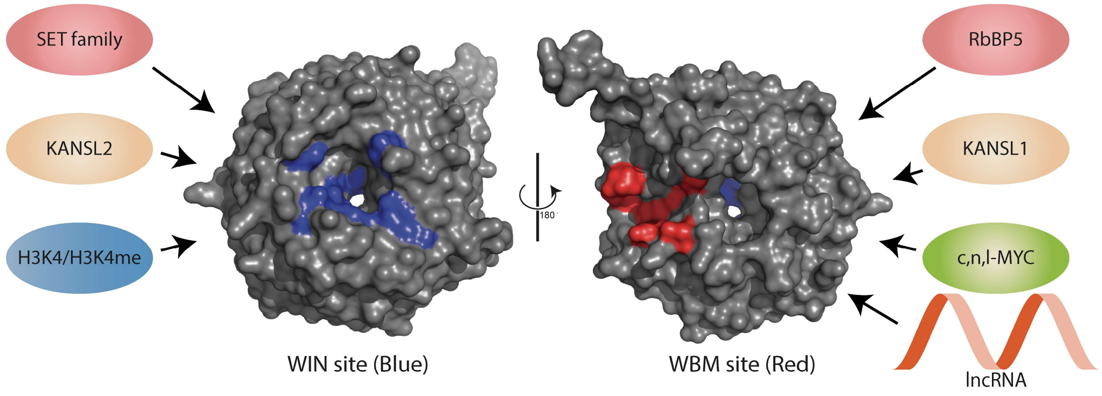


Figure 12. Cocrystal structure of WDR5 (PDB: 8Q1N) and corresponding binding partners of WIN and WBM sites[58,64,65]. A65, S91, D107, F133, Y191, Y260, F263 are colored in blue to indicate the location of the WIN binding pocket; N225, Y228, L240, F266, V268, Q289 are colored in red to indicate the location of the WBM binding pocket[57].

Since the first report of *HOTTIP*, the importance of WDR5 for *HOTTIP* to carry out its epigenetic modulation has been evaluated[54]. For *HOTTIP* to maintain the late *HOXA* gene cluster in an active state, it requires the involvement of WDR5-MLL1 to maintain the methylation level of H3K4me3 at the transcription start site (TSS) around these *HOXA* genes. The direct interaction between *HOTTIP* and WDR5 was demonstrated by several pull down experiments. Furthermore, knockdown of WDR5 leads to a general silencing of late *HOXA* clusters, including *HOTTIP* itself. In conclusion, WDR5 is essential for *HOTTIP* to maintain active chromatin around the late HOXA gene clusters.

1.5.1 The *HOTTIP*-WDR5 Interactions as a Therapeutic Target

Clinical studies have identified several cancers that overexpress *HOTTIP*, and knockdown of *HOTTIP* could reduce the tumorigenicity of cancer cells, including retinoblastoma[66], esophageal squamous cell carcinoma (ESCC)[56], acute myeloid leukemia (AML)[67], breast cancer cells (e.g., MDA-MB-231)[68] and osteosarcoma (e.g., U-2 OS)[69,70]. These data suggest that reducing the aberrant expression of HOTTIP could reduce cancer aggressiveness, suggesting a potential novel therapeutic strategy for cancer (Figure 13).

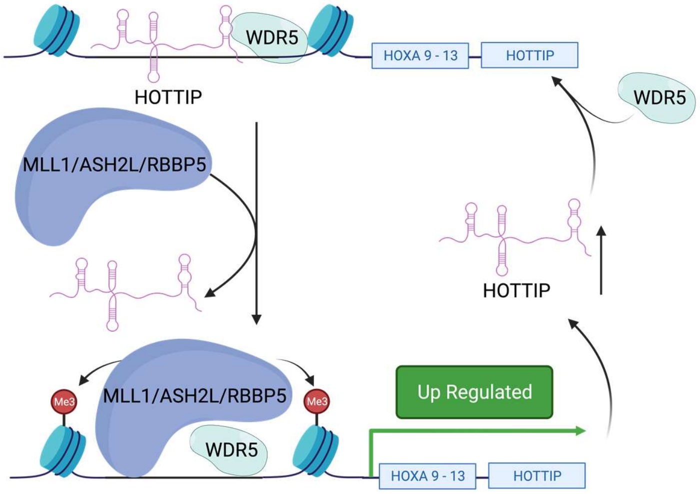


Figure 13. Positive feedback loop of *HOTTIP* to maintain its own expression level. Created with BioRender.com.

To reduce *HOTTIP* expression levels, siRNA is the most straightforward method.

However, bioavailability and selectivity between normal and cancer cells prevent its practical application[71]. An alternative strategy is to target the epigenetic modulation system. In our hypothesis, direct inhibition of *HOTTIP*-WDR5 interaction should prevent recruitment of WDR5 to the target chromatin, thereby preventing the H3K4me3 formation and leading to downregulation of *HOTTIP* (Figure 14).

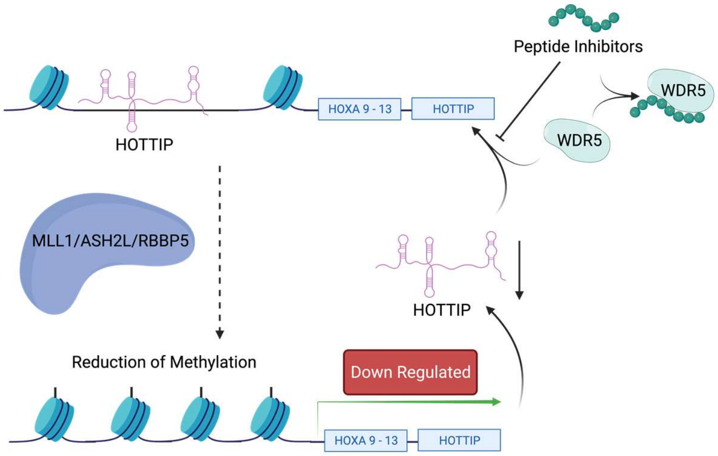


Figure 14. Hypothetical effect of inhibiting *HOTTIP*-WDR5 to prevent recruitment of WDR5, resulting in reduced levels of H3K4me3 and thus reduced levels of *HOTTIP* expression. Created with BioRender.com.

1.5.2 Targeting Strategy

To prevent the interaction between *HOTTIP* and WDR5, one can design an inhibitor that targets either *HOTTIP* or WDR5 to disturb the formation of the complex. The major difficulty in designing a *HOTTIP* inhibitor lies in the very limited information available for rational design as it is still unclear which part of *HOTTIP* is responsible for recognizing WDR5[52]. In addition, the issue of reproducibility of *in vitro* transcribed *HOTTIP* in biophysical assays also discouraged us from screening for *HOTTIP* ligands[52]. As a result, WDR5 was selected as our main target to initiate the project.

WDR5 typically acts as a scaffold protein to support complex formation, and the interaction between protein-WDR5 complexes has been well studied. Thanks to the high tendency of WDR5 to crystallize, numerous co-crystal X-ray analysis data are available, revealing the critical interactions for binding to occur. In addition to being well studied, the binding site on WDR5 responsible for recognition of *HOTTIP* has been identified and characterized. By screening *HOTTIP* and the selective mutation series of WDR5, the WBM site was identified as the binding pocket for *HOTTIP* (Figure 12)[65]. Therefore, targeting the WBM site became our focus.

1.5.2.1 Peptide Base Inhibitor for the WBM Site

The WBM site is a wide but shallow cleft between two beta propeller blades of the WD40 repeat. Numerous protein binding partners capable of recognizing the WBM site, have been identified and the minimal sequence for recognition of the WBM has been analyzed (Figure 15). To target the WBM site, derivatization of these linear peptide sequences is the most straightforward approach.

Proteins	***Sequence***	***K_D (μM)***	***Method***
RBBP5	371 380 ED**EEVDVT**SV	2.2 ± 0.2	ITC
KANSL2	406 417 EFS**DDLDVV**GDG	8.6 ± 0.3	ITC
c-MYC	258 268 DE**EEIDVV**SVE	9.3 ± 1.7	FP

Figure 15. The reported WBM binding sequences and the corresponding K_D[57,59,60,72]. The sequences are aligned and identical residues are shown in red, while residues with similar characters are shown in blue[57]. The dissociation constants (K_D) were determined by isothermal titration calorimetry (ITC) or fluorescence polarization (FP).

Solid phase peptide synthesis (SPPS) is well established due to its rapid and robust result in obtaining high quality peptide without extensive purification processes. SPPS was first reported by R. B. Merrifield in 1963 using the carbobenzoxy (Cbz) strategy[73] which was later modified to the more widely used fluorenylmethoxycarbonyl (Fmoc) strategy to facilitate deprotection and avoid the use of hydrogen fluoride (HF). The simplified reaction protocol makes it suitable for use on an automated synthesizer, further reducing the workload required for routine synthesis (Figure 16). It is also possible to selectively deprotect the side chain with an orthogonal protection strategy, allowing the scientist to perform more complex structure-activity relationship (SAR) studies. For example, macrocyclization is a common strategy to mimic the 3D conformation of peptides derived from protein-protein interactions. Unlike linear peptides, which are flexible in aqueous solution and would leave the binding pocket for a higher entropy state, macrocyclic peptides are more conformationally constrained and therefore benefit from reduced entropy loss upon binding[74]. In addition, the unnatural cyclization makes them less accessible to the proteases, thus prolonging the half-life of the peptides[74,75].

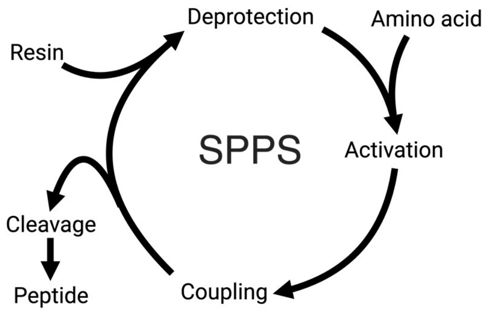


Figure 16. The reaction cycle of solid phase peptide synthesis (SPPS). Created with BioRender.com.

Peptide drugs have been used by society for more than 100 years. The first commercialized peptide drug prescribed to patients is insulin, discovered in 1921 for the treatment of diabetes, and today more than 80 peptide drugs have been approved worldwide[76]. Since 2000, thirty-seven non-insulin-based peptide drugs have been approved, and they are no longer limited to human hormone mimetics or composed only of natural amino acids[76,77]. The success of semaglutide, branded as Wegovy® and prescribed for diabetes and weight loss, has made its owner, Novo Nordisk, the most valuable company in Europe[78].

1.5.2.2 Small Molecule Inhibitors

Compared to peptide inhibitors, small molecule inhibitors are generally considered to be more stable in the cytoplasm. In addition, several prodrug strategies could be applied when the molecule has issues with cell permeability or serum stability. However, small molecule targeting of protein-protein interactions (PPIs) or RPIs has traditionally been considered a very difficult task due to size limitations and conformational rigidity. The typical contact surface area for a PPI (1500 - 3000 Å^2) is significantly larger than the solvent-accessible surface area for small molecules (approximately 500 Å^2 for 500 Da)[79–81]. To overcome the surface area limitation, a "hot-spot" strategy is developed. Specifically, it involves finding the hot-spot area, typically around 600 Å^2, that contributes the highest free energy to the PPI and designing a molecule that covers these most important interactions[82].

1.6 Aim of this Work

The aim of this research is to investigate the possibility of targeting an RNA-protein interaction to reduce the expression level of oncogenic lncRNA in cells. The *HOTTIP*-WDR5 interaction is chosen as a target because 1. *HOTTIP* upregulates several oncogenic genes[69,83]. 2. *HOTTIP* requires the recruitment of WDR5 to modulate genes[54]. 3. Knockdown of WDR5 reduces *HOTTIP* expression, suggesting that *HOTTIP* expression requires the involvement of WDR5[54]. 4. When WDR5 is mutated to be unable to recognize *HOTTIP*, the stability of WDR5 in the nucleus decreases[65].

There are several questions that need to be answered before we get to the final question. First, we have to develop our own inhibitors because they did not exist when the project started. Is it possible to adapt a protein-protein interaction to inhibit RNA-protein interactions? It is very likely, because several data suggest that *HOTTIP* shares the binding pocket of the WBM site on WDR5, which is also recognized by various proteins, but no one has tested it yet. Once the peptide inhibitors are optimized, they can be used directly as lead compounds, or in high-throughput screening to find more potential candidates.

Once a candidate is found, the next question would be how to demonstrate that the inhibitor can disrupt *HOTTIP*-WDR5 interactions. Several experiments are considered, and two experiments stand out: electrophoretic mobility shift assay (EMSA), which is a well-known experiment for studying RNA-protein interactions with a strong advantage of providing direct and comprehensible results; and *in vitro* RNA immunoprecipitation experiment, which has the advantage of compatibility with total RNA extracts and can differentiate between individual enriched RNAs by RT-qPCR. It is also interesting to test whether there is allosteric control between the two binding sites on WDR5 on RNA-protein interactions and whether different inhibitors lead to different allosteric effects.

In order for cell-based experiments to work, several practical issues need to be addressed prior to performing the experiment on cells. The most straightforward readout would be the *HOTTIP* expression level, and obviously a cell line with *HOTTIP* expression is required. For cellular experiments, the cell membrane is one of the major obstacles for peptide-based inhibitors and a strategy to improve cellular uptake is required. If small molecule inhibitors are found, their cellular uptake should be verified.

Once all the above gaps are filled, we can return to the very first question: Does inhibition of lncRNA-WDR5 lead to decreased expression of the lncRNA itself? Does the lncRNA modulation require direct lncRNA-WDR5 interaction?

Chapter 2. Targeting the lncRNA-WDR5 Interactions with Peptide-Based Inhibitors

In this chapter, a peptide-based macrocycle, designed based on the native protein-protein co-crystal structure, was shown to inhibit lncRNA-WDR5 complex formation *in vitro*. To evaluate the binding constants between the peptide and WDR5, fluorescence polarization (FP) was used as the primary assay, followed by isothermal titration calorimetry (ITC) as the secondary assay. X-ray analysis of the co-crystal structure was performed, and the binding conformation is perfectly aligned with the native linear peptide. *in vitro* RNA immunoprecipitation (*iv*-RIP) is used to demonstrate the ability to inhibit *HOTTIP*-WDR5 complex formation. Some data from this part are published in "Macrocyclic peptides as inhibitors of WDR5-lncRNA interactions" in Chemical Communications[64].

Our initial hypothesis that inhibition of *HOTTIP*-WDR5 interaction in cells could lead to *HOTTIP* downregulation was evaluated with the macrocycles. Several strategies, including conjugation with a cell-penetrating peptide (CPP), transfection reagents, induced macropinocytosis, and electroporation (EP), were applied and did indeed improve the level of cellular uptake. However, none of these strategies resulted in a significant difference between the samples and the negative control.

This part was done in collaboration, and I would like to acknowledge that protein samples were expressed by the Dortmund Protein Facility (DPF), Stefan Schmeing or Cora Neugebauer; ITC was performed by Gulshan Amrahova; X-ray analysis and protein co-crystallization were performed by Dr. Raphael Gasper-Schönenbrücher and Stefan Schmeing; *iv*-RIP was performed by Cora Neugebauer when she did her master thesis under my supervision.

2.1 Introduction and Strategy

The HOTTIP-WDR5 interaction is important for various types of cancer cells. Targeting this interaction could potentially lead to an ideal therapeutic effect[56,66–70]. To achieve this, a binder can be designed to recognize either *HOTTIP* or WDR5 and effectively target the lncRNA-protein interaction. The specific sequence of *HOTTIP* responsible for recognizing WDR5 remains unclear. However, the WDR5-Binding Motif (WBM) site on WDR5 has been shown to play a crucial role in recognizing *HOTTIP*[65]. Therefore, we chose to target WDR5 as the initial candidate for inhibiting complex formation.

2.1.1 Design and Synthesis of the Binding Sequence

The WBM site is a shallow cleft involved in numerous protein-WDR5 interactions Various studies have shown that the consensus sequence for protein-WDR5 recognition falls within the single-digit micromolar (μM) range with binding affinity, as depicted in Figure 15. To investigate the structure-activity relationship (SAR) efficiently, automated solid-phase peptide synthesis (SPPS) and Fmoc strategy were employed to modify the peptide sequence. An acetyl capping step is integrated after each coupling cycle to avoid single point truncation (Figure 17). The fluorophore is conjugated post-synthetically of the sequence.

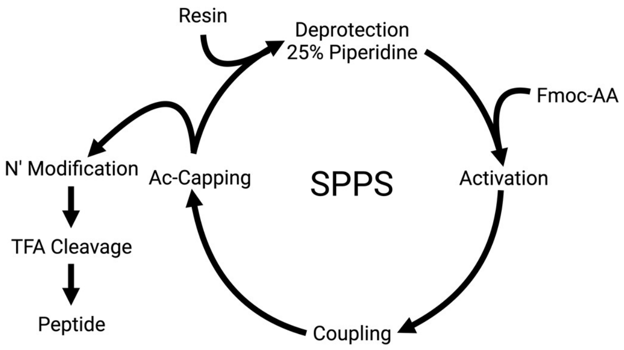


Figure 17. The reaction cycle of the Fmoc strategy for SPPS. Created with BioRender.com.

2.1.2 Primary Screening Method

The FP assay was selected as the primary assay due to its ease of use and potential for integration into a high-throughput screening (HTS) system. Figure 18 provides a brief explanation of how the FP assay operates. When polarized excitation light shines on the fluorophore, the emission light is depolarized due to the fluorophore's rotation. The rotational relaxation time, which is reliant on the molecular weight of the complex, determines the rotation speed. When a fluorophore forms a complex with a protein, its apparent molecular weight increases, resulting in a slower rotation and higher ratio of polarized light emission.

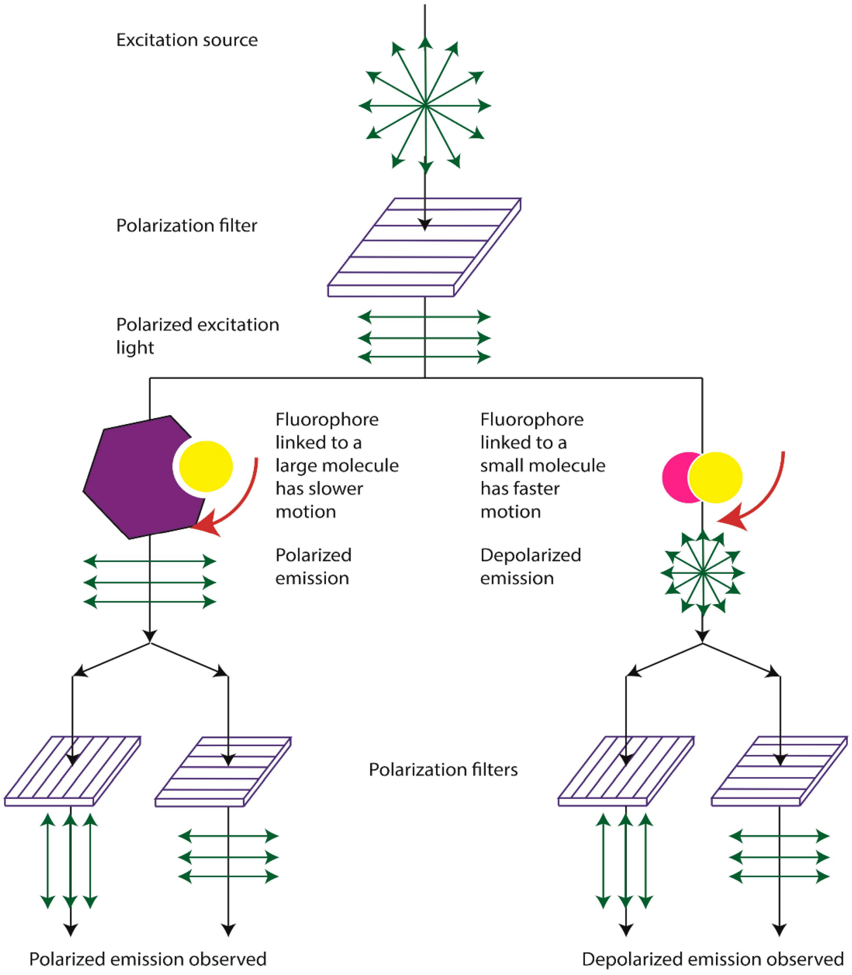


Figure 18. Illustration of the mechanism behind fluorescence polarization (FP)[84]. The figure is reprinted from (Hendrickson et al. 2020) with permission (6.7).

Compared to other conventional assays for measuring binding affinity, FP has several advantages for the current project due to the following reasons: 1. FP does not require modifying the protein for specific functional groups. 2. A fluorophore is required, which is compatible with peptide synthesis. 3. It is compatible with a small volume 384 well plate format. 4. It does not require costly substrates.

To conduct the FP assay, a fluorophore must be applied to the peptide for the measurement purposes. It is necessary to avoid using an acid labile fluorophore as it will not survive in the final stage of SPPS, which involves employing TFA for global cleavage. To label the peptide, fluorescein isothiocyanate (FITC) is used owing to its ease of chemistry and high aqueous stability. Moreover, compared to carboxyfluorescein that requires activation of the carboxylate, the isothiocyanate on FITC is readily coupled with amine. The excitation and emission wavelengths of FITC are frequently used in the green fluorescence protein (GFP) channel, enabling broad compatibility with most detectors.

When performing FP experiments with a labelled peptide, it is necessary to conduct a self-competitive experiment to demonstrate that the binding between the sequence and the target is not influenced by the attached label. To verify that the tracer can be competed away from the binding pocket, a non-labeled sequence will be synthesized and used in a competitive FP experiment.

2.1.3 Secondary Screening Methods

The primary purpose of conducting the secondary assay is to eliminate the potential influence of the sequence on the detection method and prevent false positives. To achieve this, a non-fluorescence-based method is mandatory to avoid a similar bias from FP (Figure 19). In this case, ITC is selected as the secondary assay due to its broad compatibility[85]. In addition, the nature of heat release measurement in ITC allows us to obtain data of K_D

and ΔH, which can be used to elucidate ΔG and ΔS at the given experimental temperature, which is beneficial for characterizing peptide binding properties. However, ITC has some significant drawbacks, for example, the large amount of sample required by ITC is and it is low throughput, which persuades us to use it only as secondary assay.

In addition to the previous biophysical experiment, direct evidence that a given peptide recognizes the WBM site would be desired. X-ray co-crystallization analysis would be performed if the lead sequence co-crystallizes with WDR5 to provide such evidence. Several other experiments, such as hydrogen-deuterium (H/D) exchange followed by nuclear magnetic resonance (NMR) or mass spectrometry (MS), could be used to verify the binding site of the ligand and would be performed if co-crystallization does not occur.

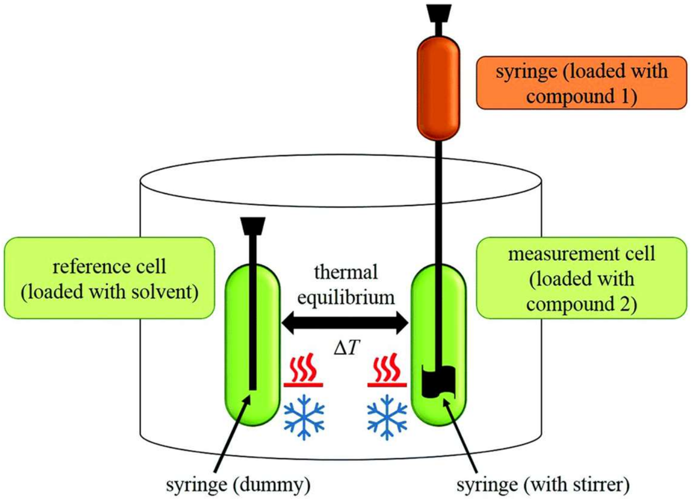


Figure 19. Principle of isothermal titration calorimetry (ITC)[86]. Figure is reprinted with permission (6.7) from (Prozeller et al., 2004).

2.1.4 *in vitro* Validation

We also deemed an orthogonal experiment necessary to demonstrate that an inhibitor is able to disrupt the *HOTTIP*-WDR5 complex before proceeding to the *in cellulo* experiments. A conventional method for analyzing the inhibitory effect on RNA-protein interactions is the electrophoretic mobility shift assay (EMSA). EMSA is typically used to determine the RNA-substrate complex formation because the separation of the analytes is determined by the mass-to-charge ratio (m/z) and the size of the complex. However, EMSA typically uses purified RNA to avoid the background signal from other RNAs. Given the potential reproducibility issues with *in vitro* transcribed *HOTTIP*, this approach was not pursued[52].

2.1.4.1 *in vitro* RNA immunoprecipitation (*iv*-RIP)

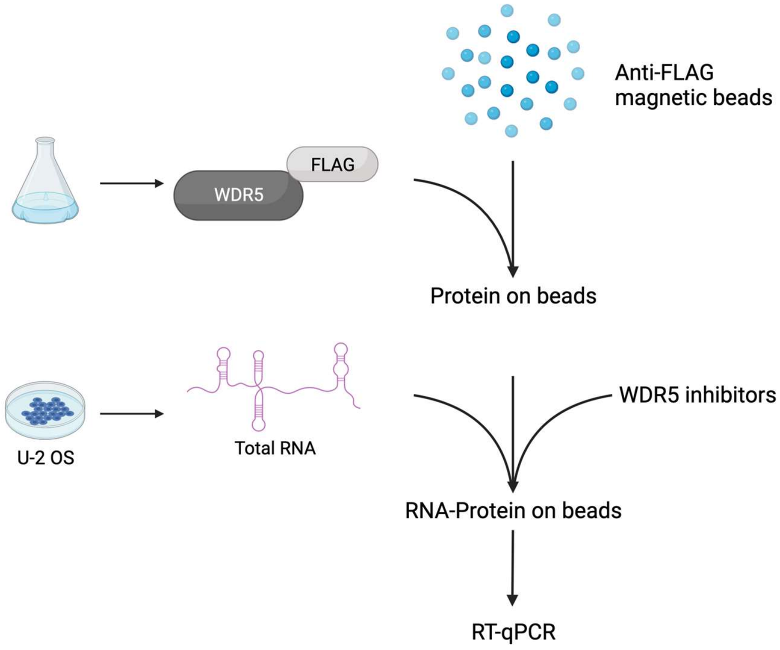


Figure 20. Illustration of the *in vitro* RNA immunoprecipitation. Created with BioRender.com.

Nevertheless, an *in vitro* validation method is necessary to evaluate the inhibition on targeting lncRNA-WDR5 complexes. Considering that the *in vitro* transcribed *HOTTIP* might have problems with misfolding, an assay compatible with total RNA extracted from mammalian cells should be used[52]. Through the literature review, *in vitro* RNA immunoprecipitation (*iv*-RIP) was selected to evaluate the inhibition of the lncRNA-WDR5 complex because of its compatibility with RNA extracts from the cell lysate (Figure 20). The inhibitor was treated with RNA at the same time to monitor the inhibition of lncRNA-WDR5 complex formation. When the inhibitor competes with the lncRNA for binding to WDR5, the amount of lncRNA enriched would reduce. A quantitative analysis of RNA enrichment can be obtained by using reverse transcription quantitative polymerase chain reaction (RT-qPCR) as the detection method, which allows us to selectively analyze the enrichment of various lncRNAs in one experiment, thus improving the efficiency of data collection.

2.2 Optimization of the Linear Sequence

RBBP5 was chosen as the starting point of the project because it had the strongest binding affinity to WBM site from the other publications (Figure 15)[57,59,60,72]. Avdic et al. used ITC to show that residues 371-380 on RBBP5 contribute most to WDR5 binding (Table 1). However, X-ray analysis of the co-crystal structure (PDB ID: 3P4F) suggests that 373-380 of the RBBP5 sequence should contribute most to the interaction with the WBM site (Figure 21). Based on these results, a library of RBBP5 peptides was synthesized and their binding affinity was monitored by FP assay. A reported WDR5 FP buffer formulation consisting of 0.1 M $K_nH_{(3-n)}PO_4$, 25 mM KCl, and 0.01% Triton-X at pH 6.5 was used to perform the FP assay (6.3, FP Buffer A)[87].

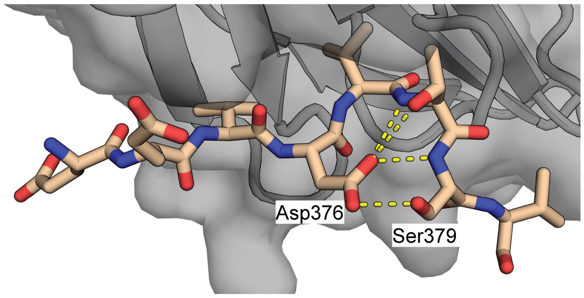


Figure 21. Sequence of RBBP5 bound to WDR5 (PDB ID: 3P4F)[64]. Figure is adapted from (Chang et al., 2023) with permission (6.7).

The full-length sequence (RBBP5 371-381) was synthesized with two versions of FITC conjugation, varying the position of labeling either at the N-terminus or C-terminus to test the influence of the position of fluorophore labeling. The results showed that the position of the label did not affect the binding affinity and both versions had a K_D similar to the reported affinity (Table 2, peptides **P1** and **P2**). To identify the essential binding sequence, two truncated series were synthesized, truncated at either the C-terminus (Table 2, peptide **P1**, **P3**-**P5**) or the N-terminus (Table 2, peptide **P2**, **P6**-**P9**). For the truncation series, the fluorophore was fixed at the terminus opposite to the truncated residue so that the observed binding affinity could better reflect the effect of truncation. A correlation of loss of binding affinity with residue deletion was observed for both the C/N truncation series, indicating that all residues within 371-381 are involved in binding events. When truncation occurred at both ends, an accumulated loss of binding affinity towards WDR5 was observed (Table 2, peptide **P10**).

Table 1. The essential binding element for RBBP5 to recognize WDR5. Figure is adapted from (Avdic et al., 2011) with permission (6.7).

Peptides	**K_D (μM)**
RBBP5 371-410	2.0 ± 0.3
RBBP5 371-400	1.3 ± 0.1
RBBP5 371-390	2.0 ± 0.2
RBBP5 371-380	2.2 ± 0.2
RBBP5 381-390	N.B[a]

[a] : no heat of binding is detected

Table 2. Series of truncated RBBP5 and the corresponding binding affinity for WDR5. K_D was determined by FP in FP buffer A. Residues within the hydrophobic core are shown in red.

Nr.	N- ***Labeling***	Sequence 371	C-	K_D (μM)
P1.	FITC-PEG-	EDEE**VDVT**SVD	-NH$_2$	2.1 ± 0.4
P2.	Ac-	EDEE**VDVT**SVD	-PEG-K(FITC)-NH$_2$	2.4 ± 0.1
	C-truncation			
P3.	FITC-PEG-	EDEE**VDVT**SV	-NH$_2$	7.1 ± 2.8
P4.	FITC-PEG-	EDEE**VDVT**S	-NH$_2$	20.9 ± 8.7
P5.	FITC-PEG-	EDEE**VDVT**	-NH$_2$	29.9 ± 15.1
	N-truncation			
P6.	Ac-	DEE**VDVT**SVD	-PEG-K(FITC)-NH$_2$	6.7 ± 1.0
P7.	Ac-	EE**VDVT**SVD	-PEG-K(FITC)-NH$_2$	23.4 ± 6.9
P8.	Ac-	E**VDVT**SVD	-PEG-K(FITC)-NH$_2$	> 37.5
P9.	Ac-	**VDVT**SVD	-PEG-K(FITC)-NH$_2$	> 37.5
	Other truncation			
P10.	FITC-PEG-	DEE**VDVT**SV	-NH$_2$	10.4 ± 2.7

To optimize the binding affinity, two other native WDR5 binding sequences (c-MYC and KANSL2) were synthesized (Table 3, peptides **P11** and **P12**), and their sequences were also considered as model templates for the following point-mutation series. Among the three native sequences, c-MYC had the strongest binding affinity (Table 3, peptide **P11**), while KANSL2 had the weakest binding affinity (Table 3, peptide **P12**). Since the hydrophobic core was considered to be the essential binding element, a library focused on screening the hydrophobic residues was generated by mimicking the c-MYC or KANSL2 hydrophobic core (Table 3, peptides **P13** - **P19**). Some of the analogs, i.e., V375I, V377I, and T378V, showed improved or similar binding affinity (Table 3, peptides **P13**, **P16**, and **P19**), while some analogs, i.e., V375F, V377L, and V377F, completely lost their binding affinity (Table 3, peptides **P15**, **P17**, and **P18**). The results showed that the binding pocket has a very limited size for specific shape recognition. This is illustrated by the fact that V377I **P16** retains binding affinity, while a methylene shift of one atom V377L **P17** interferes with recognition.

Table 3. Other WBM binding sequences and mutant series of RBBP5 and the corresponding binding affinity for WDR5. K_D was determined by FP in FP buffer A. Residues within the hydrophobic core are colored blue and the mutated or different residues are colored red.

Nr.	Series	N-	Sequence	C-	K_D (μM)
	WBM binding sequence		371		
P1.	RBBP5	FITC-PEG-	EDEE**VDVT**SVD	$-NH_2$	2.1 ± 0.4
			258		
P11.	c-MYC	FITC-PEG-	DEEE**IDVV**SVE	$-NH_2$	0.53 ± 0.06
			407		
P12.	KANSL2	FITC-PEG-	FSDD**LDVV**GDG	$-NH_2$	5.2 ± 0.3
	Point mutation at V375				
P13.	V375I	FITC-PEG-	EDEE**IDVT**SVD	$-NH_2$	1.5 ± 0.1
P14.	V375L	FITC-PEG-	EDEE**LDVT**SVD	$-NH_2$	3.8 ± 0.3
P15.	V375F	FITC-PEG-	EDEE**FDVT**SVD	$-NH_2$	> 37.5
	Point mutation at V377				
P16.	V377I	FITC-PEG-	EDEE**VDIT**SVD	$-NH_2$	2.8 ± 0.2
P17.	V377L	FITC-PEG-	EDEE**VDLT**SVD	$-NH_2$	> 37.5
P18.	V377F	FITC-PEG-	EDEE**VDFT**SVD	$-NH_2$	> 37.5
	Point mutation at T378				
P19.	T378V	FITC-PEG-	EDEE**VDVV**SVD	$-NH_2$	0.82 ± 0.07

2.3 Tailor-Made Macrocycles

In the X-ray analysis of the co-crystal structure between RBBP5 and WDR5 (PDB ID: 3P4F), a constrained conformation was observed for the RBBP5-derived peptide bound to the WBM site (Figure 22). The conformation is stabilized by a hydrogen bonding network between the carboxyl group on Asp376 with the amide backbone of Thr377 and the hydroxyl group on the side chain of Ser379. To further stabilize the conformation of the hydrogen bond between Asp376-Ser379, we devised a strategy of replacing it with a covalent bond where the cyclized product should benefit from reduced entropy loss. As a result, several macrocyclic peptides inspired by X-ray analysis were designed and verified (Figure 22).

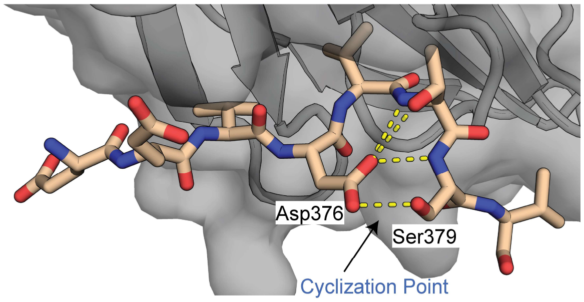
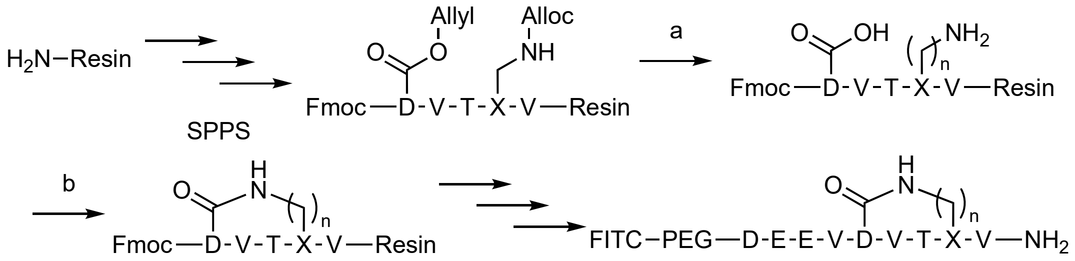


Figure 22. The cyclization point that potentially facilitates peptide to adopt the binding conformation (PDB ID: 3P4F)[64]. Figure modified from (Chang et al., 2023) with permission (6.7).

2.3.1 RBBP5 Macrocyclic Derivatives

The goal was to stabilize the conformation without altering the protein-protein interaction surface. With this in mind, the carbonyl group of Asp376 had to be retained as it is the major hydrogen bond acceptor in the constrained conformation. Finally, amide bond conjugation chemistry was chosen to replace the original hydrogen bond because of its ease of preparation and ability to retain the carbonyl in its original position.

To perform amide bond cyclization while retaining the character of serine, Ser379 was substituted to an amino acid with the primary amine on an unbranched side chain, i.e., L-2,3-diaminopropionic acid (Dap, X^1), L-2,4-diaminobutyric acid (Dab, X^2), and L-ornithine (Orn, X^3). The allyl/alloc protecting groups were used to allow orthogonal deprotection and subsequent on-bead cyclization. The general reaction scheme is shown in Scheme 1. Briefly, the peptides were synthesized up to position 376 and then the allyl/alloc was removed by using a Pd(0) catalyst. Side-to-side cyclization was performed by amide bond formation in the presence of an amide bond coupling reagent, followed by sequence extension and global cleavage to complete the peptide.



Scheme 1. Synthetic strategy for on-bead cyclization. (a) 0.25 eq $Pd(PPh_3)_4$, 25 eq $PhSiH_3$, DCM, rt, 30 min, twice. (b) 2 eq PyAOP, 2eq HOAt, base, DMF, rt, 72 h.

A series of macrocyclic peptides with different ring sizes were generated by substitution of Ser379 with either Dap, Dab or Orn (Table 4, peptides **P20** - **P22**). Peptide **P21**, the 15-membered ring macrocycle, showed an order of magnitude improvement in binding affinity compared to the linear sequence **P10**. The 16-membered ring analog peptide **P22** also showed a strong improvement, approximately four times better than the linear sequence. However, the 14-membered ring analog peptide **P20** showed a loss of binding affinity with a K_D that is about two times weaker compared to the linear sequence. Overall, cyclization improved the binding affinity, with the 15-membered ring performing the best.

Table 4. Cyclic RBBP5 series and corresponding binding affinities determined by FP in FP buffer A. Residues highlighted in bold and red contain the side chains involved in the side-to-side cyclization. The ring size indicates the number of atoms (N) within the cycle. X^1: Dap; X^2: Dab; X^3: Orn. [a]: Purity did not reach 95%.

Nr.	Ring size(N)	N-	Sequence 376	C-	K_D (μM)
Linear sequence					
P10.	linear	FITC-PEG-	DEEV**D** VT**S** V	-NH_2	10.4 ± 2.7
Ring size					
P20.	14	FITC-PEG-	DEEV**D** VT**X**^{1}V	-NH_2	21.9 ± 5.3
P21.	15	FITC-PEG-	DEEV**D** VT**X**^{2}V	-NH_2	1.18 ± 0.07
P22.	16	FITC-PEG-	DEEV**D** VT**X**^{3}V	-NH_2	2.67 ± 0.10
Regioisomer					
P23.[a]	15	FITC-PEG-	DEEV**X**^{2}VT**D** V	-NH_2	> 37.5

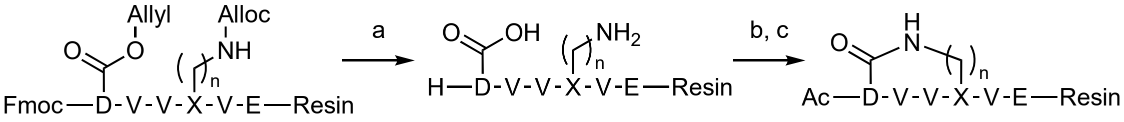
Further investigation of the effect of the carbonyl position was performed by fixing the size of the ring while exchanging the residues on the bridge (Table 4, peptides **P23**), and the results showed that the macrocycles loses binding affinity if the carbonyl group is not in the original position.

2.3.2 c-MYC Macrocyclic Derivatives

With the proof of concept from the study of the RBBP5 macrocyclic series, the same strategy was applied to the c-MYC sequence, the most potent linear peptide in our studies (Table 3, peptide **P11**).

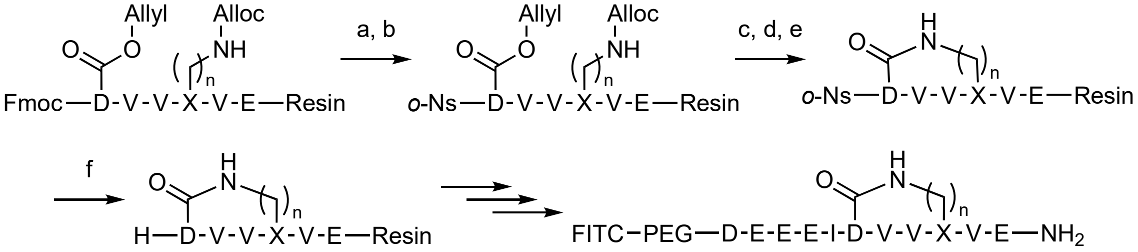
2.3.2.1 Synthesis and Purification of c-MYC Macrocycles

The same strategy from 2.3.1 was used to synthesize c-MYC macrocycles, and the bridging points were Asp263 and Ser266. However, when using the same synthetic route, a significant decrease in isolated yield was observed. The problem was later identified as premature removal of the N-terminal Fmoc group during Alloc deprotection leading to acetylation of the terminus in the capping step (Scheme 2).



Scheme 2. Self-deprotection of Fmoc during the reaction leading to premature termination of the reaction. (a) 0.25 eq $Pd(PPh_3)_4$, 25 eq $PhSiH_3$, DCM, rt, 30 min, twice. (b) 2 eq PyAOP, 2eq HOAt, base, DMF, rt, 72 h. (c) 10 eq Ac_2O, 10 eq DIPEA, DMF, rt, 30 min.

This self-deprotection occurs because Fmoc is sensitive to the primary amine residue when it is deprotected. To prevent this situation, Fmoc was substituted with 2-nitrobenzensulfonyl (*o*-Ns) to protect the N-terminus during the cyclization process. *o*-Ns is stable with the primary amine while it can be deprotected orthogonally with 2-mercaptoethanol, making it an ideal protecting group for this purpose (Scheme 3)[88].



Scheme 3. Synthetic route using *o*-Ns to protect the N-terminus during cyclization. (a) 25% piperidine in DMF, rt, 10 min, twice. (b) 4 eq *o*-NsCl, 5 eq 2,4,6-collidine, 5 eq DMAP, DMF, 15 min, twice. (c) 0.25 eq Pd(PPh_3)$_4$, 25 eq $PhSiH_3$, DCM, rt, 30 min, twice. (d) 2 eq PyAOP, 2eq HOAt, 4 eq 2,4,6-collidine, DMF, rt, 72 h. (e) 10 eq Ac_2O, 10 eq DIPEA, DMF, rt, 30 min. (f) 10 eq 2-mercaptoenthanol, 5 eq DBU, DMF, 16 h.

Due to the multiple negative charges on the sequence, this series of peptides was not easy to purify. The solubility was typically low in most acidic aqueous environments, making it impractical to use the conventional preparative high-performance liquid chromatography (HPLC) purification method with 0.1% TFA in H_2O/ACN as eluent. Instead, a binary alkaline eluent consisting of 15 mM $NH_4HCO_{3(aq)}$ and methanol (MeOH) was used for preparative HPLC purification.

2.3.2.2 Binding Affinity of Derivatized c-MYC Macrocycles

In the c-MYC derivatized macrocyclic series, a similar trend of SAR was obtained. First, macrocycles with 15-membered rings, peptides **P24** and **P26**, showed improved binding affinity compared to linear peptide **P11**, while 16-membered ring, peptides **P25** and **P27**, showed reduced or moderate difference compared to peptide **P11**. Second, deletion in the C-terminus reduced the binding affinity (Table 5, peptides **P24**/**P25** vs. **P26**/**P27**, respectively).

Further optimization was attempted on peptide **P26**, a series of 15/16-membered ring macrocyclic c-MYC undecapeptides were synthesized (Table 5, peptides **P28** - **P32**). When the four N-terminal amino acids were arranged as for RBBP5 style (peptides **P28**/**P29**) the affinity compared to c-MYC derived peptide **P26**/**P27** did not change significantly. Modifications within the hydrophobic core (V264I, V265T or both) also did not improve binding (Table 5, peptides **P30** - **P32**).

Table 5. Cyclic series of c-MYC and the corresponding binding affinity determined by FP in FP buffer A. Residues highlighted in bold and red contain the side chains involved in side-to-side cyclization, and residues highlighted in blue indicate the modification of the sequence. The ring size indicates the number of atoms (N) within the cycle. X^2: Dab; X^3: Orn. [a]: Purity did not reach 95%.

Nr.	Ring size(N)	N-	Sequence	C-	K_D (μM)
Linear sequence			263		
P11.	linear	FITC-PEG-	DEEEI**D** VV**S** VE	-NH_2	0.53 ± 0.06
Ring size & Truncation					
P24.	15	FITC-PEG-	DEEEI**D** VV**X**^{2}V	-NH_2	0.33 ± 0.02
P25.[a]	16	FITC-PEG-	DEEEI**D** VV**X**^{3}V	-NH_2	1.96 ± 0.18
P26.	15	FITC-PEG-	DEEEI**D** VV**X**^{2}VE	-NH_2	0.11 ± 0.02
P27.[a]	16	FITC-PEG-	DEEEI**D** VV**X**^{3}VE	-NH_2	0.56 ± 0.03
Hybrid sequence					
P28.	15	FITC-PEG-	**ED**EEI**D** VV**X**^{2}VE	-NH_2	0.13 ± 0.01
P29.[a]	16	FITC-PEG-	**ED**EEI**D** VV**X**^{3}VE	-NH_2	0.53 ± 0.05
P30.	15	FITC-PEG-	DEEEI**D** V**TX**^{2}VE	-NH_2	0.23 ± 0.01
P31.	15	FITC-PEG-	DEEEI**D** **ITX**^{2}VE	-NH_2	0.43 ± 0.02
P32.	15	FITC-PEG-	DEEEI**D** **I**V**X**^{2}VE	-NH_2	0.18 ± 0.01
Regioisomer					
P33.[a]	15	FITC-PEG-	DEEEI**E** VV**X**^{1}VE	-NH_2	> 37.5
P34.	15	FITC-PEG-	DEEEI**X**^{2}VV**D** VE	-NH_2	> 37.5
P35.[a]	15	FITC-PEG-	DEEEI**X**^{3}VV**E** VE	-NH_2	> 37.5

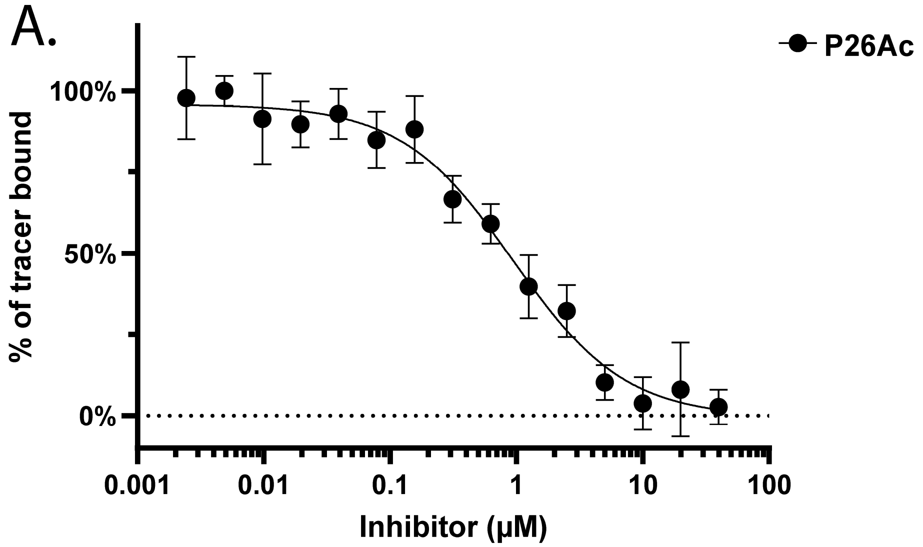
Finally, the importance of the carbonyl within the macrocycles was again tested (Table 5, **P33** - **P35**). When the amide cyclization points were shifted or transposed, no binding was observed in the FP assay. Overall, peptide **P26** was the best candidate among the tested series.

2.4 Target Validations

The direct binding FP experiment provides a primary assay that tells us which peptide had the best performance. However, this assay cannot distinguish a non-specific binding event from a selective binding event. Furthermore, it does not indicate where the binding site is located. As a result, this section is to provide the data to support the claim that peptide **P26** is specific for the WBM site.

2.4.1 Self-Competition FP Experiment

A competition experiment with an unlabeled variant of the tracer peptide **P26** can be performed to prove that the interaction between peptide **P26** and WDR5 is specific, otherwise a competition assay would not give a competition curve. Therefore, an unlabeled version of peptide **P26** (peptide **P26Ac**) was synthesized to carry out the assay. The synthetic route of peptide **P26Ac** was almost identical to that of peptide **P26**, differing only in the N-terminal modification. A competition assay was performed, and a competition curve was obtained with an apparent IC_{50} of 960 nM (Figure 23. A). Given the IC_{50}, protein concentration and probe concentration, the K_I could be calculated using the method developed by Nikolovska-Coleska et al. specifically for the competitive FP assay[89]. The K_I for peptide **P26Ac** was 97 nM (Figure 23. B), which is identical to the K_D of peptide **P26** (110 nM, Figure 23. B), indicating that peptide **P26** selectively recognizes WDR5, and that the fluorophore does not affect the interaction.



B.

Peptide	K_D(nM)	IC_{50} (nM)	K_I (nM)
P26	110 ± 20		
P26Ac		960 ± 141	97.2 ± 28.4

Figure 23. Competitive FP experiment of **P26Ac** versus **P26** for binding to WDR5 determined by FP in FP buffer A. The concentrations of WDR5 and **P26** were fixed at 500 nM and 5 nM, respectively. (A) The FP plot and fitted curve. (B) The corresponding binding affinities for **P26** and **P26Ac**. The FP signal was normalized to theoretical maximum/minimum value. The X-axis was plotted on a logarithmic scale. Figure modified from (Chang et al., 2023) with permission (6.7).

Considering that the future experiments would include small molecule inhibitors, the FP assay was performed in a DMSO-containing buffer instead of the previous DMSO-free buffer (Table 6). Although the binding affinity of peptide **P26** and **P26Ac** decreased slightly compared to the DMSO-free buffer (Table 6 vs. Figure 23. B), the calculated K_I of **P26Ac** was still very close to the K_D of **P26**, indicating that the change in of buffer did not alter the binding mechanism.

$\text{Ac—A-R-T-E-V-H-L-R-K-S-Ahx-Ahx-}\overset{\text{FAM}}{\text{K}}\text{—NH}_2$

P36

$\text{Ac—A-R-T-E-V-H-L-R-K-S—NH}_2$

P36NH₂

Figure 24. Chemical structure of WIN inhibitors[87].

A WIN selective peptide **P36** reported by Karatas et al. was synthesized as a control to monitor binding to the WIN site[87]. In addition, a non-fluorophore labeled version **P36NH₂** was synthesized based on peptide **P36** and a similar self-competitive FP experiment was performed (Figure 24). The WIN-binding peptide **P36** showed a similar K_D as reported (K_D: 0.001 ± 0.0003 μM)[87], and the competition experiment of **P36NH₂** against **P36** yielded a similar K_I, indicating that it also recognizes WDR5.

Table 6. List of binding affinities of the model peptides. FP and competitive FP experiments were performed in FP Buffer B.

Peptide	K_D(nM)	IC_{50} (nM)	K_I (nM)
P26	347 ± 14.3		
P26Ac		1298 ± 92	387 ± 37
P36	0.84 ± 0.14		
P36NH₂		2.13 ± 0.40	0.27 ± 0.14

2.4.2 Isothermal Titration Calorimetry (ITC)

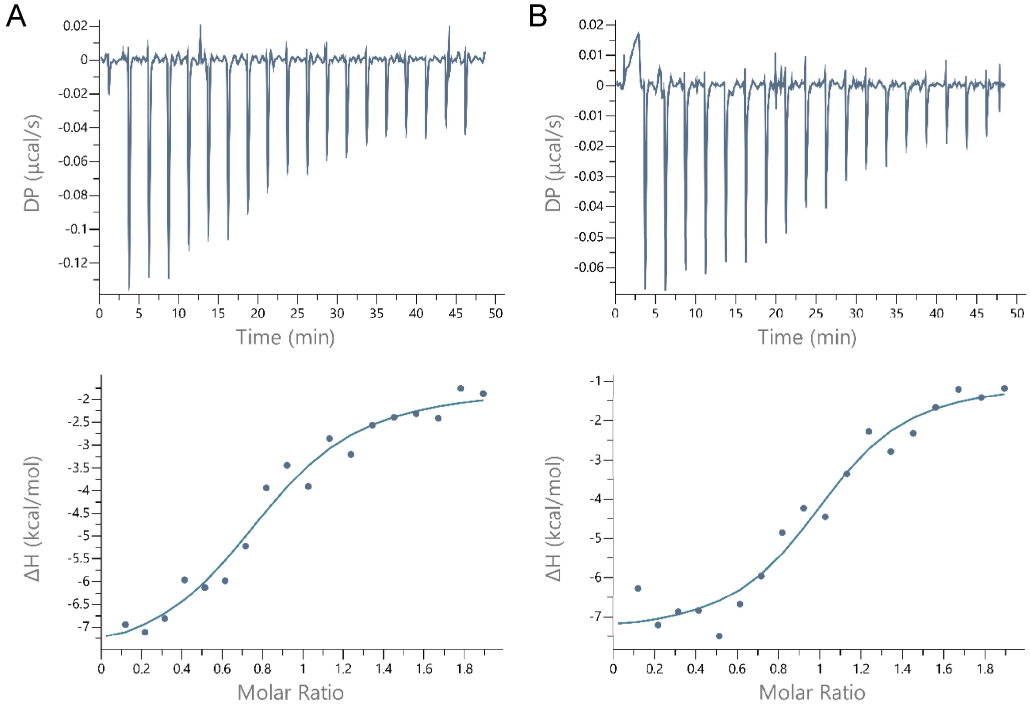
Isothermal titration calorimetry, a non-fluorescence-based assay, was chosen to eliminate the potential bias of a fluorescence-based detection method. Due to the requirement to perform ITC, a Tris-based, detergent-free pH 7.0 assay buffer (See section

6.3, ITC Buffer) was used to avoid precipitation during dialysis while meeting the ITC requirements to avoid the use of detergents in the assay.

2.4.2.1 Sample Concentration Determination

In order to perform the ITC accurately, special efforts were made to estimate the concentration of protein and peptides. Since the sequence was known and the sample was a purified protein, the absorbance at 280 nm was used to determine the protein concentration. For peptides, since the sequence does not contain any UV-active residues, the FITC-labeled version was used so that the extinction coefficient of FITC could be used to determine the concentration of the peptide. In addition, the extinction coefficient of FITC is sensitive to the environment, such as the pH of the aqueous solution. Therefore, the determination of peptides by FITC extinction coefficient was performed in 0.1 M $NaHCO_{3(aq)}$, pH 9.0 to ensure that the readout was made at the maximum extinction coefficient (wavelength (λ): 495 nm, extinction coefficient (ε): 75000 $cm^{-1} \bullet M^{-1}$) of the fluorophore[90].

Both the linear version **P11** and the cyclic version **P26** were included in the ITC experiment and the results are shown in Figure 25. For both, the stoichiometry was close to one, indicating that the binding events occurred at a one-to-one ratio. Although both peptides had a generally weaker K_D in ITC compared to FP, the cyclic peptide **P26** still had five times stronger binding affinity than the linear version **P11**. Overall, the ITC result showed that binding is independent of fluorescence detection.



C

Nr.	N	K_D (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)
P11	0.82 ± 0.01	805 ± 161	-8.32 ± 0.11	-5.67 ± 0.64	-2.66 ± 0.77
P26	0.95 ± 0.07	199 ± 64	-9.16 ± 0.20	-6.94 ± 0.62	-2.23 ± 0.42

Figure 25. The ITC results of peptide **P11** and peptide **P26**. (A) Graph of peptide **P11**. (B) Graph of peptide **P26**. (C) Thermodynamic properties of **P11** and **P26**. The ITC was performed in ITC buffer described in 6.3. Figure modified from (Chang et al., 2023) with permission (6.7).

2.4.3 X-ray Analysis of the Co-crystal Structure

To study the binding mode of the optimized peptide **P26** and to verify the binding site. **P26Ac** was co-crystallized with WDR5 and analyzed by X-ray crystallography (Figure 26. A). The crystal structure showed that **P26Ac** recognized the WBM site in an identical binding conformation as the linear c-MYC peptide, such as the macrocyclic structure, which successfully retained the native binding conformation and hydrogen bonding network (Figure 26. B). Small differences, such as the different position of the side chain of V265 and minor movements of the peptide backbone, were observed in the region of the cyclic core structure. A significant difference was observed in the highly negatively charged N-terminus, but the low electron density around this region suggests that this part is more flexible than other regions. Overall, the crystal structure shows that **P26Ac** binds to the WBM site in an identical manner to native PPI.

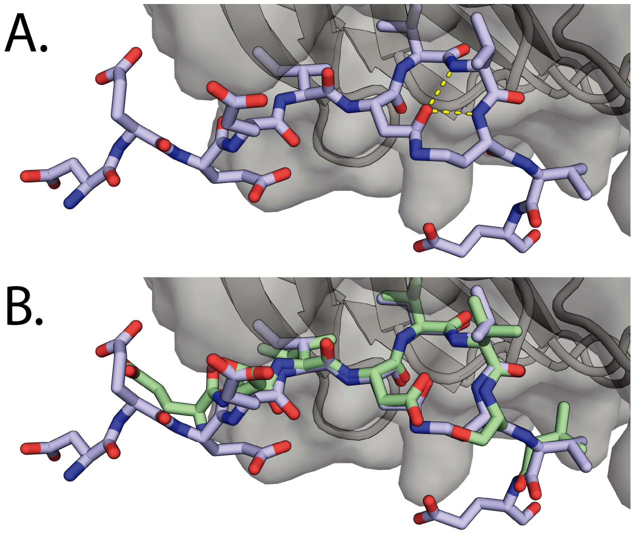


Figure 26. The cocrystal structure of (A) **P26Ac** in the WBM site (PDB ID: 8Q1N). (B) Stacking figure of **P26Ac** and linear c-MYC sequence (PDB ID: 8Q1N and 4Y7R)[64]. Figure modified from (Chang et al., 2023) with permission (6.7).

2.5 *in vitro* RNA Immunoprecipitation (*iv*-RIP)

To test whether a WBM site inhibitor could prevent lncRNA-WDR5 interactions, an *in vitro* RNA immunoprecipitation assay (*iv*-RIP) was used. In addition, several other controls such as WIN peptide and negative controls were included to show that only a WBM-selective peptide inhibitor could disrupt complex formation.

To briefly describe the experiment, FLAG-WDR5 was first loaded onto anti-FLAG magnetic beads, then RNA from cell lysates was loaded onto the WDR5-bead complex in the presence or absence of peptide inhibitors. After washing to remove unbound RNA, the RNA-protein-bead complex was analyzed and the amount of enriched RNA was quantified by RT-qPCR (Figure 27)[91].

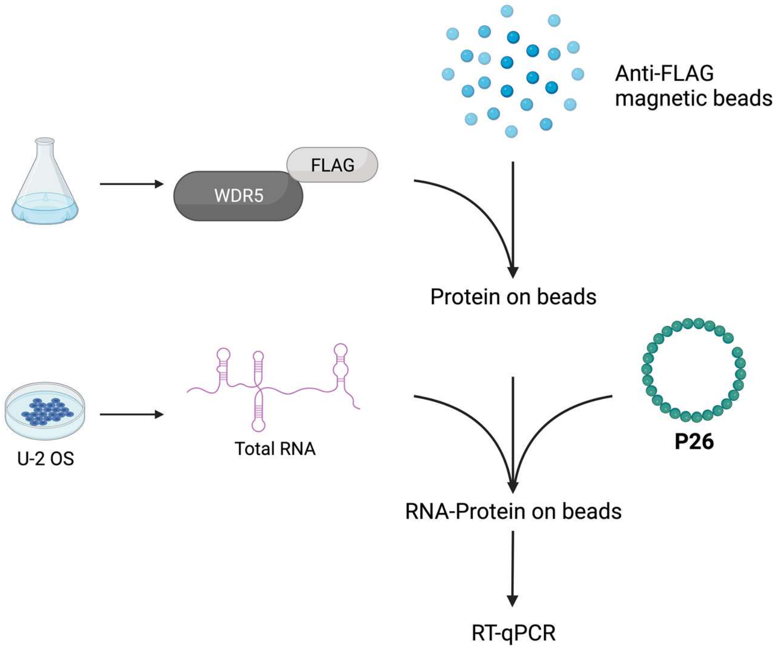


Figure 27. Flowchart of an *iv*-RIP experiment. Created with BioRender.com

2.5.1 RT-qPCR

Quantitative polymerase chain reaction (qPCR) is a widely used method for determining DNA concentrations. An RNA sample requires reverse transcribed to produce the complementary DNA (cDNA), which can then be analyzed by qPCR. When combined, this is called reverse transcription quantitative polymerase chain reaction (RT-qPCR).

To perform the RT reaction, primers must be provided to initiate the reaction, and two types of primers are typically used. For RNAs that have poly-A tails, such as mRNA, a poly-T primer could be used to selectively reverse transcribe and generate a more focused cDNA library. On the other hand, if the target RNA does not contain poly-A tails, a randomized primer could be used to generate a less biased cDNA library.

For the qPCR reaction, the typical SYBR green based method was used, which has a lower cost per sample and can be used to analyze any target by simply substituting the primer pairs to the target DNA.

2.5.1.1 Validating of qPCR Efficiency

A determination of qPCR efficiency and linear dynamic range (LDR) was performed to demonstrate that the results obtained could be used for quantification[92]. To provide a general explanation of how the assay was performed, RNA extracted from cell lysate was reverse transcribed into cDNA, and a serial dilution of cDNA was added to a fixed concentration of qPCR reagents and primers according to the manual provided with the kit. As the amplification efficiency in PCR is sequence dependent, the experiment was performed for all PCR amplicons and the efficiency was calculated individually[92].

Table 7. Result of qPCR efficiency using Promega GoTaq® Master Mix.

Target gene	Primer (µM)	Efficiency	LDR (ng)	R^2
GAPDH	0.3 µM	80%	2 - 0.125	0.997
	0.5 µM	88%	25 - 0.03	1.000
	1.0 µM	85%	25 - 0.025	0.998
ACTB	0.3 µM	65%	2 - 0.125	0.991
	0.5 µM	75%	25 - 0.03	0.999
	1.0 µM	75%	25 - 0.025	0.997
B2M	0.5 µM	86%	25 - 0.03	0.996
	1.0 µM	85%	25 - 0.025	1.000
RPL13A	0.5 µM	86%	25 - 0.03	1.000
	1.0 µM	89%	25 - 0.025	0.999

U-2 OS was reported to have a high *HOTTIP* expression level and was therefore selected for the following tests and to evaluate the accuracy of the workflow for quantitative determination of RNA[69,70]. TRIzol® was used to extract RNA (see section 5.9), followed by reverse transcription (RT) using the High-Capacity cDNA Reverse Transcription Kit (see section 5.10) to obtain the cDNA pool. The Promega GoTaq® Master Mix was initially tested for qPCR efficiency with several housekeeping genes (Table 7), but the efficiency for the selected targets never reached the tolerated range (90% - 110%)[93]. Further attempts to improve efficiency were made by checking the concentration of the primers. However, none of the conditions led to satisfactory improvements (Table 7). The homogeneity of the melting temperature within each replicate indicates that the detection was based on the same target amplification (data not shown). The coefficient of determination (R^2) for each linear regression was very close to one, indicating a properly generated titration curve. In the no template control (NTC), no signal was detected over 40 cycles of replication, indicating that the primer pair did not amplify itself. Overall, GoTaq® did not provide ideal qPCR efficiency with the primers we used.

Although several methods could potentially improve the efficiency, such as redesigning the primers and amplicons or changing the reaction time/temperature during the qPCR reaction, it was decided to try different qPCR master mixes to overcome the situation. The PowerUp™ SYBR™ Green Master Mix provided acceptable qPCR efficiency except for a few selected genes (Table 8, *B2M* and *HOXA-11*). Due to the low copy number of *HOTTIP* in the cells, a detection limit was reached when using cDNA generated from 1 ng of RNA (Table 8. *HOTTIP*). The data was unreliable when working at the concentration close to the detection limit (Ct > 32), as indicated by a lower R^2 observed only for *HOTTIP*.

Table 8. Result of qPCR efficiency of selected target genes using PowerUp™ SYBR™ Green Master Mix.

Target gene	Efficiency	LDR (ng)	R^2	Target gene	Efficiency	LDR (ng)	R^2
ACTB	102%	5 - 0.05	0.998	*HIST1H2A1*	100%	5 - 0.05	1.000
B2M	112%	5 - 0.05	1.000	*PUM*	101%	5 - 0.05	1.000
GAPDH	108%	20 - 0.03	0.998	*TBP*	98%	5 - 0.05	0.999
HMBS	100%	5 - 0.05	1.000	*GUSB*	91%	5 - 0.05	0.990
RPL13A	99%	5 - 0.05	0.998	*HOTTIP*	109%	4 - 1.0	0.989
SDHA	100%	5 - 0.05	1.000	*HOXA-13*	108%	4 - 0.25	0.999
UBC	101%	5 - 0.05	1.000	*HOXA-11*	112%	4 - 0.25	1.000
YWHAZ	98%	5 - 0.05	0.999	*CTNNB1*	108%	4 - 0.25	1.000
U6	99%	5 - 0.05	0.998				

2.5.1.2 Validating of RT-qPCR Efficiency

After checking the efficiency of qPCR, the overall efficiency of RT-qPCR was evaluated. Briefly, a dilution series of RNA was added to the RT kits as suggested by the supplier. After cDNA was generated, it was analyzed by qPCR to determine the linear relationship between RNA input and Ct output.

The supplier suggested that the kit could tolerate 2.0 µg to 0.1 µg of RNA as input and be suitable for RT-qPCR and a titration series using this range was generated. However, the response was not linear when more than 1000 ng of RNA was used in the RT reaction (Figure 28). No reverse transcriptase (NRT) control was included to monitor for genomic DNA contamination from the RNA extraction protocol, and no signal was detected for the selected target genes. The two negative controls, NRT and NTC, together demonstrate that the signal could only be generated when the selected target sequence was reverse transcribed and presented in the qPCR reaction.

Since 2.0 µg was outside the linear dynamic range, this data point was excluded from the following replicate experiment. *ACTB* and *HOTTIP* were used to determine the maximum tolerated amount of RNA in the RT reaction. The linear dose-response for *HOTTIP* fit well between the range of 1000 ng to 0.125 ng. On the other hand, *ACTB* was occasionally inhibited when 1 µg of RNA was used in the RT reaction. When this data point was excluded from the linear regression curve, an R^2 close to one was obtained. Considering the low copy number of *HOTTIP* in the used cell line, it would be better to perform the RT-qPCR analysis at a higher concentration to avoid measurements below the detection limit.

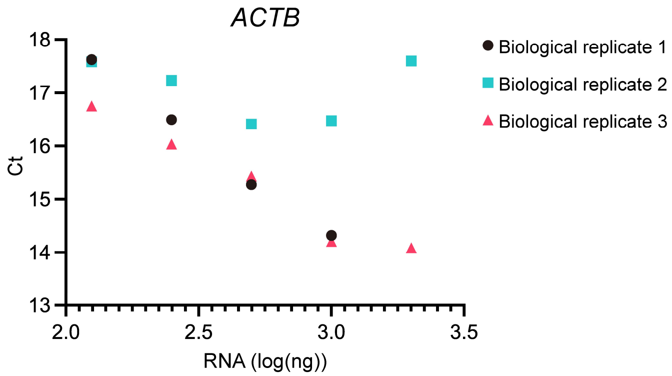
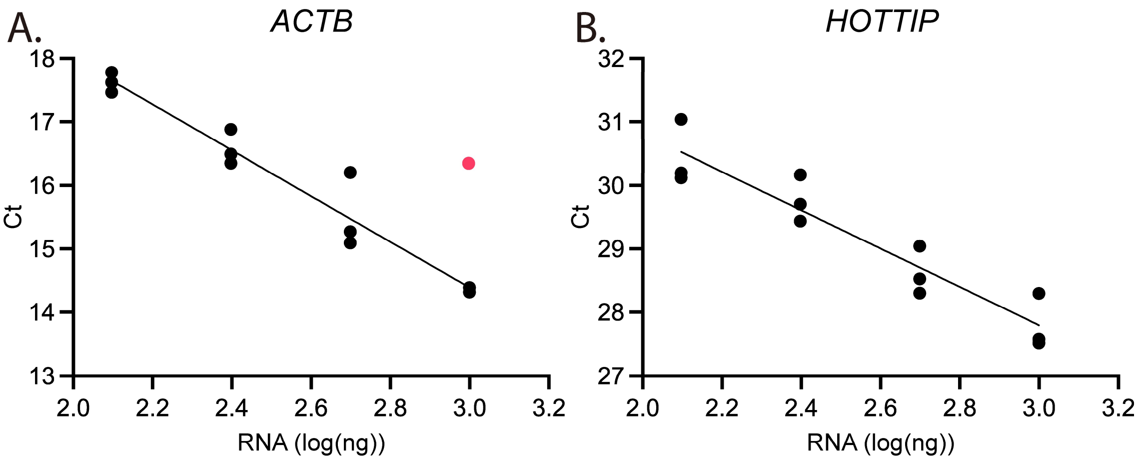


Figure 28. Plot of Ct response of *ACTB* versus amount of RNA used in the RT reaction. The experiment was performed in three biological replicates.



C.

Target gene	Efficiency	RT LDR (ng)	R^2
ACTB	93.5%	500 - 125	1.000
HOTTIP	114%	1000 - 125	0.991

Figure 29. Linear regression of Ct versus amount of RNA used in RT. (A) The plot and linear regression curve fitting for *ACTB*. (N = 3) (B) The plot and linear regression curve fitting for *HOTTIP*. (N = 3). (C) RT-qPCR parameters of *ACTB* and *HOTTIP.* The data point in red was excluded from the linear regression curve fitting.

2.5.1.3 lncRNA Expression Level

To decide which cell lines were suitable for the purpose of the *iv*-RIP experiment, we determined the RNA expression levels of three options (Table 9). These three cell lines, U-2 OS, MDA-MB-231 and MIA PACA-2, were chosen because they were used to study the effect after knockdown of *HOTTIP in previous studies*[68–70,94]. The housekeeping genes were expressed at similar levels with little variation between cell lines (Table 9. *GAPDH*, *ACTB* and *U6*). U-2 OS had the highest expression level of *HOTTIP* and *HOXC13-AS*, but rather low expression level of *FOXD3-AS1*. MDA-MB-231 had the highest expression level of *FOXD3-AS1* and a moderate level of *HOTTIP*, but HOXC13-AS was barely detectable. MIA PACA-2 showed a rather low expression level of both *HOTTIP* and *FOXD3-AS1*, and no *HOXC13-AS* could be detected. Other genes were included to monitor their expression levels and to evaluate the potential for using the cell line in treatment experiments.

The goal was to demonstrate that targeting the WBM site could inhibit lncRNA-WDR5 complex formation *in vitro*, so a cell line with a higher expression level of lncRNAs was preferred. Since U-2 OS had the highest expression level of *HOTTIP* and *HOXC13-AS*, it was the most suitable for *iv*-RIP. On the other hand, the moderate expression level of *HOTTIP* in MDA-MB-231 could lead to a potential detection limit issue, especially since the enrichment levels in the RIP assay were always rather low. Nevertheless, the expression level of *FOXD3-AS1* in MDA-MB-231 was high enough for the following study.

Table 9. The normal Ct value of RNAs in the cell line. 500 ng of RNA was used in the RT reaction and cDNA generated from 10 ng of RNA was used in the qPCR.

		Cell line		
		U-2 OS	MDA-MB-231	MIA PACA-2
Ct (Mean ± SD)	*GAPDH*	14.51 ± 0.02	15.14 ± 0.12	14.71 ± 0.08
	ACTB	14.69 ± 0.08	14.99 ± 0.06	15.17 ± 0.20
	U6	18.50 ± 0.02	18.86 ± 0.15	19.51 ± 0.26
	HOTTIP	26.53 ± 0.04	28.84 ± 0.27	30.90 ± 0.16
	HOXC13-AS	22.92 ± 0.06	32.33 ± 0.59	n.d.
	FOXD3-AS1	33.33 ± 0.41	26.84 ± 0.06	28.93 ± 0.05
	HOXA-13	22.88 ± 0.01	27.96 ± 0.07	27.87 ± 0.13
	E-CAD	27.63 ± 0.28	29.90 ± 0.25	n.d.
	SNAIL-2	25.63 ± 0.02	26.79 ± 0.04	27.55 ± 0.14

2.5.2 Result of the *iv*-RIP

2.5.2.1 Proof of Concept

As a proof of concept, a simple model assay was performed to determine how the workflow shown in Figure 27 would work. Since the peptides were freely soluble in the assay environment, the *iv*-RIP was performed in the RIP buffer without the addition of 0.1% DMSO. Protein loading was determined by denaturing 5% of the beads followed by analysis by SDS-PAGE (Figure 30. A and D). Similar levels of FLAG-WDR5 were enriched in all samples, indicating that the difference in RNA enrichment was not biased by protein loading levels. Enrichment of *HOTTIP* and *HOXC13-AS* was examined using U-2 OS RNA extracts (Figure 30. B and C). When 10 µM WBM inhibitor macrocycle **P26** was applied, the enrichment of *HOTTIP* or *HOXC13-AS* was reduced to less than 50%. In contrast, when 10 µM of the negative control macrocycle **P33** was used (K_D > 37.5 µM), a slightly higher enrichment of *HOTTIP* or *HOXC13-AS* was observed. No RNA could be detected from the

FLAG-GFP control, suggesting that the enrichment of target RNAs requires the involvement of FLAG-WDR5.

When MDA-MB-231 RNA extracts were used, the enrichment level of *HOTTIP* was below the detection limit (data not shown) and only *FOXD3-AS1* enrichment could be detected (Figure 30. E). Similar to the U-2 OS experiment, only macrocycle **P26** was able to reduce the enrichment level of *FOXD3-AS1*, while the negative control **P33** did not affect the enrichment significantly. An RNA-free control (RNA$^-$) was included to demonstrate that the lncRNA was enriched from the RNA extracts and a FLAG-GFP control to demonstrate that FLAG-WDR5 is required for RNA enrichment.

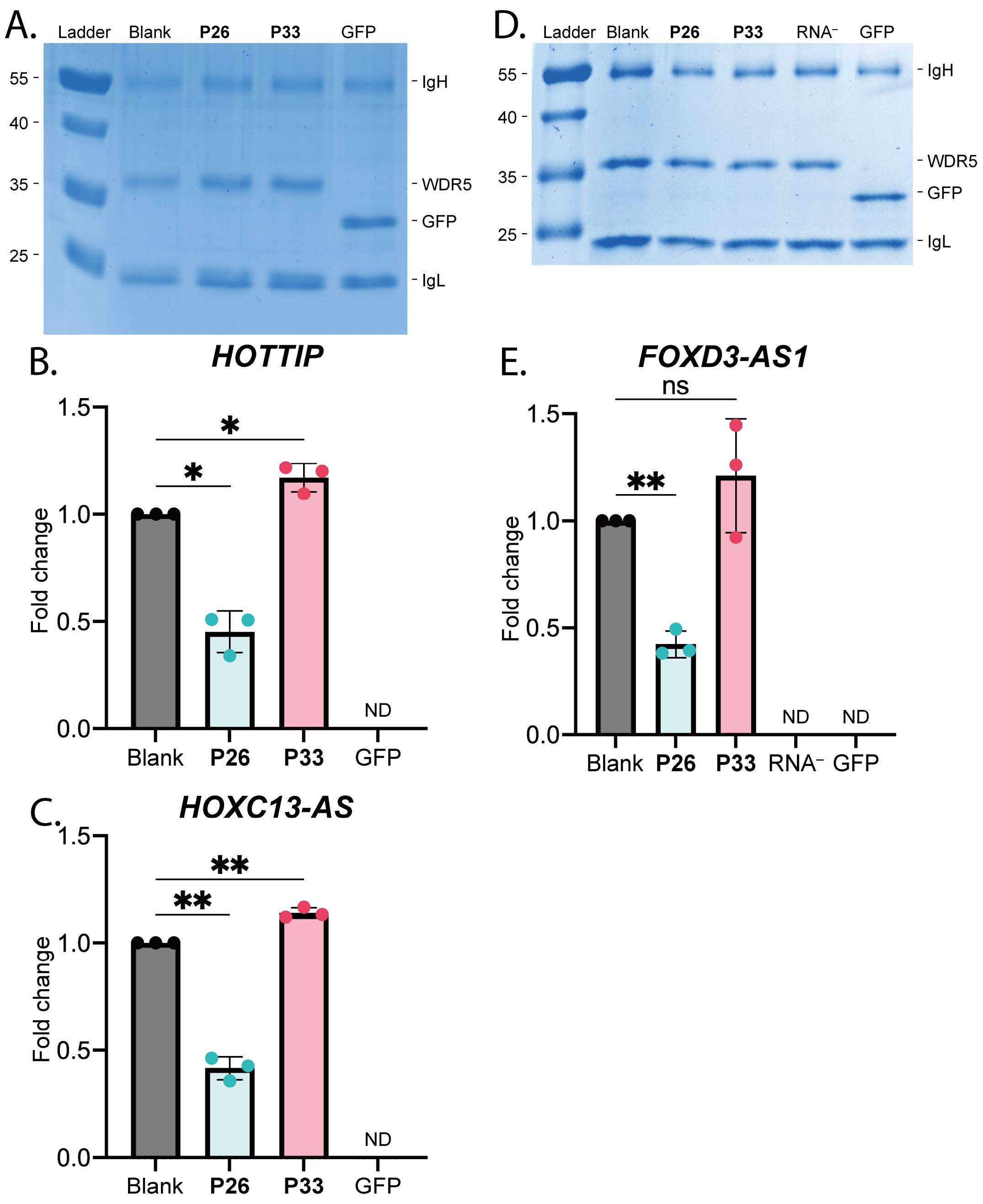


Figure 30. Results of *iv*-RIP and corresponding SDS-PAGE analysis. (A) Protein loading analysis of samples from the U-2 OS experiment. The results were consistent across all three biological replicates. (B) *HOTTIP* enrichment from U-2 OS lysate. (N = 3) (C) *HOXC13-AS* enrichment of U-2 OS lysate. (N = 3) (D) Protein loading analysis of samples from MDA-MB-231 experiment. The results were consistent across all three biological replicates. (E) *FOXD3-AS1* enrichment from MDA-MB-231 lysate. (N = 3); Experiments were performed in RIP buffer as described in 6.3. Protein bands were visualized by staining with Coomassie Blue dye. RNA^{-}: RNA-free control. GFP: FLAG-GFP control. IgH: Immunoglobulin heavy chain. IgL: immunoglobulin light chain. Analysis was done by RT-qPCR and normalized to blank. One-sample student t-test and the target value was set to one. ns: $p > 0.05$, *: $0.05 > p > 0.01$, **: $0.01 > p > 0.001$. ND: not detected.

2.5.2.2 Selectivity and Concentration Dependence

After the proof of concept in the previous section, several questions remained to be answered. For example, the use of a DMSO-free environment would limit this protocol to testing peptide inhibitors. In fact, several small molecule inhibitors targeting the WBM site have been reported over time and their inclusion in this project would be beneficial. Therefore, a DMSO-containing protocol was necessary so that these WDR5 inhibitors can be tested. The result for the small molecule inhibitors will be discussed in Chapter 3.

A titration series (1, 5, 10 μM) of peptide **P26** was generated to determine if the effect was dose dependent (Figure 31). Although the response was not as strong as in the DMSO-free condition (Figure 30. B and C), a similar inhibitory effect was observed for both *HOTTIP* and *HOXC13-AS*. The inhibitory effect of peptide **P26** on *HOXC13-AS* was indeed dose-dependent but for *HOTTIP* this was not as obvious. The non-specific negative control **P33** showed no effect on lncRNA-WDR5 complex formation. The WIN site inhibitor **P36NH$_2$** showed no effect on lncRNA-WDR5 complex formation, suggesting that the WIN site is not involved in lncRNA-WDR5 complex formation. Another negative control, FLAG-WDR5^{F266A}, a mutant that has a defect in the WBM site and lower RNA affinity, was used instead of FLAG-WDR5^{WT}, and very low levels of lncRNAs were detected, indicating that a normal WBM site is required for lncRNA-WDR5 complex formation. Taken together, the data suggested that WBM was essential for lncRNA-WDR5 complex formation while WIN site inhibition had no effect on RPI.

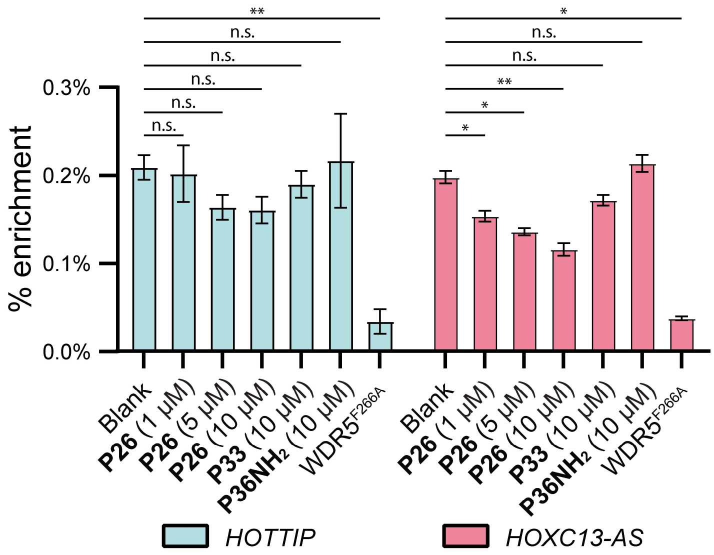


Figure 31. Results of *HOTTIP* and *HOXC13-AS* enrichment in *iv*-RIP experiment from U-2 OS lysate (N = 2)[64]. Figure modified from (Chang et al., 2023) with permission (6.7). ns: p > 0.05, *: 0.05 ≥ p > 0.01, **: 0.01 ≥ p > 0.001. Figure modified from (Chang et al., 2023) with permission (6.7).

2.6 Discussion

In this chapter, a new model for targeting RNA-protein interaction (RPI) via adaptation of an overlapping native protein-protein interaction (PPI) was demonstrated. The synthesized inhibitory peptides were verified by several biophysical assays to demonstrate their binding affinity and specificity, followed by competitive *in vitro* RNA immunoprecipitation (*iv*-RIP) to demonstrate the ability to disrupt lncRNA-WDR5 complex formation.

To optimize the binding affinity of the peptide-based inhibitors, native sequences were first truncated to find the essential binding region and then mutated to identify the important residues. In parallel to the linear sequence optimization, a custom macrocyclic structure was tested. Indeed, side-to-side cyclization preserved the original hydrogen bonding network and significantly improved the binding affinity in comparison to the linear equivalent. The combination of the strongest candidate from the linear optimization and the best cyclization model yielded macrocycle **P26** with a binding affinity (K_D) of 110 nM.

Isothermal titration calorimetry (ITC) was performed to characterize the thermodynamics of the binding. The linear peptide **P11** was included in the ITC experiment together with macrocycle **P26** in order to compare the results side by side. Efforts were made to determine the sample concentration with greater accuracy, and an almost 1:1 binding event was observed (Figure 25). The initial idea was that macrocycle should have an advantage in the ΔS component of the Gibbs free energy equation, while the ΔH remains at the same level, but the results that the macrocycle had an improved ΔH and similar ΔS (Figure 25. C). Compared with linear peptide **P11**, macrocycle **P26** had about four times better binding affinity, similar to the result observed with FP (Table 5).

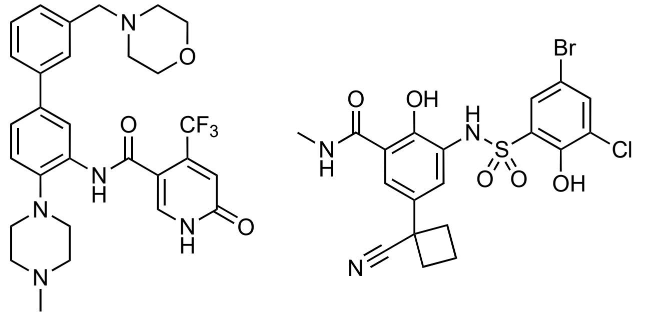
X-ray analysis of the co-crystal structure was performed to identify the position of the binding (Figure 26). In the structure, macrocycle **P26** was located at the WBM site, indicating that the binding was indeed at the target site (Figure 26. A). In addition, overlaying the original structure with the complex of WDR5 and **P26** showed an almost identical orientation at the hydrophobic core, confirming the success in maintaining the binding conformation after cyclization (Figure 26. B). This data highlighted the importance of the hydrogen bonding network within the hydrophobic core in the peptide-WDR5 binding conformation. When the ring size of the macrocycles was not ideal, the binding affinity decreased slightly (Table 5. Peptides **P25** and **P27**). However, when the carbonyl (D376 for RBBP5 and D263 for c-MYC) on the aspartate residue was relocated, the binding affinity decreased by at least two orders of magnitude (Table 5. Peptides **P33** - **P35**). Considering that this carbonyl mutated peptide **P33** had minimal difference compared to macrocycle **P26**, **P33** was later used as a negative control in *iv*-RIP instead of the conventionally used scrambled sequence.

Finally, competitive *iv*-RIP was performed to demonstrate that macrocycle **P26** could inhibit lncRNA-WDR5 interaction *in vitro*. A proof-of-concept experiment was performed

(Figure 30), which demonstrated that only **P26** could inhibit lncRNA-WDR5 complex formation, while the negative control **P33** had little effect on enrichment (Figure 30. B, C, E). Further experiments were performed to test the concentration dependence of the effect. It was observed that in the presence of 0.1% DMSO (Figure 31), the effect was not as strong as in the experiment without DMSO (Figure 30). However, a dose dependence was observed for *HOXC13-AS* (Figure 31). In addition, WIN peptide **P36NH$_2$** did not affect lncRNA-WDR5 complex formation, demonstrating that lncRNA does indeed recognize the WBM site on WDR5 (Figure 31).

Chapter 3. Targeting the lncRNA-WDR5 Interactions *in cellulo*

In Chapter 2, peptide **P26** was shown to inhibit lncRNA-WDR5 complex formation *in vitro*. The goal of this section is to demonstrate the potential therapeutic effect shown in Figure 14, which is to reduce the expression level of *HOTTIP* or *HOXC13-AS* by cell treatment[65]. In addition, the effect of targeting either the WIN or the WBM site by treatment with site-selective inhibitors will be investigated. Some data from this part are being prepared for publication under the title "Small molecule WDR5 inhibitors down-regulate lncRNA expression".



OICR-9429 (**M1**) **7k** (**M2**)

Figure 32. Chemical structures of **OICR-9429** (**M1**) and **7k (M2)**.

Since peptides are generally not considered to be membrane permeable, several attempts were made with the macrocyclic inhibitor **P26** to overcome the problem of cell permeability. Since small molecules are generally considered to be more cell permeable, two small molecules, **OICR-9429** (**M1**) and **7k** (**M2**), were included in the experiment as model inhibitors (Figure 32)[95–97]. **OICR-9429** is a WIN site inhibitor and was used to demonstrate the therapeutic effect of targeting the WIN site *in cellulo*[98]. **7k** is a WBM site inhibitor and has some cell permeability[99].

This part was done in collaboration, and I would like to acknowledge that protein samples were expressed by the Dortmund Protein Facility (DPF), Stefan Schmeing or Cora Neugebauer; fluorescence confocal microscopy was performed with the help of Dr. Stefano Maffini or Stefan Schmeing; flow cytometry analysis was performed by Damian Schiller; **7k** was prepared by Anne Mues gen. Koers when she did her bachelor thesis under my supervision; *iv*-RIP and MDA-MB-231 cell treatment was performed by Cora Neugebauer when she did her master thesis under my supervision.

3.1 General Introduction

Many parameters can influence the effect of a compound at the cellular level, and the presence of proteases in the cytoplasm is an important one for peptide-based inhibitors. Since proteases are abundant in cells, the effect of a treatment can be reduced due to the short half-life of the peptides caused by proteolytic degradation. To explore more stable modulators of lncRNA- WDR5 interactions, the small molecules **OICR-9429** (**M1**) and **7k** (**M2**) were used as compounds to demonstrate the effect of targeting the WIN or WBM site. While **OICR-9429** is a commercially available compound, **7k** is not available and had to be synthesized.

An evaluation of the site-selectivity of the ligands for the WIN/WBM sites was performed using the competitive FP assay, which allowed characterization of the function of the WIN and WBM sites individually. Since the small molecules are considered to have lower specificity compared to peptide inhibitors, **OICR-9429** and **7k** were also included in the cross-competitive FP assay. An *iv*-RIP experiment was performed for **OICR-9429** and **7k** to analyze their inhibitory effect on lncRNA-WDR5 complex formation *in vitro*.

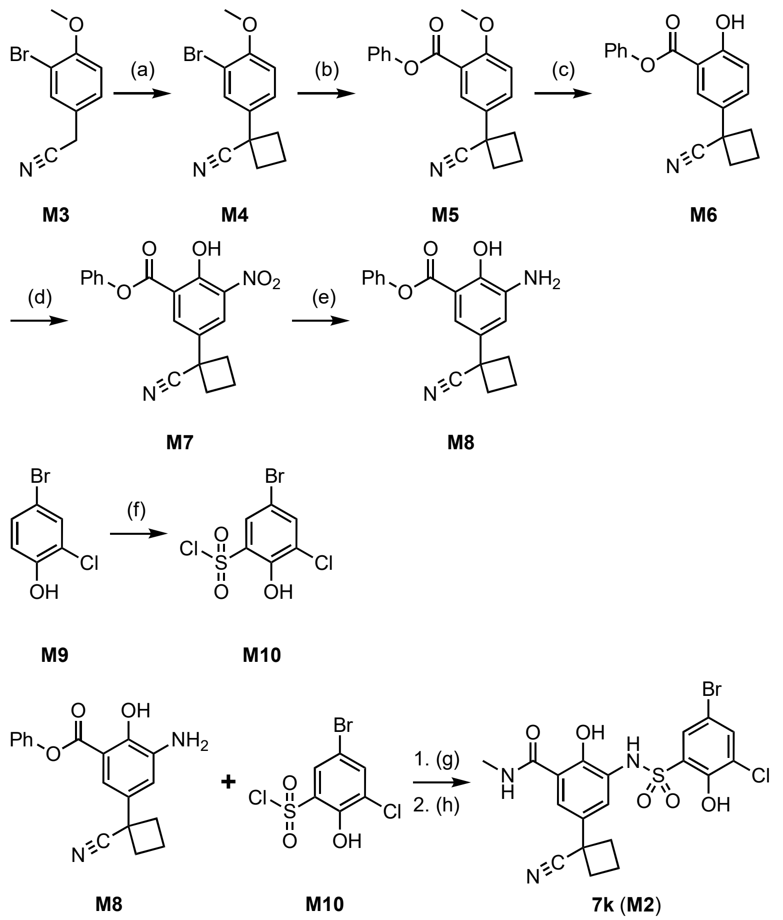
Cellular experiments were performed to verify the effect of treatment, and RT-qPCR was used to analyze the downregulation of the target lncRNAs. An RNA interference experiment targeting *HOTTIP* was performed to demonstrate that the assay could provide the readout, followed by treatment at the cellular level. The cellular uptake of the peptides was analyzed by either FACS or fluorescence confocal microscopy.

3.2 Preparing Small Molecule 7k

Molecule **7k** was synthesized according to the protocol reported by Macdonald et al. (Scheme 4)[96]. The α-proton of the nitrile group on **M3** was deprotonated with NaH and reacted with 1,3-dibromopropane to afford the spirocyclic structure **M4**. The bromo group on **M4** was substituted with a phenyl ester via palladium catalyzed carbonylation to afford **M5**[100]. The methyl ether on **M5** was deprotected with BBr_3, followed by heterogeneous phase nitration with a TBAB as phase transfer catalyst to afford **M7**. Next, hydrogenation with H_2 and Pd/C as catalyst afforded amine **M8**. To synthesize compound **M10**, starting material **M9** was reacted with freshly distilled chlorosulfuric acid in a condensation reaction.

Once the two fragments **M8** and **M10** were ready, a regioselective condensation reaction was carried out in the presence of pyridine to synthesize sulfonamide as the main product. Finally, a transamidation reaction was performed by replacing the phenyl ester with methylamine to obtain the desired product **7k** (**M2**).

During the preparation, several small details were observed and should be noted. First, for the palladium-catalyzed carbonylation reaction to proceed efficiently a properly functioning high-pressure reaction vessel was required. This reaction requires the accumulation of in situ generated CO to participate in the catalytic cycle, and a leaky or oversized reaction vessel would result in a lower partial pressure of CO, thus limiting the reaction to proceed[100]. Second, although the phase transfer catalyst TBAB was used for the heterogeneous phase nitration reaction, very vigorous stirring was required to increase the contact surface area of the two phases. In conclusion, the entire synthetic route worked well as described when these details are taken into account.



Scheme 4. Synthetic scheme of **7k**.[96] (a) NaH, 1,3-dibromopropane, DMSO, rt, 24 h; (b) Phenyl formate, phenol, $Pd(OAc)_2$, $P(t\text{-}Bu)_3 \cdot HBF_4$, NEt_3, ACN 90 ^{oC}, sealed tube, 24 h; (c) BBr_3, DCM, -78 ^{oC}. (d) HNO_3, TBAB, DCE/H_2O, 60 ^{oC}, 24 h; (e) H_2, Pd/C, MeOH, rt; (f) HSO_3Cl, 0 ^{oC} - rt; (g) Pyridine, DCM, rt; (h) NH_2Me, DIPEA, THF, 65 ^{oC}.

3.3 *in vitro* Validations

The selectivity of the peptides was tested in Chapter 2.4 using a DMSO-free buffer environment. In this section, a buffer containing 0.1% DMSO was used because small molecules were to be tested as well for comparison. In addition, all compounds were tested for both WIN and WBM sites to evaluate the selectivity of the inhibitors towards their target pocket. Finally, competitive *iv*-RIP experiments were performed to validate the effect of small molecules on lncRNA-WDR5 complex formation.

In addition to wild-type WDR5 ($WDR5^{WT}$), F266A-WDR5 ($WDR5^{F266A}$) is a mutant reported to have significantly lowered binding affinity toward lncRNA but retain the ability to form protein-WDR5 complexes[65]. In this section, the binding affinity between **P26**/**P36** toward WDR5^{F266A} is verified by FP experiments. Since F266 is located at the WBM site, the single point mutation of F266A is expected to affect WBM recognition. The quantified results would provide a better understanding of the binding events for future studies[65].

3.3.1 Evaluation of Target Specificity

The FP buffer B used in Table 6 is used for this section. To evaluate the effect of the F266A mutation on peptide WDR5 recognition, the binding affinity of compounds **P26** and **P36** toward WDR5^{F266A} was evaluated (Figure 33. A). Indeed, the binding affinity of **P26** toward WDR5^{F266A} decreased by 32 times compared to WDR5^{WT}. In addition, the binding affinity of **P36** toward WDR5^{F266A} decreased by 92-fold compared to WDR5^{WT}, indicating that the WIN binding site is affected by the F266A mutation.

A

K_D (nM)		Protein	
		WDR5^{WT}	WDR5^{F266A}
Tracer	**P26**	347 ± 14.3	11200 ± 1900
	P36	0.84 ± 0.14	78.0 ± 13.8

B

K_I (nM)		Inhibitor			
		P36NH$_2$	**OICR-9429**	**P26Ac**	**7k**
Tracer	**P26**	> 40000	> 40000	387 ± 37.7	107 ± 20.0
	P36	0.27 ± 0.14	1.24 ± 0.39	> 40000	> 40000

Figure 33. Results of the FP assay from the cross-titration experiments. A: Table of K_D from the direct binding assay. B: Table of K_I from the competitive binding assay. K_D/IC$_{50}$ were curve-fitted using Graphpad Prism 9. K_I were calculated from IC$_{50}$ by the method of Nikolovska-Coleska et al.[89]

To evaluate the specificity of the inhibitors, WIN site ligands were tested for binding to the WBM site and vice versa (Figure 33. B). When the WBM inhibitors **P26Ac** or **7k** were titrated against the WBM tracer **P26**, the observed K_I was close to the reported values[96]. The observed K_I for WIN inhibitors **P36NH$_2$** and **OICR-9429** against WIN tracer **P36** were slightly lower than the reported values[87,95], but the trend of affinity remains the same.

Additionally, no competition was observed when WBM inhibitors were titrated against the WIN tracer, and no competition was observed when WIN inhibitors were titrated against WBM tracer. Overall, all inhibitors were highly selective for their binding site and no cross-inhibition was observed.

3.3.2 Results of *iv*-RIP

The small molecule inhibitors **OICR-9429** and **7k** were tested for their ability to inhibit lncRNA-protein complex formation by *iv*-RIP (Figure 34). The direct WBM inhibitor **7k** resulted in a significantly reduced enrichment of *HOTTIP* and *HOXC13-AS*, indicating that these two lncRNAs recognize WDR5 at the WBM site. A small effect was observed when WIN inhibitor **OICR-9429** was used, but the reduction in enrichment was not significant by Welch's t-test. Since similar protein loading levels of WDR5^{WT} were observed for all samples, the amount of RNA enriched from *iv*-RIP did indeed reflect lncRNA-WDR5 complex formation *in vitro* (Figure 34. C). Using the negative control WDR5^{F266A} instead of WDR5^{WT} for *iv*-RIP resulted in a significant reduction in lncRNA enrichment, despite a higher level of WDR5^{F266A} loading on the beads (Figure 34. C). The no RNA control indicated that the enrichment of target lncRNA required the input of RNA extracts, and the GFP control indicated that the enrichment of target lncRNA also required WDR5^{WT}.

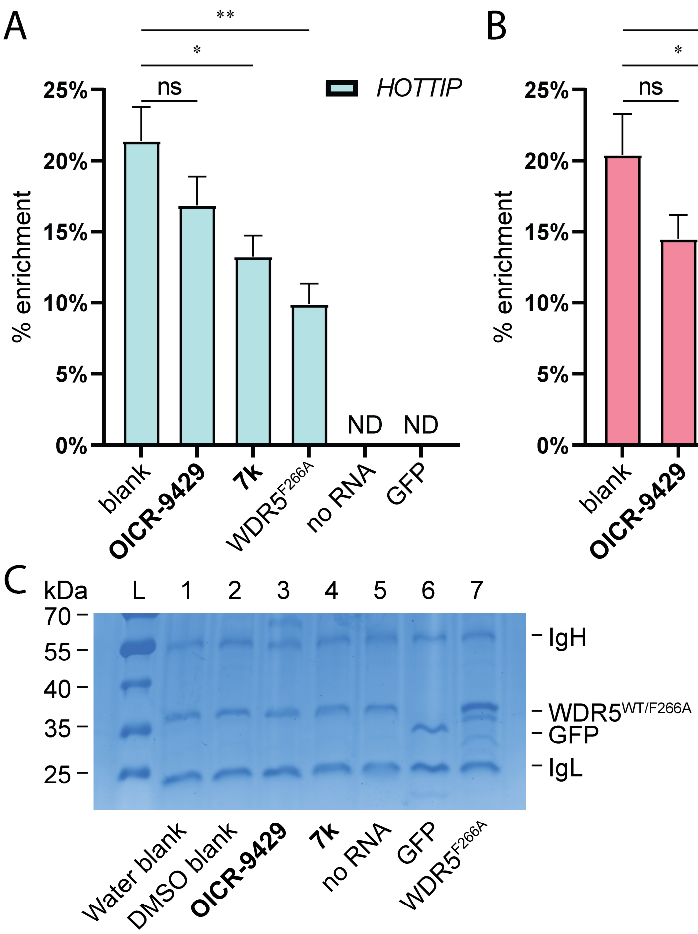


Figure 34. *iv*-RIP results using RNA isolates from U-2 OS cells and WDR5^{WT}, WDR^{F266A}, or GFP. Compounds **OICR-9429** and **7k** were tested at 10 μM in combination with WDR5^{WT}. No RNA control was tested with WDR5^{WT}. Significance of difference between samples was analyzed by Welch's t-test. ND: not detected. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

3.4 Target Validation for Macrocycles *in cellulo*

After having obtained promising *in vitro* data, the next step was to demonstrate the potential therapeutic effect at the cellular level. Most importantly, we wanted to show that lncRNAs such as *HOTTIP* can be downregulated by targeting WDR5.

3.4.1 Experimental Design and Model Testing

To study *HOTTIP* downregulation, a cell line with a higher *HOTTIP* expression level is preferred to make sure the detection limit of qPCR is reached. Therefore, U-2 OS was selected because it had the highest *HOTTIP* expression level among the three cell lines tested in Chapter 2.5.1.3 (Table 9). In addition, several publications demonstrated the effect of *HOTTIP* knockdown in U-2 OS, resulting in several different phenotypic changes[69,70].

To verify whether our RT-qPCR could detect down-regulation of *HOTTIP*, we knocked *HOTTIP* down with siRNA as a positive control (Figure 35). Indeed, transfection of siHOTTIP with Lipofectamine 2000 successfully reduced the expression level of *HOTTIP* in U-2 OS cell after incubation for one day. The difference was significant when compared to the non-targeting siRNA control or blank. The negative control non-targeting siRNA showed a significant *HOTTIP* reduction compared to blank, a sample with no exposure to lipofectamine or siRNA, indicating that some optimization was required to prevent non-specific effects on *HOTTIP* expression levels. Overall, the experiment demonstrated the system's ability to detect changes in *HOTTIP* expression levels.

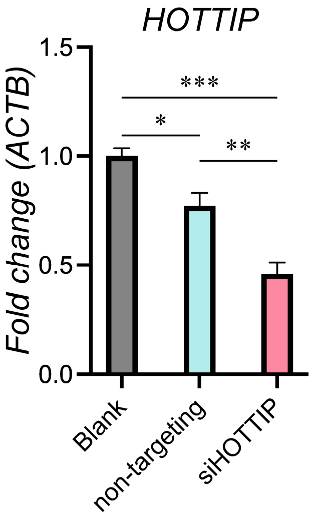


Figure 35. Knockdown of *HOTTIP* with siRNA reduces the expression level of *HOTTIP* after one day of incubation. Lincode Non-Targeting pool (Dharmacon) was used for as a non-targeting control. The significance of the difference between samples was analyzed by Welch's t-test. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$. ***: $0.001 \geq p > 0.0001$.

3.4.2 Attempts on use Macrocycles *in cellulo*

To overcome the low membrane permeability of the peptides, a few methods to cross the membrane were evaluated. These were 1. Direct conjugation with a sequence of trans-activator of transcription (Tat) with several different conjugation strategies[101]. 2. Lipofectamine 2000 transfection reagent. 3. Mixing the peptides with the phase-separating peptide, **HB*pep*-SP** (**P38**), to form cell permeable microparticles[102]. 4. Direct treatment of cells that undergo macropinocytosis[103]. 5. Transfection by electroporation.

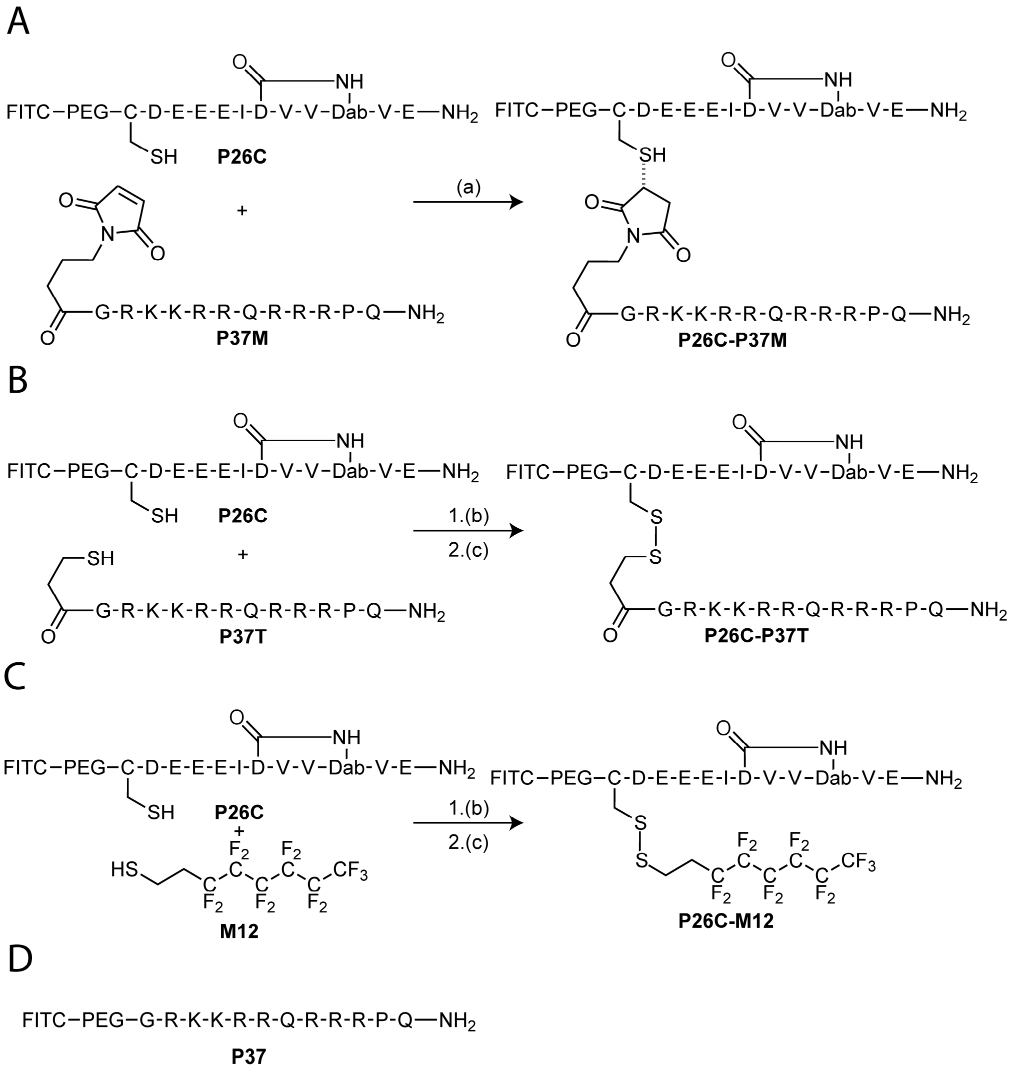
3.4.2.1 Cell-Penetrating Peptides (CPP)

Considering that peptide **P26** is highly negatively charged, its ability to directly diffuse though the membrane into the cell was low[104]. To improve on this limitation of inhibitor **P26**, a cell-penetrating peptide (CPP) moiety could be added to the structure to improve permeability and thus overcome the membrane barrier[105].

The most prominent CPP, Tat, was selected as the CPP moiety[101,105]. However, the highly positive charge of Tat could be neutralized by the highly negatively charged peptide **P26** which has previously been shown to reduce its cell-permeability enhancement effects [106,107]. Therefore, a neutrally charged membrane-permeable group, perfluoro-1-octanethiol, was also tested[108].

3.4.2.1.1 Preparation of P26-CPP Conjugated Product

Despite the well-established SPPS protocols, difficulties arose when attempting direct extension of peptide sequences to combine **P26** and Tat. Both possible combinations were tested (Tat at C- or N-terminus), but none of the syntheses yielded the desired product. Therefore, a conjugation strategy was adopted to overcome the difficulties encountered. The Maleimide-Thiol "click" chemistry was chosen to conjugate Tat with peptide **P26** (Scheme 5. A)[109].



Scheme 5. Synthetic scheme of **P26** conjugated with membrane permeable group. A. The reaction between **P26C** and maleimido-Tat(**P37M**). B. The reaction between **P26C** and 3-mercaptopropionyl Tat (**P37T**). C. The reaction between **P26C** and perfluoro-1-octanethiol (**M12**). D. The structure of Tat (**P37**). (a). 0.1M NH_4HCO_3(aq.), MeOH, pH 9; (b). 2,2'-dithiodipyridine, MeOH; (c). 0.1M NH_4HCO_3(aq.), MeOH, pH 9.

To perform the maleimide-thiol reaction, a thiol group was installed at the N-terminus of peptide **P26** by incorporating a cysteine (**P26C**) into the sequence, while a maleimide group was installed at the N-terminus of Tat (**P37**) by coupling the 4-maleimidobutanoic acid (**P37M**) at the N-terminus. The maleimide-thiol "click" conjugation was performed by mixing the starting materials **P26C** and **P37M** in 0.1M ammonium bicarbonate buffer with some methanol to give the product **P26C-P37M** (Scheme 5. A). A reversible linker strategy was employed by disulfide bridge conjugation. Briefly, the thiol group at the N-terminus of

P37T was first reacted with 2,2'-dithiodipyridine to give an activated disulfide, followed by a thiol-disulfide exchange in an alkaline environment to give the product **P26C-P37T** (Scheme 5. B). The same disulfide strategy was applied to perfluoro-1-octanethiol (**M12**) to afford the product **P26C-M12** (Scheme 5. C).

3.4.2.1.2 Cellular Uptake Analysis by Flow Cytometry

The improvement of the CPP conjugates was first evaluated by flow cytometry. Since many CCPs are membrane-lytic[110], propidium iodide (PI) staining was included in the protocol to evaluate the proportion of apoptotic cells. A trypan blue quenching procedure was performed to reduce the signal resulting from the non-specific charge-charge interaction between Tat and the outer cell membrane[111].

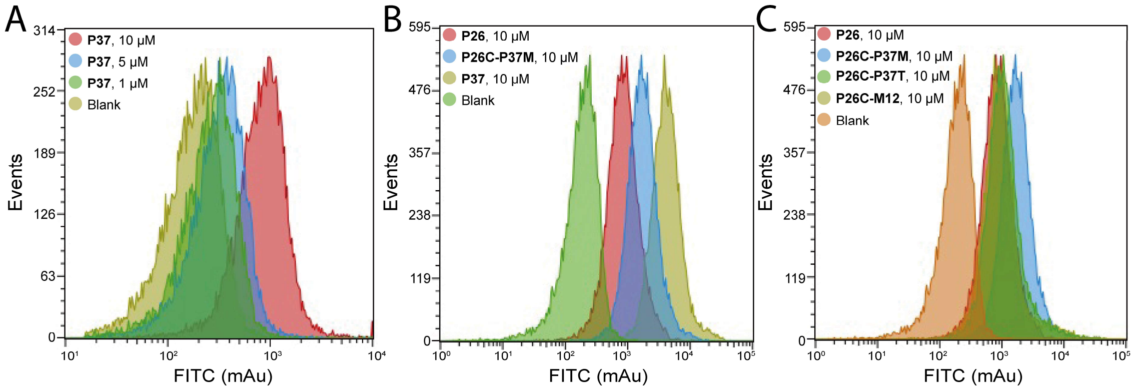


Figure 36. Histograms of flow cytometry analysis for four-hour treatments on U-2 OS. A. Dilution series of **P37**. B. Model assay of **P26C-P37M** with **P26**, **P37** and blank. C. Combined result of all conjugate series with controls. The x-axis is the intensity of the FITC channel on a logarithmic scale and the y-axis is the number of events.

To determine the dynamic range of the flow cytometry analysis, a titration series of positive control **P37** was generated. Briefly, a three-point dilution series (1/5/10 μM) of **P37** was used to treat U-2 OS cells for four hours at 37 ºC to examine the signal intensity during the flow cytometry analysis. The results demonstrate that the system could verify the fluorescence intensity at the current treatment concentration range (Figure 36. A). In addition, a significant shift in the FITC peak was distinguishable at the highest tested concentration which was then used for further experiment (Figure 36. A). Using these experimental settings, the three compounds **P26**, **P26C**-**P37M**, and **P37** were tested to verify the effect of CPP-conjugation (Figure 36. B). Interestingly, the unmodified peptide **P26** was significantly higher than the background, indicating that some of **P26** was taken up by the cells (Figure 36. B, red). Treatment with **P26C-P37M**, the peptide irreversibly conjugated with CPP, showed a significantly higher signal compared to unmodified **P26**, demonstrating that conjugation with a CPP does indeed promote cellular uptake (Figure 36. B, blue). However, the positive control **P37** showed an even stronger signal compared to **P26C-P37M**, suggesting that conjugation of the cargo affected the uptake efficiency of **P37** (Figure 36. B, yellow). Finally, we tested the other two conjugated peptides, **P26C-P37T** or **P26C-M12**, and observed that they were taken up at levels similar to unmodified **P26**, concluding that the disulfide conjugation strategy did not improve cellular uptake in this case (Figure 36. C, green and yellow).

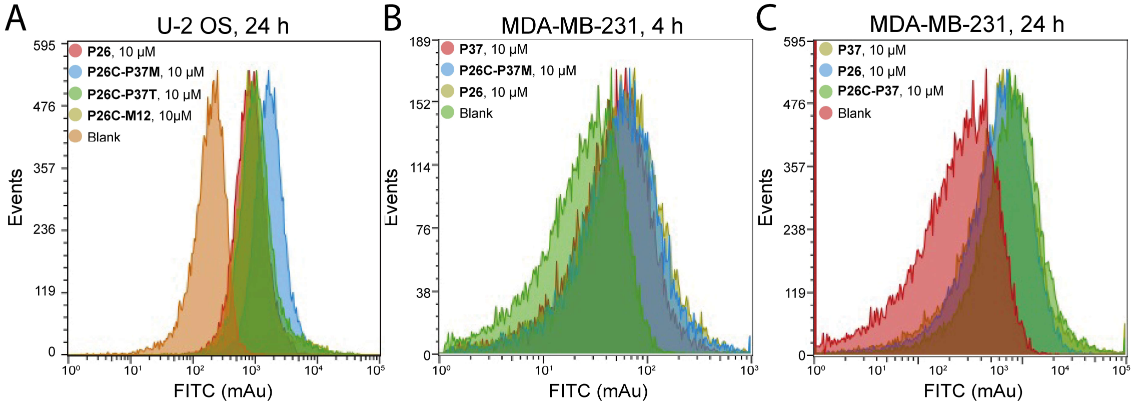


Figure 37. Histograms of flow cytometry analysis for different conditions. A. Treatment with U-2 OS at 24 h. B. Treatment with MDA-MB-231 at 4 h. C. Treatment with MDA-MB-231 at 24 h. The x-axis is the intensity of the FITC channel on a logarithmic scale and the y-axis is the number of events.

In addition to four hours incubation, a one-day incubation on U-2 OS was performed to monitor the time dependence of cellular uptake (Figure 37. A), and the result was similar to the four-hour incubation (Figure 36. C), suggesting that uptake rate was not time dependent.

The effect of cellular uptake was evaluated in a second cell line MDA-MB-231. Interestingly, when the measurements were performed using MDA-MB-231, all peptides showed a similar levels of cellular uptake (Figure 37. B and C) and the signal was stronger than the blank (Figure 37. B, green; C, red), indicating that the cell took up the peptide without preference. This non-selective uptake phenomenon was observed in both four-hour and one-day incubations (Figure 37. B and C).

3.4.2.1.3 Evaluation of Cellular Uptake by Fluorescence Confocal Microscopy

Fluorescence confocal microscopy was performed to determine the localization of the peptides after cellular uptake (Figure 38. A). The unmodified **P26** showed a minimal signal with sporadic spots, while the conjugated **P26C-P37M** showed a stronger overall and cytosolic intensity, indicating that the conjugation with CPP improved the cellular uptake and the peptide was at least partially escaped from the endosome. The images were analyzed by corrected total cell fluorescence (CTCF) to provide a more quantitative result (Figure 38. B). Again, the averaged CTCF observed with **P26C-P37M** was significantly higher than that of unmodified **P26**, indicating that the improvement in cellular uptake was due to conjugation with CPP. Nevertheless, similar to the results in flow cytometry analysis, peptide **P37** itself showed a better uptake in fluorescence confocal microscopy, indicating that conjugation of **P26C** to **P37** compromises its cell permeability.

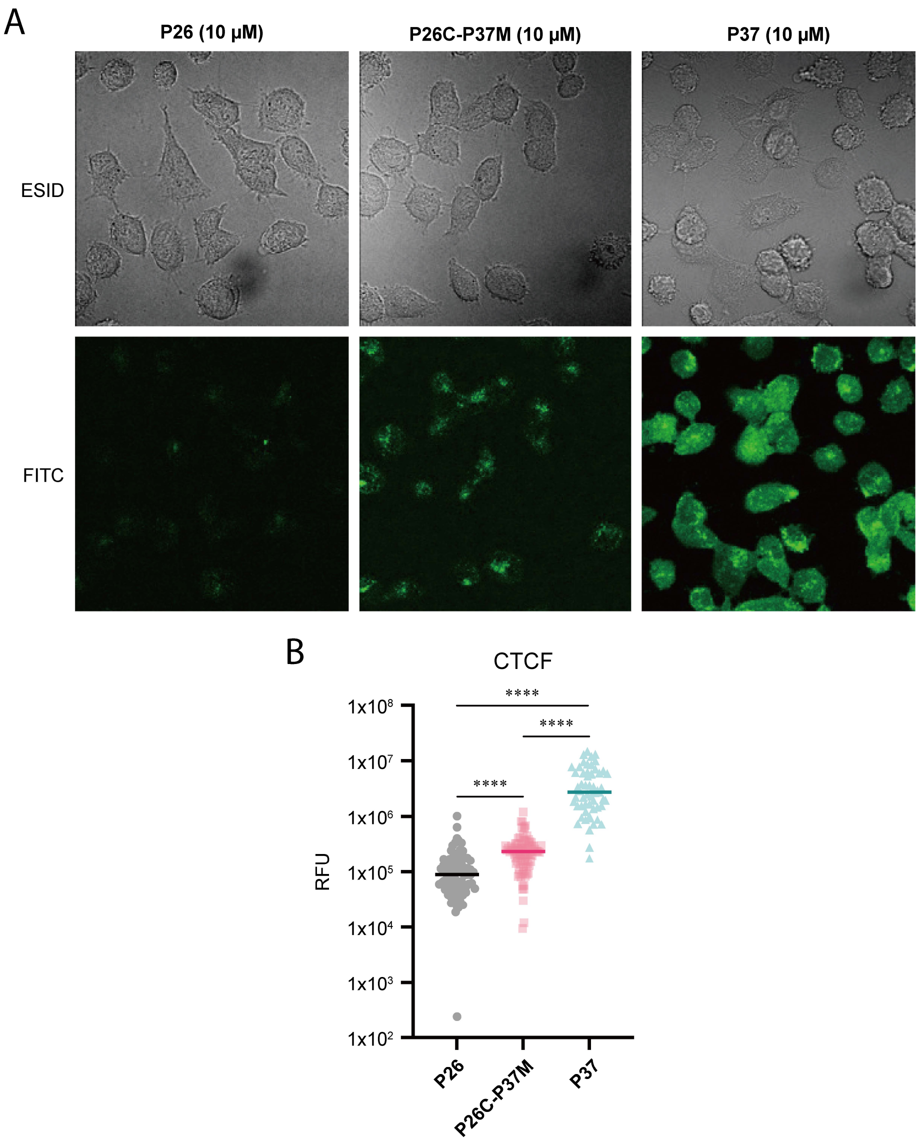


Figure 38. Results of fluorescence confocal microscopy. A. Pictures of HEK293T taken by Electronically switchable illumination and detection module (ESID) or GFP channel. The pictures were taken using a ZEISS LSM 800. B. CTCF values calculated from the images. The y-axis is the relative fluorescence unit (RFU) on a logarithmic scale. The significance of the difference between samples was analyzed by Welch's t-test. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$, ***: $0.001 \geq p > 0.0001$, ****: $0.0001 \geq p$

3.4.2.1.4 Effect of Treatment with CPP Conjugates

RT-qPCR was used to monitor the effect of inhibitor treatment on the cells. Although the peptide **P26C-P37M** was taken up by the cells and not trapped in the endosome, it did not reduce the expression level of *HOTTIP* or other monitored lncRNAs. The following section focuses on the conditions tested.

In the first experiment, a one-day treatment with 10 μM on U-2 OS was tested, and nothing changed among the selected lncRNAs (Figure 39. A). Considering that the effect of epigenetic mechanisms sometimes requires a longer incubation time, it was extended to a three-day treatment, or two continuous three-day treatments with the medium refreshed on the third day, still with no significant effect on lncRNA expression (Figure 39. B).

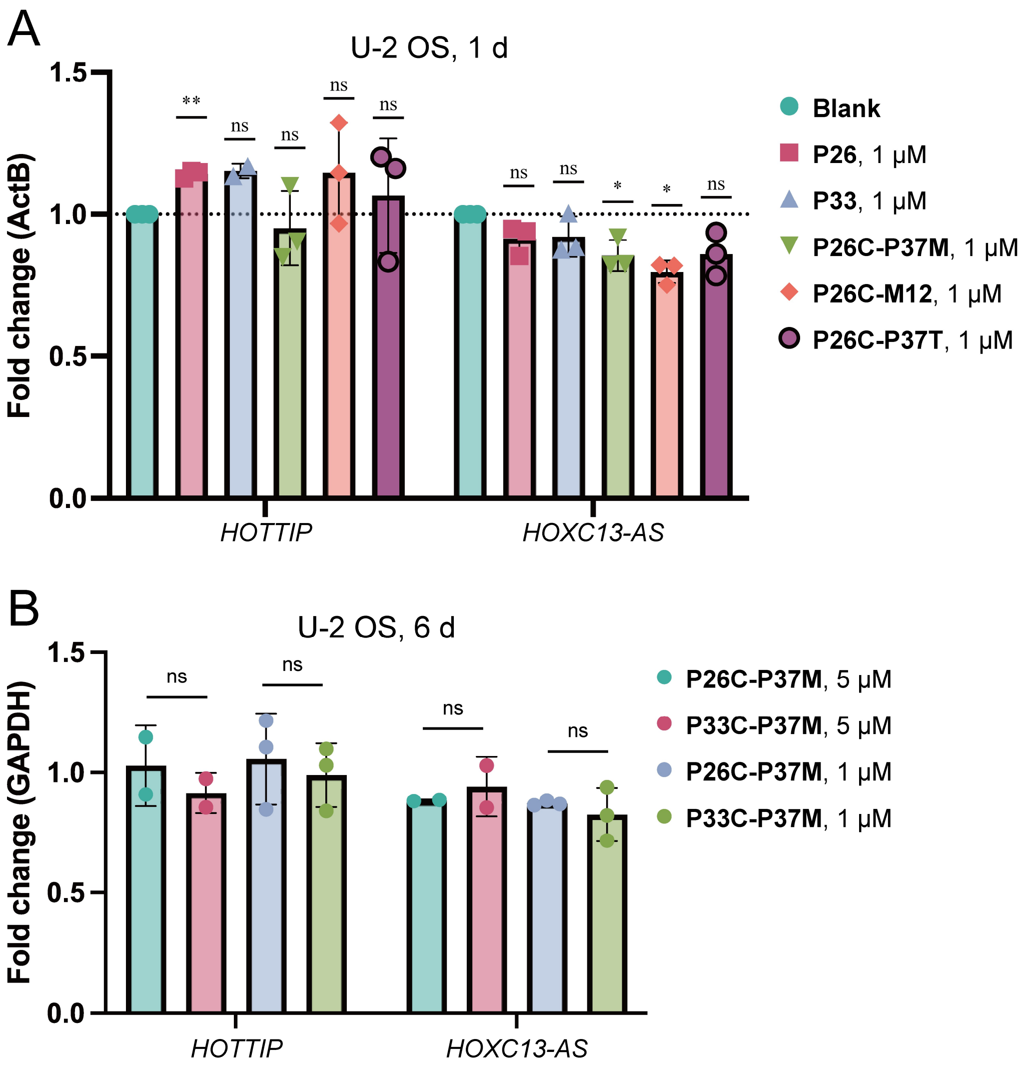


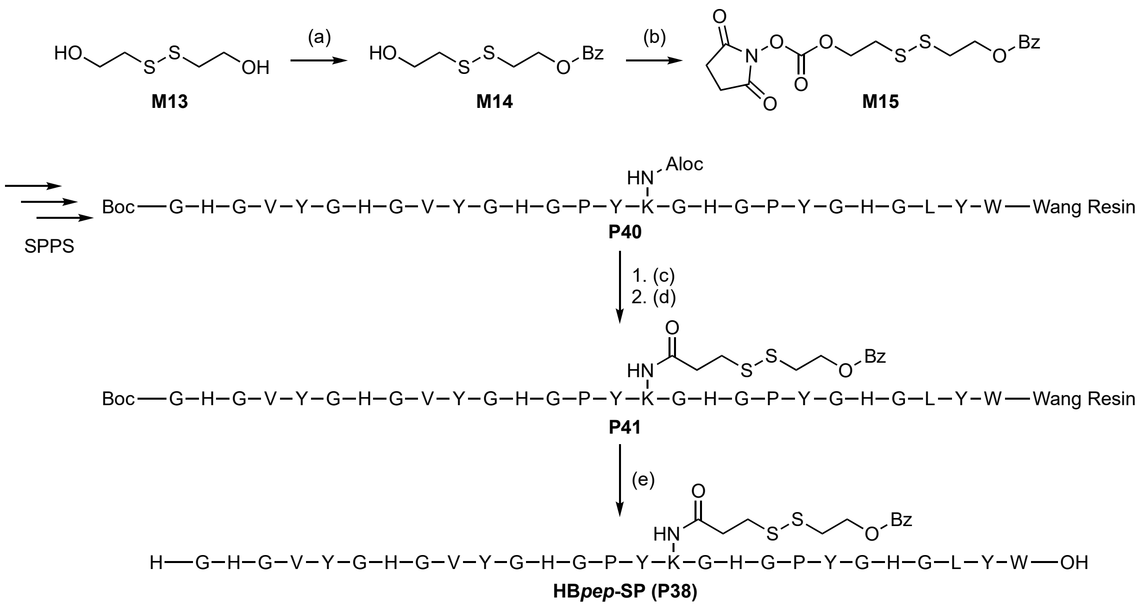
Figure 39. Results of treatment on U-2 OS. A. Cells were treated 1 μM for one day (N = 3). B. Cells were treated twice with 5/1 μM for three days (N = 2/3). RT-qPCR results were analyzed by ΔΔCt method. The significance of the individual samples was analyzed by one-sample Student t-test with the target value set at one, and the significance of differences between samples was analyzed by Welch's t-test. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

Treatment of these CPP conjugates with 10 µM on MDA-MB-231 was tested under different incubation times, including one, two or three days. However, no changes in the selected lncRNAs were detected.

3.4.2.2 Phase-Separating Peptide to Facilitate Cell Uptake

Since the first attempts to use CPP conjugates did not work, several alternative methods to promote cellular uptake were tested on U-2 OS. A phase-separating peptide, **HB*pep*-SP** (**P38**, Scheme 6), was used for direct cytosolic delivery[102]. Briefly, **HB*pep*-SP** is a peptide that can undergo liquid-liquid phase separation and form peptide coacervates in a basic environment. During the coacervate formation, this peptide coacervate can recruit peptides into the microdroplets and deliver the cargo to the cytoplasm via an endocytosis-independent pathway[112]. The reducing environment in the cytoplasm can reduce the disulfide bridge of self-immolative group, leading to the exposure of the lysine side chain, which increases the pH required for coacervation to occur, thus leading to the breakdown of coacervate and releasing the cargos in the cytoplasm. The broad ability to transport peptides, folded and functional proteins, and mRNA is noteworthy.

The protocol for the synthesis of **HB*pep*-SP** was slightly modified to allow the self-immolative side chain (**M14**) to be coupled to the peptide on the bead, thus reducing the purification process (Scheme 6). Briefly, Fmoc-K(Alloc)-OH is used so that the lysine residue on the bead can be selectively deprotected, and then reacted with self-immolative side chain. Boc-Gly-OH is used for the final residue so that the N-terminus can be deprotected along with any other protecting groups in the global cleavage.



Scheme 6. Synthesis scheme of **HB*pep*-SP**[102]. (a) Bz_2O, NEt_3, DMAP, DCM. (b) *N*,*N'*-disuccinimidyl carbonate, NEt_3, DMAP, THF/DMF. (c) $Pd(PPh_3)_4$, $PhSiH_3$, DCM. (d) **M14**, DIPEA, DMF. (e) $TFA/H_2O/TIPS$.

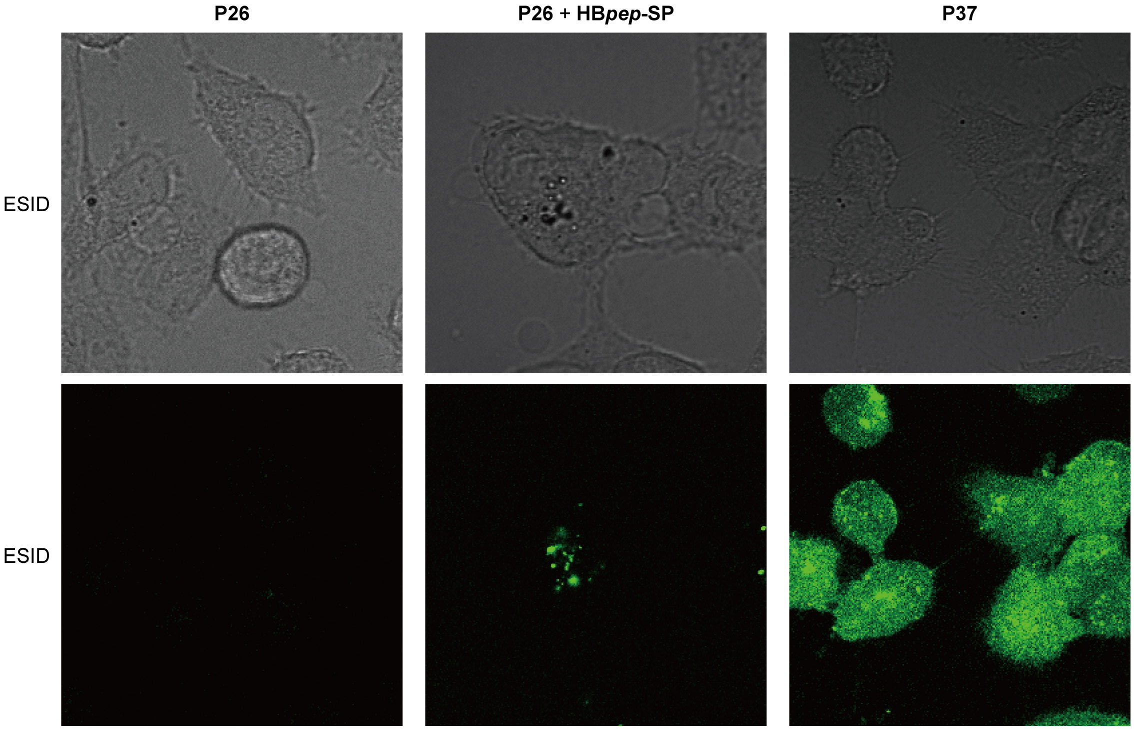


Figure 40. Results of fluorescence confocal microscopy. Pictures of U-2 OS taken by Electronically switchable illumination and detection module (ESID) or GFP channel. The pictures were taken using a ZEISS LSM 800.

First, coacervate formation was tested by measuring OD600 to determine the turbidity and indeed a solution containing **HB*pep*-SP** became turbid when the pH was adjusted to pH 6.5, which is the pH required for coacervate formation. Second, the loading efficiency for macrocycle **P26** into the coacervates was tested and an estimated 50% of the cargo was recruited when 2 µmol of macrocycle **P26** was mixed with 20 µg of **HB*pep*-SP**. Third, fluorescence confocal microscopy was performed after four hour of treatment of U-2 OS with coacervate and indeed the coacervate internalization was observed, but the microdroplet was not yet disassembled (Figure 40). However, treatment of **P26** with **HB*pep*-SP** did not alter the expression level of *HOTTIP* in U-2 OS cells after three days of incubation.

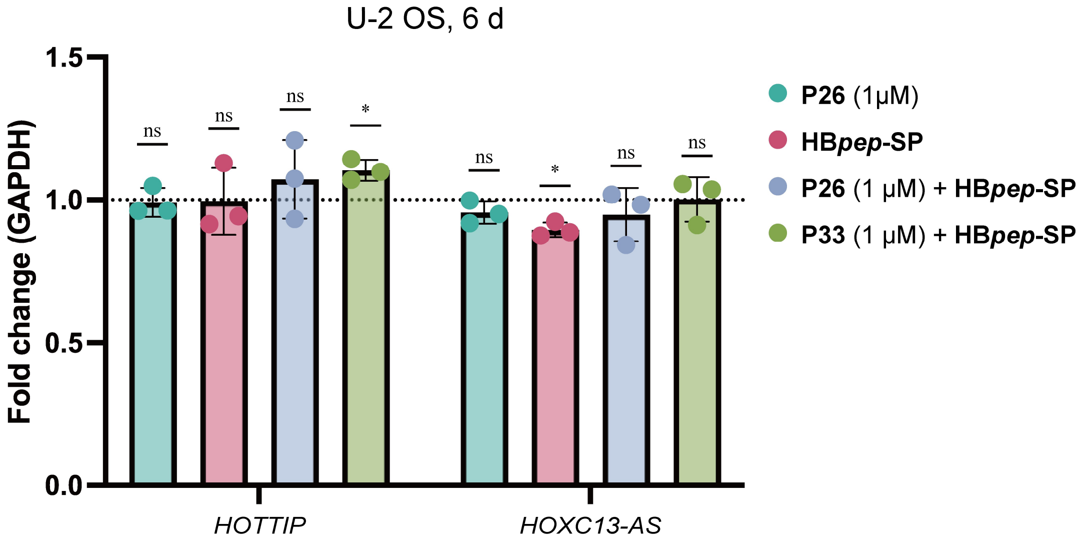


Figure 41. Results of using **HB*pep*-SP** to deliver peptide into U-2 OS. The treatment was performed twice with three days interval (N = 3). RT-qPCR results were analyzed by ΔΔCt method. The significance of the individual samples was analyzed by one-sample Student t-test with the target value set at one. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

3.4.2.3 Comparing Cell Lines

Although internalization was observed for both **P26C-P37M** and **P26** combined with **HB*pep*-SP** strategy, the lncRNAs expression level remained the same. Considering that the expected change was not observed, a positive control was needed to determine whether the inhibitor or the system did not work. The small molecule inhibitors **OICR-9429** and **7k** were tested to verify whether the problem was due to the peptidic nature of our inhibitors. Meanwhile, the MDA-MB-231 cell line was tested in parallel to see if the problem was with the cell type (Figure 42). When cells were treated with the WIN inhibitor **OICR-9429** at 5 µM, a strong reduction in *HOTTIP* was observed for MDA-MB-231 in a three-day treatment, but only a slight and non-significant reduction in *HOTTIP* was observed for U-2 OS in a double three-day treatment. On the other hand, treatment of MDA-MB-231 with WBM inhibitor **7k** at 5 µM showed a non-significant reduction of *HOTTIP* expression level in a three-day experiment, and no effect on *HOTTIP* expression level when U-2 OS was treated double three-day incubation. Overall, the result showed that the *HOTTIP* expression level of MDA-MB-231 was sensitive to inhibition of WDR5 in the single-digit micromolar concentration range, while U-2 OS was rather insensitive under the current conditions.

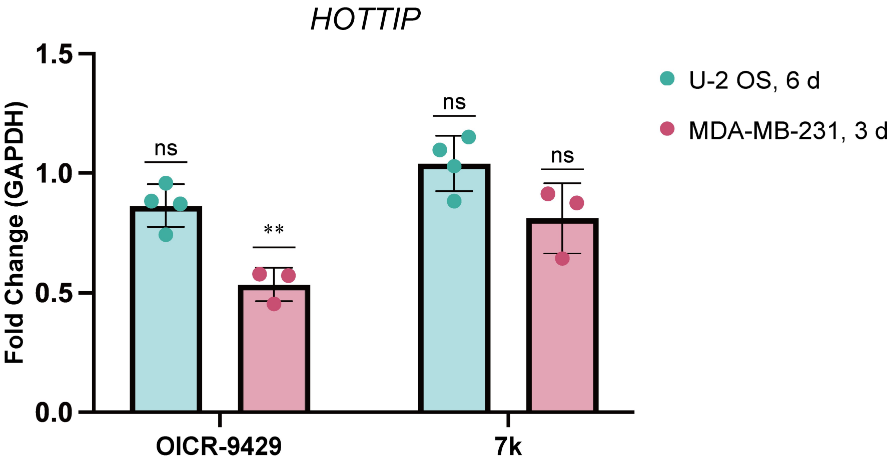


Figure 42. Results of treatment of different cells. U-2 OS was treated twice with 5 µM for three days (N = 4). MDA-MB-231 was treated once with 5 µM for three days (N = 3). RT-qPCR results were analyzed by ΔΔCt method. The significance of the individual samples was analyzed by one-sample Student t-test with the target value set at one. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

3.4.2.4 Macropinocytosis

MDA-MB-231 is a macropinocytic cancer cell and can consume large extracellular materials such as cell debris via macropinocytosis[113]. This also explains the result observed in Figure 37 that the fluorescence intensity and therefore peptide uptake is independent of

CPP conjugation. As a result, a series of experiments based on macropinocytosis were performed.

Since MDA-MB-231 is a macropinocytic cell type, it can undergo macropinocytosis without further stimulation. Therefore, the first experiment was direct treatment with peptide w/wo CPP conjugation (**P26** or **P26C-P37M**). However, *HOTTIP* expression levels remained the same in the three-day treatment, but the positive control treatment of 10 μM **OICR-9429** and **7k** significantly reduced *HOTTIP* expression (Figure 43).

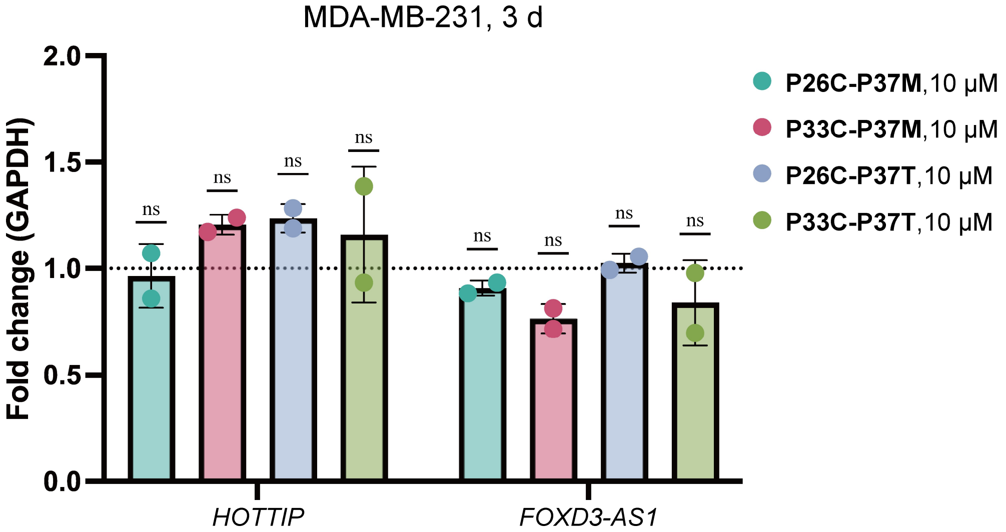


Figure 43. Results of treatment on MDA-MB-231. Cells were treated at 10 μM for three days (N = 2). RT-qPCR results were analyzed by ΔΔCt method. The significance of the individual samples was analyzed by one-sample Student t-test with the target value set at one. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

To exclude the possibility that the peptides were not able to escape from the endosome, **LK15-oxP4A** (**P42**), a macropinocytosis inducing peptide oxP4A conjugated to the membrane-lytic sequence LK15, was used as an additive to enhance macropinocytosis while promoting endosomal escape at the same time[99]. Nevertheless, the combined treatment of 10 μM **P26** and **LK15-oxP4A**(**P42**) provided similar results as for the direct treatment where the lncRNA expression level remained unchanged.

3.4.2.5 Transfection by Electroporation

The final attempt to deliver **P26** into the cell was done by electroporation, a technique that creates pores in the cell membrane and can be used for transfer of large biological molecules such as DNA, RNA, and proteins into the cell. In a nutshell, cells are resuspended between two electrodes and treated with an electric pulse that created pores in the membrane and allows material to diffuse through. Indeed, electroporation led to a high level of internalization when using MDA-MB-231 cell when **P26** was used at 5 or 20 µM (Figure 44). However, the lncRNA expression level again did not change comparison to the blank or negative control **P33** for one, two or four days of incubation, indicating that the peptide inhibitors could not provide the desired treatment effect for the current treatment conditions (Figure 45).

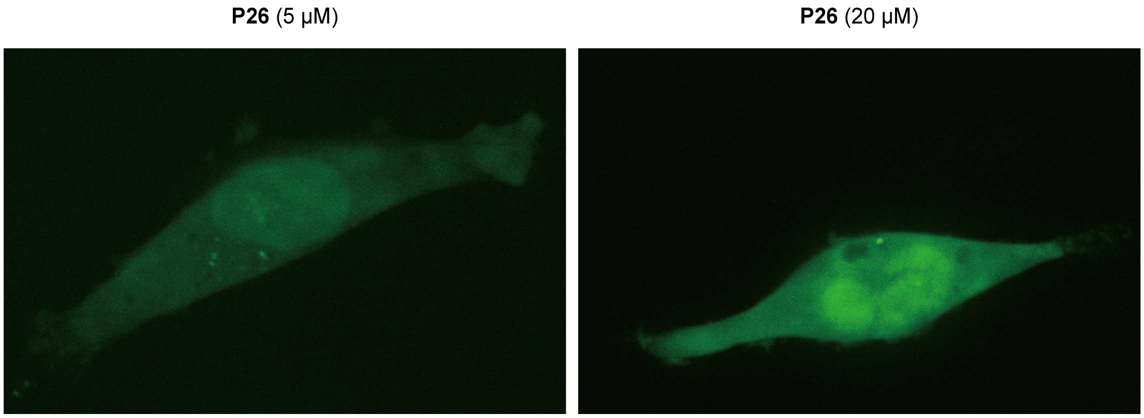


Figure 44. Confocal fluorescence microscopy images of MDA-MB-231 after electroporation with 5 or 20 µM **P26** taken through the FAM channel. The images were taken by using a 3i Marianans Confocal Fluorescence Microscope.

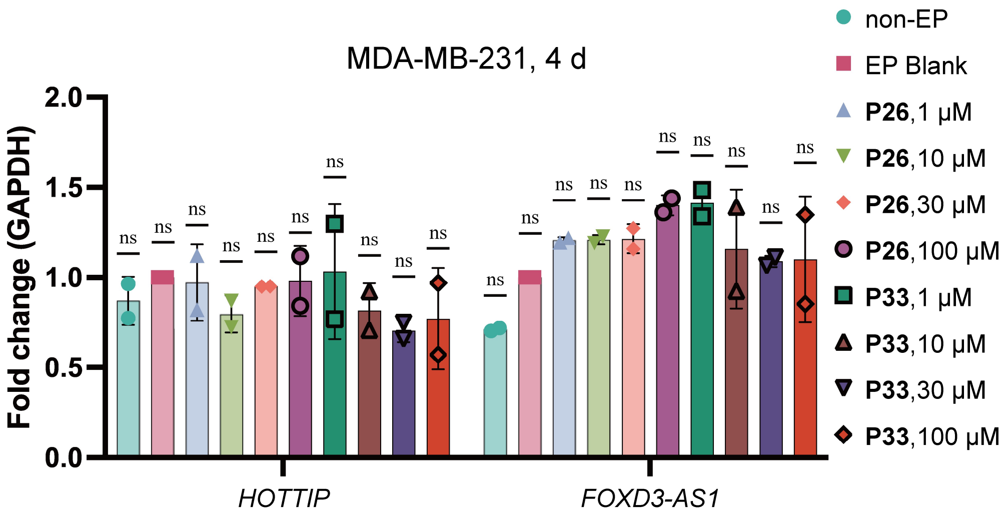


Figure 45. Results of using electroporation (EP) to deliver peptides into MDA-MB-231. Cells were treated at several concentrations for four days (N = 2). RT-qPCR results were analyzed by ΔΔCt method. Samples were normalized to EP Blank. The significance of the individual samples was analyzed by one-sample Student t-test with the target value set at one. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

3.5 Effect of Small Molecule WDR5 Inhibitors on lncRNA Expression *in cellulo*

The small molecule inhibitors demonstrated their potential to reduce lncRNA expression levels in MDA-MB-231 cells in Section 3.4.2.3. Next, several conditions were tested to further characterize the effect of small molecule inhibitor treatment. In addition, the difference between targeting WIN or WBM sites was also investigated.

3.5.1 Time Dependency of the Treatment

A time-dependent experiment was performed to investigate the kinetics of lncRNA expression. The small molecule inhibitors were used at 10 μM and incubated for one or three days (Figure 46). Both **OICR-9429** and **7k** showed a clear time-dependent effect on the downregulation of *HOTTIP* (Figure 46. A), with the three-day treatment providing a greater reduction than the one-day treatment. Interestingly, only **OICR-9429** was able to reduce the expression level of *FOXD3-AS1* in the one-day and three-day treatments, while

7k did not significantly affect the expression level (Figure 46. B).

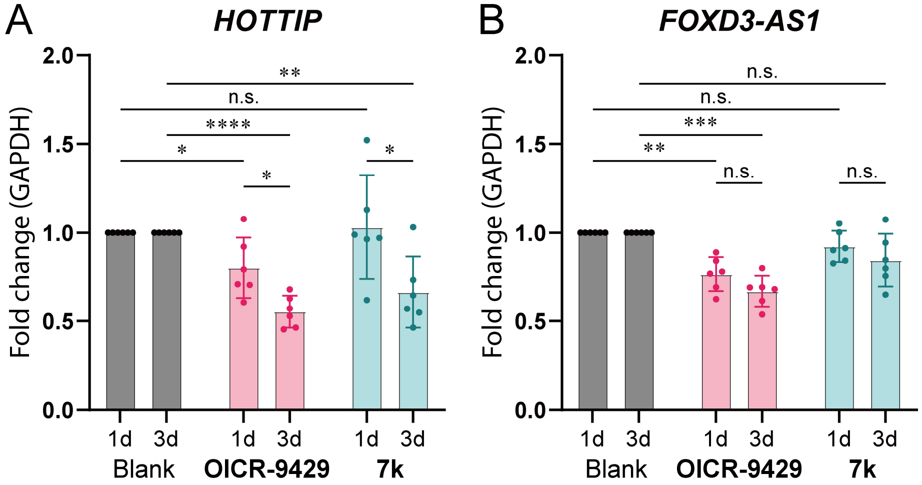


Figure 46. Effect and time dependency of lncRNA expression level after treatment with small molecule WDR5 inhibitors for three days (N=6). A. Effect of small molecule treatment on *HOTTIP* expression level. B. Effect of small molecule treatment on *FOXD3-AS1* expression level. The blank was treated with 0.1% DMSO. RT-qPCR results were analyzed by ΔΔCt method. Target genes were normalized to *GAPDH* and then normalized to blank. The significance of individual samples was analyzed by one-sample Student t-test with the target value set at one, and the significance of differences between samples was analyzed by Welch's t-test. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$, ***: $0.001 \geq p > 0.0001$, ****: $0.0001 \geq p$.

3.5.2 Concentration Dependency of the Treatment

A concentration dependency experiment was performed on MDA-MB-231 with an incubation time of three days to monitor the effect *in cellulo* (Figure 47). Treatment with the WIN inhibitor **OICR-9429** at 1 μM resulted in a significant downregulation of *HOTTIP*, and the effect of *HOTTIP* reduction was stronger but started to saturate at 5 μM (Figure 47. A). A similar effect was observed for *FOXD3-AS1* for **OICR-9429** treatment (Figure 47. B).

On the other hand, treatment of MDA-MB-231 with **7k** showed a similar result in *HOTTIP* downregulation. Although treatment with 1 or 5 μM **7k** did not result in significant downregulation, 10 μM resulted in a significantly reduced expression level of *HOTTIP*, and the level of reduction was similar to that of 5/10 μM **OICR-9429** treatment on *HOTTIP* expression level (Figure 47. A). Similar to the results shown in Figure 46. B, treatment with **7k** did not result in significant changes in *FOXD3-AS1* among the three concentrations tested (Figure 47. B).

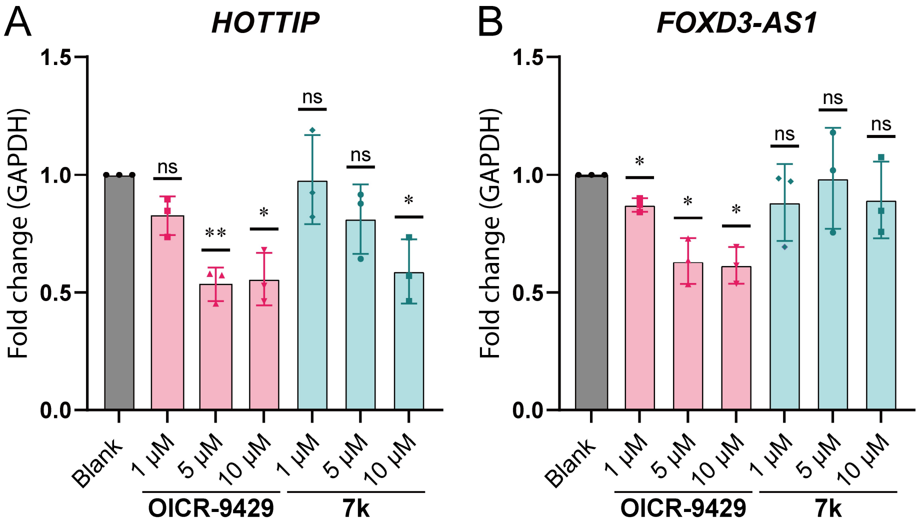


Figure 47. Effect and dose-dependency of lncRNA expression level after treatment with small molecule WDR5 inhibitors for three days (N=3). A. Effect of small molecule treatment on *HOTTIP* expression level. B. Effect of small molecule treatment on *FOXD3-AS1* expression level. The blank was treated with 0.1% DMSO. RT-qPCR results were analyzed by ΔΔCt method. Target genes were normalized to *GAPDH* and then normalized to the blank. The significance of individual samples was analyzed by one-sample Student t-test with the target value set at one, and the significance of differences between samples was analyzed by Welch's t-test. ns: $p > 0.05$, *: $0.05 \geq p > 0.01$, **: $0.01 \geq p > 0.001$.

3.6 Discussion

In this chapter, the allosteric control between WIN and WBM sites was evaluated by testing both peptide and small molecule WDR5 inhibitors selective for each site. The results showed that binding events between WIN and WBM sites to their respective inhibitors are independent and no off-target inhibition was detected. The *iv*-RIP experiments showed similar results that targeting the WBM site can disrupt lncRNA-WDR5 complex formation *in vivo*, while targeting the WIN site by **OICR-9429** was not significant although a decrease in average enrichment was observed.

Several experiments were performed to verify that targeting WDR5 could reduce lncRNA expression in cells. First, peptide inhibitors were conjugated to CPPs and their internalization was evaluated by flow cytometry and confocal microscopy. The level of cellular uptake was increased compared to the native peptide and the peptide was not trapped in endosomes. However, no effect on lncRNA expression was observed in U-2 OS or MDA-MB-231. A phase-separating peptide coacervate **HB*pep*-SP** (**P38**) was tested and internalization was observed, but still no effect on the expression level of lncRNAs in U-2 OS could be detected. Further tests with spontaneous macropinocytosis or induced macropinocytosis using **LK15-oxP4A** (**P42**) did not change the expression level of lncRNAs in MDA-MB-231. Finally, transfection was performed by electroporation, and clear internalization with wide distribution was observed. However, no effects on lncRNAs expression level could be detected. From this we concluded that the peptide did not work under the current treatment conditions.

Finally, small molecule inhibitors were tested in parallel and the results showed that U-2 OS was insensitive to compound treatment at the concentrations tested, while MDA-MB-231 showed a strong down-regulation of *HOTTIP*. To monitor the difference between targeting the WIN and WBM sites, time-dependent and concentration-dependent experiments were performed. In the time-dependent experiment, the effect of **OICR-9429** and **7k** treatment on *HOTTIP* was still evolving from one-day to three-day incubation, while the *FOXD3-AS1* expression level remained similar between one-day and three-day treatments. In the concentration dependency experiment, treatment with **OICR-9429** resulted in strong downregulation of *HOTTIP* and *FOXD3-AS1*. However, treatment with **7k** decreased the expression level of *HOTTIP* at 10 μM, while *FOXD3-AS1* was not affected at all. Combining the previous results, *HOTTIP* was sensitive to both WIN and WBM inhibition, while *FOXD3-AS1* was sensitive to inhibition at the WIN site but not at the WBM site.

In conclusion, a strategy to target RNA-protein interaction was demonstrated and indeed targeting *HOTTIP*-WDR5 complex leads to downregulation of *HOTTIP in cellulo*.

Chapter 4. Summary

4.1 Conclusion

In this work, a strategy and workflow for the design of an inhibitor of RNA-protein interactions was demonstrated. When an RPI shares the same pocket with a PPI, a PPI-derived peptide inhibitor could be used as a competitive inhibitor to disrupt the RPI. The peptide-based inhibitors can be used in high throughput screening (HTS) to find small molecule inhibitors. A competitive *iv*-RIP experiment could be used to verify the disruption of RNA-protein complex formation *in vitro*. In the *iv*-RIP experiment, the tolerance of using RNA extracts from cell lysates helps to avoid the use of high purity *in vitro* transcribed lncRNA, thus avoiding the potential problem of lncRNA misfolding during *in vitro* transcription.

Numerous attempts were made to demonstrate our hypothesis that disruption of lncRNA-protein interactions could lead to downregulation *in cellulo* using macrocycle **P26**. Strategies such as conjugation with a CPP moiety to enhance cell permeability, delivery with a phase-separating coacervate peptide, and electroporation to induce membrane permeability were all tested and demonstrated good internalization. Furthermore, electroporation showed a strong signal in the nucleus, indicating that accumulation in the nucleus was not the problem for the peptides. However, none of these strategies reduced the expression level of *HOTTIP*. Considering that small molecules require three days of incubation to develop downregulation, a single treatment method may not provide enough dose to induce changes, this could be further influenced by the fact that peptides tend to have shorter half-lives in cells.

The therapeutic effect at the cellular level was demonstrated by treatment with small molecule inhibitors **OICR-9429** and **7k**. Treatments of MDA-MB-231 could lead to downregulation of lncRNA, while U-2 OS was insensitive to these treatments, suggesting that different cells have different control methods of lncRNA expression levels. In addition, it is interesting to find that different lncRNA also have different dependencies on the disruption of lncRNA-WDR5 complexes. The expression of *HOTTIP* requires fully functional WDR5 with free WIN and WBM sites, while the expression of *FOXD3-AS1* tolerates the inaccessibility of WBM sites.

4.2 Future Outlook

Our results demonstrate that the lncRNA-WDR5 interaction is a druggable target. This method could be used to target other lncRNA-protein interactions if complex formation is required for lncRNA to exert its epigenetic modulatory function. Since different cell lines have different sensitivity to treatment and lncRNA have higher tissue specificity, it is possible to find a candidate that targets only cancer cells without affecting normal cells. Overall, I hope this work can inspire other scientists studying lncRNA-protein interactions and develop a new strategy for cancer treatment.

Chapter 5. Materials and Methods

5.1 General Information

All commercially available reagents and solvents were purchased and used directly without any further treatment unless specified. Solvents for chromatography were laboratory grade. Thin-layer chromatography (TLC) was performed using Supelco silica gel aluminum sheets with F254 indicator, visualized by irradiation with UV light. Column chromatography was performed using Acros Organics silica gel 60 Å (particle size 0.035 – 0.070 mm). A Thermo Scientific HAAKE EK90 with an acetone bath in a Dewar vessel was applied together to cool reactions to – 78 ºC. Preparative HPLC was performed using a BÜCHI Pure C-850 FlashPrep equipped with a C18 column (Macherey-Nagel, 5 μm, 125 x 21 mm, flow rate: 20 mL/min). Analytical HPLC-MS was performed by using Agilent 1200 Infinity II LC system equipped with a C18 column (Agilent Poroshell 120, 2.7 μm, 3 x 100 mm, flow rate: 0.4 mL/min, temperature: 33ºC) and Agilent InfinityLab LC/MSD G6125C, and the compounds were eluted with a linear gradient from 5% to 95% (percentage of B. A: 0.1% TFA in H2O; B: 0.1% TFA in ACN) over 20 min. The whole UV-VIS spectrum was recorded during the run, and the selected channel (210 / 254 nm), was used to plot and calculate the purity of the samples. High-resolution mass spectra were recorded on LTQ-XL Orbitrap mass spectrometer coupled to an Accela HPLC System (HPLC column: Hypersyl GOLD, 50 mm x 1 mm, 1.9 μm). ^{1}H-NMR and ^{13}C-NMR, COSY-NMR, HSQC-NMR, HMBC-NMR were recorded on a Bruker Avance III HD NanoBay (400 MHz) or a Bruker Avance III HD (500 MHz, 600 MHz, or 700 MHz) spectrometer in $CDCl_3$ or DMSO-d6. Chemical shifts (d) are reported in ppm, coupling constants (J) are reported in Hertz (Hz), splitting patterns are indicated as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet). Gels were imaged using a Bio-Rad ChemiDoc MP imaging system. Reverse transcription (RT) was performed by Eppendorf Mastercycler ep gradient, and quantitative polymerase chain reaction (qPCR) was performed by Bio-Rad CFX Connect Real-Time PCR Detection System.

5.2 Preparation of peptide

In general, peptides were synthesized through the conventional Fmoc strategy, and Rink Amide AM resin (0.3 – 0.6 mmol/g) was used as solid support unless otherwise specified. The reactions were conducted at room temperature unless otherwise specified.

5.2.1 Method A – Manual linear peptide synthesis

Fmoc was deprotected with 25% Piperidine in DMF for 10 min twice. The new amino acid was coupled by incubating resin with 4 eq amino acids, 4 eq PyBOP, and 8 eq DIPEA in DMF for 40 min. Ac capping was done after each coupling for peptides with incomplete coupling by incubating resin with 10 eq Ac_2O, and 10 eq DIPEA in DMF for 15 min.

5.2.2 Method B – Automated linear peptide synthesis

Peptide synthesis was performed using an automated Syro I parallel peptide synthesizer. Fmoc was deprotected by incubating resin with 25% Piperidine in DMF for 3 min twice. The new amino acid was coupled by incubating resin with 4 eq amino acids, 4 eq PyBOP, and 8 eq DIPEA in DMF for 1 h. Ac capping was performed by incubating resin with 10 eq Ac_2O, and 10 eq DIPEA in DMF for 15 min.

5.2.3 Method C – Automated linear peptide synthesis

Peptide synthesis was performed using an automated PurePep Chorus parallel peptide synthesizer. Fmoc was deprotected by incubating resin with 25% Piperidine in DMF for 3 min twice at 50ºC. The new amino acid was coupled by incubating resin with 5 eq amino acid, 5 eq HCTU, and 10 eq DIPEA in DMF for 10 min at 75ºC. Ac capping was performed by incubating resin with 10 eq Ac_2O, and 10 eq DIPEA in DMF for 10 min.

5.2.4 Method D – Cyclization with Fmoc at N-terminus

Allyl/Alloc protecting groups were used for orthogonal side chain protection. The elongation of peptides was paused when the second position for cyclization was reached. The side chains were deprotected by incubating the resin with 0.25 eq $Pd(PPh_3)_4$, 25 eq $PhSiH_3$ in DCM for 30 min twice. Then the resin was washed with 0.5% sodium diethyldithiocarbamate in DMF for 5 min and a total of five times. The carboxylic acid group on the side chain was activated by incubating the resin with 2 eq PyAOP, and 2 eq HOAt in DMF for 15 min, then 4 eq 2,4,6-collidine was added directly to the reaction and continued for 48 h. Capping of the unreacted amine side chain was carried out by incubating the resin with 10 eq Ac_2O, and 10 eq DIPEA in DMF for 30 min. Sequence elongation was then continued after cyclization was done.

5.2.5 Method E – Cyclization with *o*-Ns at N-terminus

This protocol was applied to those sequences that suffer from self-deprotection during the cyclization process. Allyl/Alloc protecting groups were used for orthogonal side chain protection. The elongation of peptides was paused when the second position of cyclization was synthesized. Fmoc was deprotected with 25% Piperidine in DMF for 10 min twice. Reprotection was performed by incubating resin in 4 eq *o*-NsCl, 5 eq 2,4,6-collidine, and 5 eq DMAP in DMF for 15 min twice.[88] The side chain was deprotected by incubating resin with 0.1 eq $Pd(PPh_3)_4$, 0.4 eq PPh_3, and 8 eq pyrrolidine in DCM for 1 h.[115] Then the resin was washed with 0.5% sodium diethyldithiocarbamate in DMF for 5 min and a total of five times. The carboxylic acid group on the side chain was activated by incubating the resin with 2 eq PyAOP, and 2 eq HOAt in DMF for 15 min, then 4 eq 2,4,6-collidine was added directly to the reaction and continued for 48 h. Capping of the unreacted amine side chain was carried out by incubating the resin with 10 eq Ac_2O, and 10 eq DIPEA in DMF for 30 min. *o*-Ns group was deprotected by incubating resin with 10 eq 2-mercaptoethanol and 5 eq of DBU in DMF for 16 h. Sequence elongation was then continued after cyclization and *o*-Ns deprotection were done.

5.2.6 Modifications at N-terminus

In general, the N-terminus of the peptides was either capped with FITC-linker for use in FP assays, or with acetylation to make a fluorescence-label free version, or other special modification (3-mercaptopropionic acid or 4-maleimidobutyric acid) for the following solution-phase conjugation reactions.

5.2.6.1 FITC labeling

Sequences were further elongated with Fmoc-Linker-OH after the sequence was done. Fmoc was deprotected with 25% Piperidine in DMF for 10 min twice. FITC labeling was performed by incubating resin with 2 eq 6-FITC, 4 eq DIPEA in DMF for 16 h.

5.2.6.2 Ac capping at N-terminus

After the sequence was fully elongated, Fmoc was deprotected with 25% Piperidine in DMF for 10 min twice. Ac capping was performed by incubating resin with 10 eq Ac_2O, 10 eq DIPEA and DMF for 30 min.

5.2.6.3 Other capping at N-terminus

After the sequence was fully elongated, Fmoc was deprotected with 25% Piperidine in DMF for 10 min twice. The final carboxylic acid was coupled by incubating resin with 2 eq carboxylic acid, 2 eq PyBOP, and 4 eq DIPEA in DMF for 12 h.

5.2.7 Global cleavage

After sequence elongation and N-terminal modification, peptides without thiol groups were cleaved from the resin by incubating the resin in the cleavage solution A (TFA/H_2O/TIPS, 95/2.5/2.5), while peptides with a thiol group were cleaved from the resin by incubating them in the cleavage solution B (TFA/EDT/H_2O/TIPS, 90/5/2.5/2.5). Cleavage normally took 1 h, while peptides which contain Arg(Pbf) took 4 h. An additional cleavage cycle was performed for peptides that were difficult to deprotect. After the cleavage was complete, the resin was filtered off and the filtrate was concentrated by blowing N_2 over the surface followed by precipitation of the peptide in cold diethyl ether three times. The product was then purified by preparative LC with either acidic or basic eluent.

5.2.8 Preparative LC with acidic eluent

Crude peptides were resuspended in a solution of H_2O/ACN/DMSO/TFA (70/20/10/0.1), centrifuged to remove insoluble material, then purified by reverse-phase preparative LC. Products were eluted using a linear gradient from 10% to 60% (percentage of B. A: 0.1% TFA in H_2O; B: 0.1% TFA in ACN) over 60 min. The elution was monitored using a diode array detector at 254 nm for the FITC-labeled peptides, or 210 nm for non-FITC-labeled peptides. The pure fractions were pooled together and lyophilized to give the desired products.

5.2.9 Preparative LC with basic eluent

The purification with alkaline eluent was used for peptides that didn’t dissolve in an acidic aqueous solution. Crude peptides were resuspended in a solution of H_2O/MeOH (90/10), and the pH was adjusted by gradually adding 200 mM $NH_4HCO_{3(aq.)}$ until the overall pH was above 7. The crude solution was centrifuged to remove insoluble material, then purified by reverse-phase preparative LC. Products were eluted using a linear gradient from 5% to 60% (percentage of B. A: 15 mM $NH_4HCO_{3(aq.)}$; B: MeOH) over 60 min. The elution was monitored using a diode array detector at 254 nm for the FITC-labeled peptides, or 210 nm for non-FITC-labeled peptides. The pure fractions were pooled together and lyophilized to give the desired products.

5.2.10 Solution phase conjugation

Some of the longer peptides were synthesized in fragments and combined using conjugation protocol to improve the efficiency of screening for a membrane permeability strategy while preventing the intrinsic limitation of SPPS protocol to synthesize difficult peptides.

5.2.10.1 Conjugating by Thiol-Maleimide reaction

Both starting peptides bearing mercapto or maleimido groups were purified as previously described before performing the conjugation reaction in solution phase. In general, a solution of 1.3 eq maleimido peptide in 100 mM $NH_4HCO_{3(aq)}$, pH 9.0 buffer were added to a solution of 1 eq mercapto peptide which was dissolved in 100 mM $NH_4HCO_{3(aq)}$, pH 9.0 buffer and the reaction was stirred at rt for 1 – 4 h. After the mercapto peptide was depleted (monitored by HPLC-MS), the reaction was lyophilized, and the residue was purified using the preparative HPLC method with acid eluent (5.2.8) to afford the desire product.

5.2.10.2 Conjugating by disulfide bond formation

Both starting materials bearing mercapto groups were purified as previously described before performing the conjugation reaction in solution phase. In general, the more available substrate (1 eq) was first reacted with 2,2’-Dipyridyl disulfide (3 eq) in methanol for 12 h, then purified with either a manual column or preparative HPLC to afford activated material A[108]. Solution phase selective disulfide bond formation was carried out by stirring 1 eq peptide B and 2 eq activated material A in a solution of 50% methanol with 50% 100 mM $NH_4HCO_{3(aq)}$, pH 9.0 buffer at 4ºC for 4 h. After the peptide B was depleted (monitored by HPLC-MS), the reaction was lyophilized, and the residue was purified by the preparative HPLC with acid eluent (5.2.8) to afford the desire product.

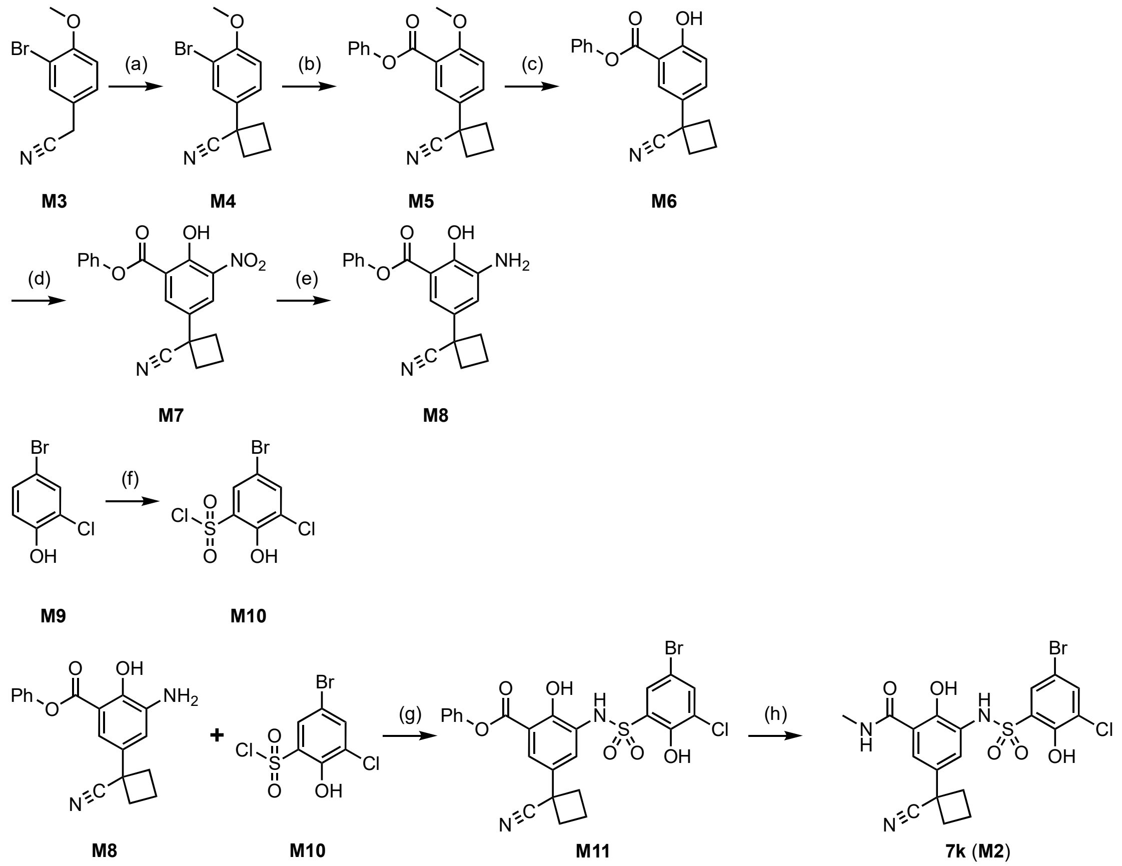
5.2.11 Preparation of HB*pep*-SP

The synthetic route for HB*pep*-SP was slightly modified from the original method reported by Sun et al.[102] In short, peptides HB*pep* were synthesized through the conventional Fmoc strategy, and Wang resin (1g, 0.5 mmol/g) was used as solid support. The first amino acid was loaded by incubating the resin with Fmoc-AA (10 eq), DIC (5 eq) and DIPEA (10 eq) in DMF for 4 h at rt. The sequence was elongated with the Method C, which was described in 5.2.3. Fmoc-His(Boc)-OH, Fmoc-Lys(Alloc)-OH and Boc-Gly-OH were used during the sequence elongation instead of the conventional protecting groups for better performance and selective deprotection/modification. After the sequence was fully elongated, Alloc was deprotected by incubating resin with 0.2 eq $Pd(PPh_3)_4$, 20 eq $PhSiH_3$ in DCM for 1 h[114]. Then the resin was washed with 0.5% sodium diethyldithiocarbamate in DMF for 5 min and a total of five times. After the resin was washed with DMF, Boc-HB*pep*-resin (**P40NH$_2$**, approximately 0.5 mmol) was ready for conjugation.

In the meantime, the self-immolative side chain was prepared through a procedure similar to the reported method. The mono-benzylation was carried out by slowly adding a solution of Bz_2O (1.35 g, 6 mmol, 0.6 eq) in 9 mL DCM to a solution of 2-hydroxyethyl disulfide (1.08 mL, 9 mmol, 1 eq, **M13**) DMAP (110 mg, 0.9 mmol, 0.1 eq) and NEt_3 (1.25 mL, 9 mmol, 1 eq) in 13 mL DCM while stirring at 0ºC. After the addition was finished, the reaction was continued at rt for another 3 h. After the reaction was complete (monitored by TLC), it was diluted with EtOAc, then washed with $NaHCO_{3(sat.\ aq)}$, $NH_4Cl_{(sat.\ aq)}$, water and brine successively. The organic layer was collected, dried over $MgSO_4$, concentrated, and purified with normal phase column chromatography (petroleum ether / EtOAc gradient from 19:1 to 3:2) to afford the mono-benzylated product (1.14 g, 4.4 mmol, yield 73%, **M14**). Then the mono-benzylated product (1.03 g, 4 mmol, 1 eq) was added to a solution of *N*,*N′*-disuccinimidyl carbonate (2.05 g, 8 mmol, 2 eq), DMAP (48.8 mg, 0.4 mmol, 0.1 eq) and NEt_3 (1.12 mL, 8 mmol, 2 eq) in THF/DMF (24 mL, 4:1) at 0ºC and stirred for 10 min. Then the reaction was continued at rt for another 16 h. After the reaction was completed (monitored by TLC), it was concentrated, diluted with EtOAc and washed with $NH_4Cl_{(sat.\ aq)}$, water and brine, successively. The organic layer was collected, dried over $MgSO_4$, concentrated, and purified with normal phase column chromatography (petroleum ether / EtOAc gradient from 19:1 to 3:2) to afford the activated carbonate product (1.43 g, 3.58 mmol, yield 89.5% **M15**).

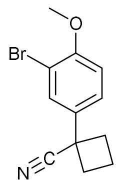
The assembly of self-immolative group was coupled by incubating the Boc-HB*pep*-resin (0.25 mmol, 1 eq, **P40NH$_2$**) with activated succinimidyl carbonate (150 mg, 0.375 mmol, 1.5 eq, **M15**) and DIPEA (435 µL, 2.5 mmol, 10 eq) in DMF for 24 h. After the peptide was ready, it was cleaved and deprotected by the global cleavage protocol (5.2.7) with the TFA/H_2O/TIPS method, then purified by the preparative HPLC method with acid eluent (5.2.8) to afford the desire HB*pep*-SP (12.2 mg, **P38**).

5.3 Preparation of small molecules



Supplementary Scheme 1. Synthetic scheme for compound **7k** (**M2**)[96]. (a) NaH, 1,3-dibromopropane, DMSO, rt, 24 h; (b) Phenyl formate, phenol, Pd(OAc)$_2$, P(*t*-Bu)$_3$ • HBF$_4$, NEt$_3$, ACN 90ºC, sealed tube, 24 h; (c) BBr$_3$, DCM, -78ºC. (d) HNO$_3$, TBAB, DCE/H$_2$O, 60ºC, 24 h; (e) H$_2$, Pd/C, MeOH, rt; (f) HSO$_3$Cl, 0ºC - rt; (g) Pyridine, DCM, rt; (h) NH$_2$Me, DIPEA, THF, 65ºC.

1-(3-Bromo-4methoxyphenyl) cyclobutane-1-carbonitrile (M4)



NaH (60% in mineral oil, 0.6g, 15 mmol, 3 eq) was added portion wise to a stirring solution containing 2-(3-Bromo-4-methoxyphenyl)acetonitrile (**S1**, 1.13g, 5 mmol, 1.0 eq) and 1,3-dibromopropane (612 μL, 6 mmol, 1.2 eq) in DMSO (50 mL). After the reaction was stirred for 16 h at rt, the reaction was quenched with EtOAc/Et_2O (1:1), then the organic phase was washed with water (3 X). The aqueous layer was back extracted with EtOAc. The organic layers were combined, washed with $\text{brine}_{(\text{sat.})}$, concentrated, and purified by normal phase column chromatography (petroleum ether / EtOAc gradient from 19:1 to 0:1) to provide the **S2** as yellowish oil (867 mg, 3.27 mmol, 65% yield).

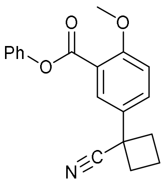
R_f = 0.30 (Petroleum ether: EtOAc = 9:1)

LRMS: calculated for $[M+Na]^+$ $C_{12}H_{12}BrNONa$: 288.0, found: 287.8

^{1}H NMR (600 MHz, Chloroform-*d*) δ 7.58 (d, *J* = 2.4 Hz, 1H), 7.32 (dd, *J* = 8.5, 2.4 Hz, 1H), 6.90 (d, *J* = 8.6 Hz, 1H), 3.91 (s, 3H), 2.84 – 2.76 (m, 2H), 2.63 – 2.53 (m, 2H), 2.48 – 2.35 (m, 1H), 2.12 – 2.00 ppm (m, 1H).

^{13}C NMR (151 MHz, Chloroform-*d*) δ 155.52, 133.31, 130.66, 125.92, 124.11, 112.16, 111.99, 56.38, 39.30, 34.77, 16.99 ppm.

Phenyl 5-(1-Cyanocyclobutyl)-2-methoxybenzoate (M5)



Phenyl formate (290 µL, 2.67 mmol, 2.0 eq) and NEt_3 (558 µL, 4.0 mmol, 3.0 eq) were added to a sealed tube reactor containing a solution of **S2** (335 mg, 1.34 mmol, 1.0 eq), $Pd(OAc)_2$ (15.1 mg, 0.067 mmol, 0.05 eq), $(tBu)_3PH{\cdot}BF_4$ (78.4 mg, 0.26 mmol, 0.2 eq) and phenol (117 µl, 1.34 mmol, 1.0 eq) in ACN (5 mL) with a stirring bar, then the reactor was sealed, heated to 90ºC and stirred for 24 h. After the reaction was cooled down to rt, the reaction was filtered through a pad of celite, and the filtrate was collected and concentrated. The residue was redissolved in EtOAc and washed with water, then the organic layer was collected, concentrated, and purified by normal phase column chromatography (petroleum ether / EtOAc gradient from 9:1 to 0:1) to provide the product as a colorless solid (**S3**, 328.9 mg, 1.07 mmol, 80% yield).

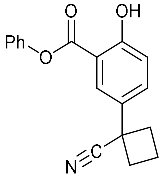
R_f = 0.04 (Petroleum ether: EtOAc = 9:1)

LRMS: calculated for $[M+H]^+$ $C_{19}H_{18}NO_3$: 308.1, found: 308.0

^{1}H NMR (700 MHz, Chloroform-*d*) δ 8.02 (d, *J* = 2.6 Hz, 1H), 7.59 (dd, *J* = 8.7, 2.6 Hz, 1H), 7.45 – 7.40 (m, 2H), 7.29 – 7.26 (m, 1H), 7.25 – 7.22 (m, 2H), 7.06 (d, *J* = 8.8 Hz, 1H), 3.95 (s, 3H), 2.88 – 2.83 (m, 2H), 2.68 – 2.60 (m, 2H), 2.50 – 2.40 (m, 1H), 2.13 – 2.06 ppm (m, 1H).

^{13}C NMR (176 MHz, Chloroform-*d*) δ 164.19, 159.47, 151.01, 131.80, 131.73, 129.58, 129.40, 126.03, 124.29, 121.97, 119.75, 112.92, 56.45, 39.59, 34.86, 17.18 ppm.

Phenyl 5-(1-Cyanocyclobutyl)-2-hydroxybenzoate (M6)



A solution of **S3** (123 mg, 0.4 mmol, 1.00 eq) in anhydrous DCM (1 ml) was cooled to – 78ºC in an acetone bath cooled using a HAAKE Ek90 device and kept under Argon. BBr_3 (1M in hexane, 0.8 ml, 0.8 mmol, 2 eq) was added dropwise to the stirring solution, and the reaction was continued for another 1 h at – 78ºC, then the reaction was allowed to warm up to rt. After the reaction returned to rt, it was quenched by pouring it over an ice/water slurry. The whole slurry was then extracted with EtOAc, and the organic phase was washed with water and brine successively. The organic phase was then collected, dried over anhydrous $MgSO_4$, concentrated, and purified by normal phase column chromatography (petroleum ether / EtOAc gradient from 19:1 to 0:1) to provide the product as a colorless oil (89.9 mg, 0.31 mmol, 76.5% yield).

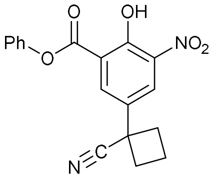
R_f = 0.26 (Petroleum ether: EtOAc = 9:1)

LRMS: not found

^{1}H NMR (700 MHz, Chloroform-*d*) δ 10.55 (s, 1H), 8.09 (d, *J* = 2.5 Hz, 1H), 7.59 (dd, *J* = 8.8, 2.5 Hz, 1H), 7.51 – 7.44 (m, 2H), 7.37 – 7.29 (m, 1H), 7.25 – 7.20 (m, 2H), 7.09 (d, *J* = 8.8 Hz, 1H), 2.89 – 2.83 (m, 2H), 2.68 – 2.59 (m, 2H), 2.51 – 2.41 (m, 1H), 2.14 – 2.06 ppm (m, 1H).

^{13}C NMR (176 MHz, Chloroform-*d*) δ 168.70, 161.92, 150.09, 133.94, 131.14, 129.87, 127.44, 126.75, 124.29, 121.80, 118.93, 111.99, 39.63, 34.79, 17.19 ppm.

Phenyl 5-(1-Cyanocyclobutyl)-2-hydroxy-3-nitrobenzoate (M7)



Compound **S4** (88 mg, 0.3 mmol, 1 eq) was dissolved in DCE (300 µl) and the solution was cooled with ice/water bath. A solution of tetrabutylammonium bromide (4.83 mg, 0.015 mmol, 0.05 eq) and concentrated nitric acid (38 µl) in H_2O (300 µL) was also added, followed by vigorous stirring at 60ºC for 24 h to carry out the reaction. The reaction was allowed to cool down to rt, then diluted with DCM, and the organic phase was washed with water, collected, dried over anhydrous $MgSO_4$, concentrated, and purified by normal phase column chromatography to provide the product as pale-yellow solid (71.5 mg, 0.21 mmol, 70% yield).

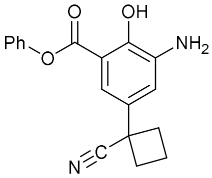
R_f = 0.84 (EtOAc)

LRMS: calculated for $[M+Na]^+$ $C_{18}H_{14}N_2O_5Na$: 361.1, found: 361.0

^{1}H NMR (500 MHz, Chloroform-*d*) δ 11.74 (s, 1H), 8.40 (d, *J* = 2.6 Hz, 1H), 8.29 (d, *J* = 2.5 Hz, 1H), 7.56 – 7.44 (m, 2H), 7.44 – 7.31 (m, 1H), 7.25 – 7.21 (m, 2H), 2.96 – 2.87 (m, 2H), 2.73 – 2.61 (m, 2H), 2.59 – 2.45 (m, 1H), 2.22 – 2.08 ppm (m, 1H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 167.06, 155.58, 149.71, 138.21, 133.60, 131.00, 130.04, 129.18, 127.22, 123.13, 121.51, 116.38, 39.33, 34.77, 17.18 ppm.

Phenyl 3-Amino-5-(1-cyanocyclobutyl)-2-hydroxybenzoate (M8)



Compound **S5** (66.5 mg, 0.196 mmol, 1 eq) was added to a slurry of Pd/C (5% by weight, 10 mol%) in a MeOH/EtOAc (1:1, 4 ml), then the hydrogenation was carried out by stirring under H_2 atmosphere for 16 h. After the starting material was consumed (monitored by TLC), the reaction was filtered through a pad of celite, the filtrate was collected, concentrated, and purified by normal phase column chromatography (petroleum ether / DCM gradient from 1:1 to 0:1, then DCM / EtOAc gradient from 9:1 to 0:1) to give the product (27.4 mg, 0.09 mmol, 45%).

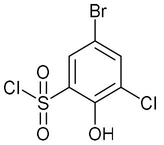
R_f = 0.19 (DCM)

LRMS: calculated for $[M+H]^+$ $C_{18}H_{17}N_2O_3$: 309.1, found: 309.0

^{1}H NMR (500 MHz, DMSO-d_6) δ 10.94 – 9.55 (br, s, 1H) 7.53 – 7.45 (m, 2H), 7.38 – 7.30 (m, 3H), 7.22 (d, *J* = 2.3 Hz, 1H), 7.01 (d, *J* = 2.4 Hz, 1H), 6.00 – 4.65 (br, s, 2H) 2.75 – 2.65 (m, 2H), 2.62 – 2.53 (m, 2H), 2.33 – 2.19 (m, 1H), 2.07 – 1.93 (m, 1H).

^{13}C NMR (126 MHz, DMSO-d_6) δ 168.29, 150.01, 148.08, 138.84, 130.88, 129.66, 126.40, 124.57, 122.08, 115.43, 112.58, 110.84, 39.73, 33.67, 16.71 ppm.

5-Bromo-3-chloro-2-hydroxybenzenesulfonyl Chloride (M10)



Chlorosulfuric acid (932 µl, 14 mmol, 7 eq) was cooled down to 0°C. 4-bromo-2-chlorophenol (**S7**, 514 mg, 2,48 mmol, 1 eq) was added portion-wise and stirred for 1 h at 0°C. After warming to room temperature and stirring for 23 h, the reaction was quenched by carefully pouring over a slurry of ice, DCM, and brine. The whole slurry was extracted with DCM, and the organic layer was collected, dried over anhydrous $MgSO_4$, concentrated then purified by normal phase column chromatography (petroleum ether / DCM gradient from 1:1 to 0:1) to provide the pure product as a white crystal (560 mg, 1.83 mmol, 73% yield).

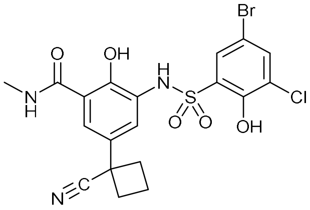
R_f = 0.89 (DCM)

LRMS: Not found.

^{1}H NMR (500 MHz, DMSO-d_6) δ 11.94 (s, 1H), 7.64 (d, *J* = 2.4 Hz, 1H), 7.48 ppm (d, *J* = 2.4 Hz, 1H).

^{13}C NMR (126 MHz, DMSO-d_6) δ 148.89, 133.52, 132.98, 128.15, 121.78, 109.18 ppm.

3-((5-Bromo-3-chloro-2-hydroxyphenyl)sulfonamido)-5-(1cyanocyclobutyl)-2-hydroxy-N-methylbenzamide (7k, M2)



To compound **M8** (27.4 mg, 90 µmol, 1 eq), **M10** (33.6 mg, 110 µmol, 1.2 eq) and pyridine (21.8 µl, 270 µmol, 3 eq) in DCM (0.45 ml) were added. The resulting reaction mixture was stirred for 5 h at room temperature. Additional compound **M10** (16.8 mg, 55 µmol, 0.6 eq) and pyridine (7.3 µl, 90 µmol, 1 eq) were added and the mixture was stirred for another 2 h at room temperature. The mixture was concentrated under reduced pressure, diluted with EtOAc, and washed with 0.1 M $HCl_{(aq)}$ and brine. The organic layer was collected, dried over anhydrous $MgSO_4$, concentrated under reduced pressure, then purified by normal phase column chromatography (MeOH / DCM gradient from 0:1 to 3:17) to give the intermediate product **M11** (20 mg, 35 µmol, 35% yield).

To a solution of compound **M11** (20 mg, 35 µmol, 1 eq) in THF (1 mL) was added with NH_2Me (2M in methanol, 86 µL, 173 µmol, 5 eq) in one portion. The reaction was sealed and stirred at rt for 1 h. The reaction was concentrated, then purified by preparative HPLC (0.1% TFA in H_2O / 0.1% TFA in ACN gradient from 18:7 to 1:49) to give the product **7k** (**M2**, 6.49 mg, 12.6 µmol, 37% yield).

Retention time on analytical HPLC: 15.21 min

Purity of product on analytical HPLC: 98.0% (UV 254 nm, see Supplemental Figure 41)

HRMS: calculated for $[M+H]^+$ $C_{19}H_{18}O_5N_3{}^{79}Br^{37}ClS$: 515.98041, found: 515.98074

^{1}H NMR (700 MHz, DMSO-*d*$_6$) δ 13.75 (br, s, 1H), 11.09 (br, s, 1H), 9.39 (br, s, 1H), 9.12 (d, *J* = 4.5 Hz, 1H), 7.93 (d, *J* = 2.5 Hz, 1H), 7.71 (d, *J* = 2.2 Hz, 1H), 7.71 (d, *J* = 2.5 Hz, 1H), 7.35 (d, *J* = 2.2 Hz, 1H), 2.82 (d, *J* = 4.6 Hz, 3H), 2.72 – 2.66 (m, 2H), 2.55 – 2.51 (m, 2H), 2.29 – 2.20 (m, 1H), 2.00 – 1.92 (m, 1H).

^{13}C NMR (176 MHz, DMSO-*d*$_6$) δ 169.51, 153.43, 150.53, 136.50, 130.30, 129.63, 129.18, 125.62, 124.30, 124.09, 123.98, 120.78, 114.45, 109.48, 38.87, 33.71, 25.97, 16.60 ppm.

5.4 Recombinant Protein Expression and Purification

WDR5 (Δ1-21), WDR5-FLAG (Δ1-21), and WDR5^{F266A}-FLAG (Δ1-21) were expressed in BL21 DE RIL *Escherichia coli*. *E. coli* was grown in LB medium supplemented with 0.1% ampicillin and the expression was induced with 200 µM isopropyl-beta-D-thiogalactopyranoside (IPTG) at 20ºC for 16 h. After centrifugation, the pellets were resuspended in Buffer A additionally supplemented with 1 mM PMSF and a spatula tip of DNase I (Roche, #04716728001), then disrupted using a sonicator CL-334 (70% amplitude, 10 s pulse with 6 min interval) while the solution was kept cold by placing it on ice. The insoluble cell components were removed by centrifugation at 60000 rcf at 4ºC for 45 min. The supernatant containing the desired protein was filtered with a 0.22 µm syringe filter, loaded on HisTrap HP His tag protein purification column (Cytiva, #17524801) with Buffer A and eluted by a linear gradient to Buffer B (100%, 100 min). The desired fractions were collected, pooled, and dialysis with Prescission-3C-protease (3.5 mg/mL) at 4ºC for 16 h. The dialyzed protein was then purified by reverse Ni-NTA with a linear gradient of Buffer A to Buffer B in 50 min and concentrated with Amicon® ultra centrifugation filter (4000 rcf, 10 min, 4 °C). The concentrated protein was cleared by centrifugation (10 min, 4ºC, 20000 rcf), and the supernatant was loaded directly on a Superdex 75 16/600 column eluting with SEC buffer (isocratic, flow = 1 mL/min). For each purification step, the protein of interest was analyzed by 15% SDS-Page polyacrylamide gels and staining by Coomassie Blue staining solution.

5.5 Fluorescence Polarization (FP)

Fluorescence polarization was measured using a Tecan Spark Microplate Reader with 384-well plates (Corning, #4514). The FP Buffer A contained 0.1 M $K_nH_{(3-n)}PO_4$, 25 mM KCl, and 0.01% Triton-X at pH 6.5[87], while FP Buffer B contained 25 mM HEPES, 150 mM NaCl and 0.01% Triton-X at pH 7.0. Monochromator was applied for the fluorescence measurement and the excitation wavelength was set to 485 nm with a bandwidth of 20 nm, while the emission wavelength was set to 535 nm with a bandwidth of 20 nm.

5.5.1 FP measurement for direct binding assay

A 15-point dose-response curve was generated for direct binding assays by titrating a twofold dilution series of protein against a fixed concentration of FITC-labelled peptide (0.6 nM or 5 nM). A sample of protein-free corresponding FITC-labelled peptide in the assay buffer was included as non-binding control for each set of replicates. After the plate was prepared, it was sealed and incubated at 25ºC for 1 h before measurement. The raw values of fluorescence polarization were used, and K_D (dissociation constant) was calculated by fitting with GraphPad Prism 9 using the [Agonist] vs. Response –variable slope (four parameter) function for curve fitting.

5.5.2 FP measurement for competitive assay

Competitive fluorescence polarization measurements were performed by titrating a 2-fold dilution series of the inhibitor against a fixed concentration of WDR5 and tracer. The result was plotted using a logarithmic x-axis and normalized using the theoretical maximum and minimum binding values as 100 and 0% respectively. The raw value of fluorescence polarization was used, and the IC_{50} was calculated by fitting with log(Inhibitor) vs. normalized response –variable slope function for curve fitting using Graphpad Prism 9. The K_I was then calculated from the IC_{50} by the method reported by Nikolovska-Coleska et al.[89] using the following equation:

$$K_I = [I]_{50}/(\frac{[L]_{50}}{K_D} + \frac{[P]_0}{K_D} + 1)$$

In this equation, $[I]_{50}$ is the concentration of the unbound inhibitor at 50% inhibition, $[L]_{50}$ is the concentration of the unbound tracer at 50% inhibition, and $[P]_0$ refers to the concentration of the free protein without any inhibition. The calculations were made using the excel sheet provided by the authors which can be found at:

http://websites.umich.edu/~shaomengwanglab/software/calc_ki/index.html

5.6 Isothermal Titration Calorimetry

Isothermal titration calorimetry was measured using a MicroCal PEAQ-ITC, following the general instructions provided by the instrument. WDR5 was freshly dialyzed against the ITC buffer for 16 h, and the peptide was freshly prepared by dissolving in the corresponding dialysate. The concentration of WDR5/peptide was determined by OD_{280}/ OD_{495} on Nanodrop and diluted to the target concentration with dialysate. A total of 19 injections were made at 25ºC to record the titration curve. The thermodynamic parameters were determined using the PEAQ-ITC analysis software. Measurements were performed in duplicate.

5.7 Co-Crystallization and Structure Determination

5.7.1 Co-Crystallization of WDR5 with P26Ac

Purified WDR5 was concentrated to 23.25 mg/mL and ligand **P26Ac** dissolved in gel filtration buffer and a 1.5-fold molar excess of peptide was added. Crystallization was set up in MRC-3 drop plates by adding 100 nL protein/ligand complex to 100 nL reservoir solution and the plates were stored at 20°C. Crystals were obtained after a few days with reservoir conditions of 0.2 M Li_3-citrate and 20 % w/v PEG3350 and were cryo-conserved in reservoir solution supplemented with 20 % glycerol.

5.7.2 X-ray data collection and processing

Diffraction data was collected at beamline ID30A-3 at ESRF Grenoble (acquisition date 18.11.2022). Datasets were integrated using XIA2/DIALS (ccp4).[116,117] The structure was solved using Phaser (Phenix) and an AlphaFold model of apo-WDR5. [118,119]

5.7.3 Structure solution and refinement

WDR5-**P26Ac** co-crystallized in space group $P\ 4_3\ 2_1\ 2$ with dimensions 82.1737 x 82.1737 x 201.707 Å with two protein-peptide dimers in the asymmetric unit. The solved structure was refined by iterations of phenix.refine and manual model building in coot.[119,120] The structure was refined to a final Rfree of 24 % at 1.84 Å.

5.7.4 X-ray model analysis

The protein crystallized with two protein-peptide dimers in the asymmetric unit (Supplemental Figure 68) and the crystal showed a dense packing of monomers (shown by the symmetry related molecules in 20 Å distance; Supplemental Figure 69). We observed a big discrepancy in the quality of the $2F_O-F_C$ map between the two WDR5 chains probably caused by a higher disorder and flexibility in chain B, which corresponds to the increased B-factors of this chain (Supplemental Figure 68 B). Both chains share the same typical fold consisting of seven beta-propeller blades with four-stranded antiparallel sheets.[121] Overall, the two WDR5 chains still show a high similarity with an RMSD of 0.285 Å (Supplemental Figure 68 C). The biggest differences of both chains were observed in the *N*- and *C*-terminal regions, where we observed a higher ordered ß-sheet conformation for chain B, while the loops of chain A point away from each other, which could be caused by additional contacts to symmetry related molecules. The presence of ligand **P26Ac** was verified by a well resolved density of the central amino acids including the position of macrocyclization, but the density of the terminal Glu and Asp was less well resolved (Supplemental Figure 70 C/D). Further on, the ligands a and b bind in an identical manner to the reported native peptide of RBBP5 (Supplemental Figure 70 A/B). The binding of **P26Ac** to the WBM site does not lead to any significant changes of the conformation of this site compared to the reported apo-structure (PDB 2H14) (Supplemental Figure 70 E). Furthermore, we observed numerous structured waters in the structure, most of them were located in the central cavity of the protein, which was also observed in previous reports (Supplemental Figure 70 F)[122]. Furthermore, we recognized a potential disulfide bridge of chain A and chain B with a symmetry related molecule of chain B and A, respectively (Supplemental Figure 70 G).

5.8 Cancer cell lines

U-2 OS cells (DSMZ, ACC 785) were grown in DMEM high glucose medium (Sigma-Aldrich, #D6429) supplemented with 10% Fetal Bovine Serum (FBS, Cellsera, #AU-FBSPG) and 1% Penicillin-Streptomycin (Pen-Strep, Gibco, #15140-122) in a humidified atmosphere with 5% CO_2 under 37ºC. The subcultivation ratio suggested by ATCC was followed.

MDA-MB-231 cells (DSMZ, ACC 732) were grown in DMEM, high glucose, GlutaMAX supplement, pyruvate (Gibco, 31966-021) supplemented with 10% Fetal Bovine Serum (FBS, Cellsera, #AU-FBSPG) and 1% Penicillin-Streptomycin (Pen-Strep, Gibco, #15140-122) in a humidified atmosphere with 5% CO_2 under 37ºC. The subcultivation ratio suggested by ATCC was followed.

5.9 Total RNA isolation

For isolating total RNA, cells in a Petri dish were washed with PBS (Gibco, #10010-023), trypsinized (0.05% Trypsin, Gibco, #25300-054), followed by pelleting of the cells by centrifugation (3 min, rt, 200 rcf). The pellet was then washed with PBS, resuspended in TRIzol® (Invitrogen, #15596026), and the protocol from the manufacturer was followed to obtain total RNA isolates.

5.10 Quantitative reverse transcription-polymerase chain reaction (RT-qPCR)

In general, cDNA was synthesized by using high-capacity cDNA reverse transcription kit

(Applied Biosystems™, 4368814) following the manufacturer's protocol. For RNA from RNA-IP, all the RNA for each sample was directly put in the reaction with a reaction volume of 20 µL. On the other hand, the concentration of RNA from each sample originating from cell treatment was determined by nanodrop, then 0.5 µg of RNA was put in 20 µL of reaction to generate the cDNA. After cDNA was generated, it was analyzed by PowerUp™ SYBR™ Green Master Mix (Applied Bioscience™, #A25742) via the fast-cycling mode and supplemented with 500 nM of both forward and reverse primer according to the target amplicon.

5.10.1 Validation of qPCR efficiency

RNA was isolated as previously described in Section 5.9. 1.0 µg of RNA underwent reverse transcription in 20 µL RT reaction volume to afford the cDNA. To verify the qPCR efficiency, a titration curve of cDNA was generated by putting different amounts of cDNA into the well in 10 µL qPCR reaction volume. A straight line should be obtained when plotting the X as log(concentration) versus Y as Ct(observed), and the slope could be estimated by the function "linest" from excel. The qPCR efficiency could be calculated from the slope with the following equation[123]:

$$\text{Efficiency} = \left(10^{\left(-\frac{1}{slope}\right)} - 1\right) \times 100\%$$

During curve fitting, the $R^2 > 0.985$ was the minimum requirement and the data below this limit were abandoned for lacking reliability.

5.10.2 Validation of RT efficiency

RNA was isolated as previously described in 5.9. To verify the RT efficiency, a titration curve of RNA was generated by adding different amounts of RNA into 20 µL RT reaction volume to afford a titrated cDNA sample. Then 0.5 µL of cDNA samples was analyzed with qPCR in a 10 µL qPCR reaction setting. A straight line should be obtained when plotting the X as log(concentration) versus Y as Ct(observed), and the slope could be estimated by the function "linest" provided from excel. When the qPCR efficiency was between 90 – 110%, the RT efficiency could be calculated from the slope with the same equation as mentioned above.[123]

In most of the cases, the response of RT production was not linear when the amount of RNA input was ≥ 1.0 µg for 20 µL reaction. When this phenomenon occurred, those data points were not included in the curve fitting.

5.11 RNA immunoprecipitation (RNA-IP)

Total RNA isolates, purified proteins, and inhibitors were obtained as previously described. The protocol was adapted from Fuentes-Iglesias et al.[91] 10 µL of ANTI-FLAG® M2 Magnetic Beads (Millipore, #M8823) were washed with RIP Buffer C, then loaded with target protein (0.2 mg/mL, in Buffer C) supplemented with 0.1% IGEPAL, 0.1% BSA(Serva, #11930.03) and 10 ppm yeast tRNA (Biosciences, 058Y) at 4ºC for 3 h. After the incubation, the beads were washed with 0.1% IGEPAL in Buffer C five times. 1 µg of total RNA isolated from U-2 OS or MDA-MB-231, w/wo inhibitor, and w/wo 0.1% DMSO in RNA-immunoprecipitation buffer were incubated with protein loaded beads at rt for 1 h. After the incubation, the beads were washed with RNA-immunoprecipitation buffer 5 times and then rinsed with 50 µL of RNase free water once. The RNA/protein bound beads were resuspended in 50 µL of RNase free water. From this resuspended solution, 5 µL was transferred and mixed with 2X Laemmli buffer, loaded on a 15% SDS-PAGE polyacrylamide gel to analyze the protein loading of each sample. The other 45 µL was treated with 450 µL of TRIzol following the TRIzol protocol to extract the RNA. After the extraction process, the RNA pellet was dissolved in 10 µL of RNase free water and used for RT-qPCR analysis. A sample of 0.5 µg RNA input underwent the same treatment starting from resuspension in 50 µL RNase free water, including sampling for gel and TRIzol extraction, to control for sample loss between RIP product and non-treated RNA input.

5.11.1 Method for analyzing the percentage of enrichment after RNA-IP

Percentage of enrichment was calculated by normalization with 0.5 µg (50%) of RNA input. The percentage of enrichment was calculated by the method reported by Fuentes-Iglesias et al.[91] In short, the difference of Ct between the sample after enrichment and the 0.5 µg of RNA input (compensated with the percentage of dilution, 50%) was calculated.

$$Ct_{sample} - (Ct_{input} - \log_2(1/percentage\ of\ dilution)) = \Delta Ct_{corrected}$$

Then the percentage of target RNA enrichment from the RNA pool can be calculated by the following equation:

$$\text{Percentage of enrichment} = 2^{-\Delta Ct_{corrected}} \times 100\%$$

The student t-test was carried out by putting the average of mean (Mean) and the standard deviation (SD) with numbers of replicate (N) into GraphPad Prism 9 and calculated from multiple unpaired t tests with no assumption about consistent SDs (Welch's t-test) for each set of comparisons.

5.12 lncRNA expression level assay

5.12.1 Knockdown with siRNA

U-2 OS cells were seeded in a Biolite 6 well Multidish (Fisher Scientific, 11825275) for overnight (37ºC, 5% CO_2) in normal growth medium (DMEM + 10% FBS + 1% Pen-Strep). After the cells were adhered, knockdowns with Lincode Human *HOTTIP* siRNA smart pool (Dharmacon, R-188311-00-0005) or negative control Lincode Non-targeting Pool (Dharmacon, D-001320-10-05) were performed with Lipofectamine 2000 (ThermoFisher, 11668027) according to the manufacturer's instructions. After cells were incubated with transfecting medium for 4 h, the transfecting medium was changed into normal growth medium for another 24 h. Samples were harvested by RNA extraction (5.9) to afford the RNA, followed with RT-qPCR (5.10) then analyzed with the method described in 5.12.5 to obtain the fold change of the target gene.

5.12.2 Direct treatment with U-2 OS

U-2 OS cells were seeded in a Biolite 6 well Multidish (Fisher Scientific, 11825275) or Biolite 12 well Multidish (Fisher Scientific, 130185) for overnight (37ºC, 5% CO_2) in normal growth medium (DMEM + 10% FBS + 1% Pen-Strep). After the cells were settled, the selected compound concentration or blank (water or DMSO 0.1%) was prepared in fresh normal growth medium. Then the old growth medium was changed for the compound containing medium and incubated at 37ºC, 5% CO_2 for the desired incubation time.

Plate	Seeding density	Settling time	Incubation time
6 well	3.0×10^5	18 h	1 d
12 well	2.5×10^5	18 h	3 d
12 well	1.0×10^4	18 h	3 d[a] + 3 d

[a]: Refreshing the treatment and medium by exchanging everything fresh.

Samples were harvested by RNA extraction (5.9) to afford the RNA, followed with RT-qPCR (5.10) then analyzed with the method describe in 5.12.5 to obtain the fold change of the target gene.

5.12.3 Direct treatment with MDA-MB-231

MDA-MB-231 cells were seeded (1 x 10^5 cells/well for one-day treatment, or 5 x 10^4 cells/well for three-day treatment.) in a Biolite 12 well Multidish (Thermo Scientific, 130185) for one day (37ºC, 5% CO_2) in normal growth medium (DMEM-GlutaMAX + 10% FBS + 1% Pen-Strep). After the cells were settled, the selected compound or blank (water or DMSO 0.1%) was prepared in fresh normal growth medium. Then the old growth medium was changed into the treatment medium and incubated at 37ºC, 5% CO_2 for the desired incubation time.

Samples were harvested by RNA extraction (5.9) to afford the RNA, followed with RT-qPCR (5.10) then analyzed with the method described in 5.12.5 to obtain the fold change of the target gene.

5.12.4 Enhance Cellular Uptake with Additives

5.12.4.1 Direct cytosolic delivery by phase-separating peptides

The original protocol was described by Sun et al in 2022.[102] For verifying the effect of HB*pep*-SP delivery system, the experiment was started with seeding 3.0 x 10^4 of U-2 OS into a 8-well Ibidi dish (Ibidi, #80826) and incubated for 18 h (37ºC, 5% CO_2) in normal growth medium (DMEM + 10% FBS + 1% Pen-Strep). After the cells were settled, the medium was replaced with 180 µL of Opti-MEM and was ready for the delivery. The peptide delivering HB*pep*-SP resuspension was freshly prepared by adding HB*pep*-SP (10 mg/ml, 2 µL in 10 mM $AcOH_{(aq)}$) to the peptide (11 µM peptide, 18 µL Coacervate Formatting buffer.), then the resuspension was transferred to the cells and incubated for 4 h (37 °C, 5% CO_2). After the cargos were settled, the medium was removed and washed with PBS twice, then resupplied with 200 µL normal growth medium and the incubation was continued for another 20 h (37 °C, 5% CO_2). The free, unabsorbed coacervates were washed with pH 5.0 phosphate buffer, followed with PBS washing for five times, then fixation with ROTI® Histofix 4.5% (Roth, #2213.4) for 10 min (37 °C, 5% CO_2). After the fixation was completed, the fixation buffer was washed PBS for two times, then resupplied with 100 µL PBS to prevent the cells from drying out. The release of peptide in cytosol was observed by following the protocol 5.13.2.1.

For detecting lncRNA expression levels after treatment, the experiment was started with seeding 3.0 x 10^5 of U-2 OS into a 12-well dish and incubated for 18 h (37ºC, 5% CO_2) in normal growth medium (DMEM-high glucose + 10% FBS + 1% Pen-Strep). After the cells were settled, the medium was replaced with 900 µL of Opti-MEM and was ready for the delivery. The peptide delivering HB*pep*-SP suspension was freshly prepared by adding

HB*pep*-SP (10 mg/ml, 10 µL in 10 mM $AcOH_{(aq)}$) to the peptide (11 µM peptide, 90 µL Coacervate Formatting buffer.), then the suspension was transferred to the cells and incubated for 4 h (37 °C, 5% CO_2). After the particles were settled, the medium was removed and washed with PBS twice, then resupplied with 1 mL normal growth medium and the incubation was continued for another 20 h (37 °C, 5% CO_2). The free coacervates were washed away with pH 5.0 phosphate buffer, then the cells were harvested followed by RNA extraction (5.9) to afford the RNA. RT-qPCR (5.10) was used to measure lncRNA levels and analyzed with the method described in 5.12.5to obtain the fold change of the target genes.

5.12.4.2 Macropinocytosis

The original protocol was described by Arafiles et al in 2021.[99] In short, MDA-MB-231 cells were seeded (1 x 10^5 cells/well for one-day treatment, or 5 x 10^4 cells/well for three-day treatment.) in a Biolite 12 well Multidish (Thermo Scientific, 130185) for one day (37ºC, 5% CO_2) in normal growth medium (DMEM-GlutaMAX + 10% FBS + 1% Pen-Strep). After the cells were settled, the selected compound concentration or blank (water) with a constant concentration of LK15-oxP4A (5 µM) was prepared in Opti-MEM. The old growth medium was changed into the selected medium and incubated for 4 h (37ºC, 5% CO_2) to allow the internalization. After 4 h incubation, the treatment medium was washed with PBS for two times, then resupplied with normal growth medium and incubated for another 20 h (37ºC, 5% CO_2). Samples were harvested by RNA extraction (5.9) to afford the RNA, followed with RT-qPCR (5.10) then analyzed with the method described in 5.12.5 to obtain the fold change of the target gene.

5.12.4.3 Lipofecatime-2000

The protocol provided by the manufacturer was followed. In short, MDA-MB-231 cells were seeded (1 x 10^5 cells/well for one-day treatment, or 5 x 10^4 cells/well for three-day treatment.) in a Biolite 12 well Multidish (Thermo Scientific, 130185) for one day (37ºC, 5% CO_2) in normal growth medium (DMEM-GlutaMAX + 10% FBS + 1% Pen-Strep). The transfection solution was freshly prepared before the treatment by incubating the selected compound or blank (water) with lipofectamine-2000 (1 µL) in Opti-MEM (50 µL) for 5 min at rt. The old growth medium was refreshed with fresh normal growth medium (450 µL), then the transfection solution was added and incubated for 4 h (37ºC, 5% CO_2). After 4 h incubation, the treatment medium was washed with PBS twice, then resupplied with normal growth medium and incubated at 37ºC, 5% CO_2 for the desire treatment period. Samples were harvested by RNA extraction (5.9) to afford the RNA, followed with RT-qPCR (5.10) then analyzed with the method described in 5.12.5 to obtain the fold change of the target gene.

5.12.4.4 Electroporation

The Neon NxT electroporation system was used to perform the electroporation and the protocol provided by the manufacturer was followed. In short, MDA-MB-231 cells were seeded in eight 15 cm dishes with 50% confluency for one day (37ºC, 5% CO_2) in normal growth medium (DMEM-GlutaMAX + 10% FBS + 1% Pen-Strep). Cells were detached by trypsination, collected, and centrifuged at 200 rcf, rt for 5 min. The supernatant was discarded, and the cells were resuspended in PBS, and centrifuged at 200 rcf for 5 min. The PBS resuspension was performed twice more to remove residual trypsin from previous steps. The cells were resuspended in PBS and the cell density in the resuspension was measured by using a cell counting kit (Countess™ Cell Counting Chamber Slides). 3×10^6 cells were centrifuged at 200 rcf for 5 min, the supernatant removed, and the electroporation pipette was used to resuspend the cell pellet with 100 µL electroporation solution, which was prepared by diluting the stock solution of peptide with EP buffer in 1:9 ratio. The electroporation was carried out using a pulse of 1000 V, 35 msec twice. After the pulse, suspensions were diluted with 1 mL PBS, centrifuged at 200 rcf for 5 min, followed by removal of the supernatant and seeding into 6 a well plate (1×10^6 cells/well) with normal growth medium and incubated for 24 h. The old medium was removed and the cells were washed with PBS twice to remove the non-adherent cells, then either harvested directly or subcultivated into a 6-well plate (50% for another one day, 20% for another three day) with normal growth medium and incubated at 37ºC, 5% CO_2 until the desired treatment period. Samples were harvested by RNA extraction (5.9) to afford the RNA, followed with RT-qPCR (5.10) then analyzed with the method described in 5.12.5 to determine the fold change of the target genes.

To verify the effect of electroporation, a portion of samples were seeded in 8 well Ibidi plate (1×10^5 cells/well) and followed the protocol 5.13.2.2 to qualitatively measure peptide uptake.

5.12.5 Fold change of gene expression level

The raw Ct value obtained from the qPCR instrument was used in $2^{-\Delta\Delta Ct}$ method to calculate the fold change of target gene expression level and GAPDH was used as the internal reference gene[124]. In short, the expression level of the target gene was first normalized to GAPDH, then normalize to the expression level of the blank to give $\Delta\Delta Ct_{sample}^{gene}$.

$$Ct^{gene} - Ct^{GAPDH} = \Delta Ct^{gene}$$

$$\Delta Ct_{sample}^{gene} - \Delta Ct_{blank}^{gene} = \Delta\Delta Ct_{sample}^{gene}$$

Then the fold change can be calculated from the following equation:

$$\text{Fold change} = 2^{-\Delta\Delta Ct_{sample}^{gene}} \times 100\%$$

Each data point was represented by the mean of an individual biological replicate, and the standard deviation was calculated from the group of means by Graphpad Prism 9. The significance of the difference of each sample to the reference were calculated by Graphpad Prism 9 using the one sample student t-test and the target value was set to one (reference was normalized to one.).

5.13 Measurement of cellular uptake

The method for preparing the samples of interest was described in the previous section and the scale of experiment was sometime adapted to match the requirement of performing the following experiment.

5.13.1 Flow cytometry

U-2 OS cells were seeded in a 6 well Multidish (5 x 10^5 cells/well) for 18 h (37ºC, 5% CO_2) in normal growth medium (DMEM + 10% FBS + 1% Pen-Strep). After the cells were settled, the selected compound concentration or blank (water) was prepared in the fresh normal growth medium. Then the old growth medium was changed into the selected treatment medium and incubated for the desired incubation time (37ºC, 5% CO_2). Cells were harvested by trypsinization (0.05% Trypsin, 37ºC, 5% CO_2) for 3 min, then collected and centrifuged at 200 rcf, rt for 5 min. The supernatant was discarded, and the cells were resuspended with PBS, and centrifuged at 200 rcf for 5 min. The PBS resuspension was performed twice more to remove residual trypsin from previous steps. After the detachment, cells were resuspended in 300 µL of 1 µg/mL propidium iodide (PI) in PBS and kept on ice until the measurement. The cell suspension was supplemented with 100 µL of 0.4% trypan blue (Sigma Aldrich, T8154) and filtered through a cell strainer (Falcon, 352235). The flow cytometry measurement was performed with the Sony Cell Sorter (model LE-SH800SFP) using 405, 488 and 561 nm lasers coupled with 525/50 (FL2) and 600/60 (FL3) nm filters to detect FITC, and PI, respectively. FCM results were exported as flow cytometry standard files (FCS 3.0 or 3.1) by the cell sorter software (v. 2.1.3 or v. 2.1.5, Sony Biotechnology) and the student t-test was carried out by GraphPad prism 9 and calculated from multiple unpaired t tests with no assumption about consistent SDs (Welch t-test) for each set of comparisons.

5.13.2 Fluorescence confocal microscopy

5.13.2.1 Protocol for fixed cells

The day before the experiment, HEK293T cells were seeded at a concentration of 10,000 cells in 8-well μ-Slides (Ibidi). The cells were washed with 1x PBS and HKR solution the following day.[125] Next, FITC-labeled peptides were added in a final concentration of 10 μM and incubated for 1 hour. After 5 washes with HKR buffer, the cells were fixed with 4% formaldehyde (ROTHI®Histofix, ROTH) at 37 °C for 10 minutes, then washed with HKR buffer and stored in 1x PBS. Cells were imaged using a ZEISS LSM 800 at 40x magnification with oil immersion.

The image was analyzed by ImageJ Fiji to calculate the corrected total cell fluorescence (CTCF). In brief, the area of cells (A_{cell}) was defined in ESID channel, and the total fluorescence (I_{total})of the cell location was calculated in GFP channel. A random cell-free dark area was analyzed to determine the background intensity (I_{blank}). The CTCF can be calculated by the following formula[126]:

$$CTCF = I_{total} - A_{cell} \times I_{blank}$$

5.13.2.2 Protocol for living cells

Samples were seeded in 8 well Ibidi plate (1×10^5 cells/well) and incubated in normal growth medium for the desire incubation time (37ºC, 5% CO_2). The medium was exchanged into phenol-free DMEM, supplied with HEPES and the picture was taken by using confocal microscopy (3i Marianas Spinning Disk Confocal fluorescence microscope) at 40x magnification with oil immersion. The figures were exported by ImageJ Fiji.

5.14 Reference

(1) Campbell, N. A.; Reece, J. B. *Biology*; Pearson, Benjamin Cummings, 2005.

(2) Mattick, J. S.; Amaral, P. P.; Carninci, P.; Carpenter, S.; Chang, H. Y.; Chen, L.-L.; Chen, R.; Dean, C.; Dinger, M. E.; Fitzgerald, K. A.; Gingeras, T. R.; Guttman, M.; Hirose, T.; Huarte, M.; Johnson, R.; Kanduri, C.; Kapranov, P.; Lawrence, J. B.; Lee, J. T.; Mendell, J. T.; Mercer, T. R.; Moore, K. J.; Nakagawa, S.; Rinn, J. L.; Spector, D. L.; Ulitsky, I.; Wan, Y.; Wilusz, J. E.; Wu, M. Long Non-Coding RNAs: Definitions, Functions, Challenges and Recommendations. *Nat. Rev. Mol. Cell Biol.* **2023**, 1–17. https://doi.org/10.1038/s41580-022-00566-8.

(3) Hahn, M. W.; Wray, G. A. The G-Value Paradox. *Evol. Dev.* **2002**, *4* (2), 73–75. https://doi.org/10.1046/j.1525-142X.2002.01069.x.

(4) Moore, D. S. *The Developing Genome: An Introduction to Behavioral Epigenetics*; Oxford University Press: Oxford, New York, 2015.

(5) Waddington, C. H. The Epigenotype. *Int. J. Epidemiol.* **2012**, *41* (1), 10–13. https://doi.org/10.1093/ije/dyr184.

(6) Berger, S. L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. An Operational Definition of Epigenetics. *Genes Dev.* **2009**, *23* (7), 781–783. https://doi.org/10.1101/gad.1787609.

(7) Piovesan, A.; Pelleri, M. C.; Antonaros, F.; Strippoli, P.; Caracausi, M.; Vitale, L. On the Length, Weight and GC Content of the Human Genome. *BMC Res. Notes* **2019**, *12* (1), 106. https://doi.org/10.1186/s13104-019-4137-z.

(8) Gillooly, J. F.; Hein, A.; Damiani, R. Nuclear DNA Content Varies with Cell Size across Human Cell Types. *Cold Spring Harb. Perspect. Biol.* **2015**, *7* (7), a019091. https://doi.org/10.1101/cshperspect.a019091.

(9) Arents, G.; Burlingame, R. W.; Wang, B. C.; Love, W. E.; Moudrianakis, E. N. The Nucleosomal Core Histone Octamer at 3.1 A Resolution: A Tripartite Protein Assembly and a Left-Handed Superhelix. *Proc. Natl. Acad. Sci.* **1991**, *88* (22), 10148–10152. https://doi.org/10.1073/pnas.88.22.10148.

(10) Thomas, J. O.; Kornberg, R. D. An Octamer of Histones in Chromatin and Free in Solution. *Proc. Natl. Acad. Sci.* **1975**, *72* (7), 2626–2630. https://doi.org/10.1073/pnas.72.7.2626.

(11) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* **1997**, *389* (6648), 251–260. https://doi.org/10.1038/38444.

(12) Williams, S. K.; Tyler, J. K. Transcriptional Regulation by Chromatin Disassembly and Reassembly. *Curr. Opin. Genet. Dev.* **2007**, *17* (2), 88–93. https://doi.org/10.1016/j.gde.2007.02.001.

(13) Morrison, O.; Thakur, J. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. *Int. J. Mol. Sci.* **2021**, *22* (13), 6922. https://doi.org/10.3390/ijms22136922.

(14) Musselman, C. A.; Lalonde, M.-E.; Côté, J.; Kutateladze, T. G. Perceiving the Epigenetic Landscape through Histone Readers. *Nat. Struct. Mol. Biol.* **2012**, *19* (12), 1218–1227. https://doi.org/10.1038/nsmb.2436.

(15) Brander, S. M.; Biales, A. D.; Connon, R. E. The Role of Epigenomics in Aquatic Toxicology. *Environ. Toxicol. Chem.* **2017**, *36* (10), 2565–2573. https://doi.org/10.1002/etc.3930.

(16) Clapier, C. R.; Iwasa, J.; Cairns, B. R.; Peterson, C. L. Mechanisms of Action and Regulation of ATP-Dependent Chromatin-Remodelling Complexes. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (7), 407–422. https://doi.org/10.1038/nrm.2017.26.

(17) Santos-Rosa, H.; Schneider, R.; Bannister, A. J.; Sherriff, J.; Bernstein, B. E.; Emre, N. C. T.; Schreiber, S. L.; Mellor, J.; Kouzarides, T. Active Genes Are Tri-Methylated at K4 of Histone H3. *Nature* **2002**, *419* (6905), 407–411. https://doi.org/10.1038/nature01080.

(18) Dimopoulos, K.; Gimsing, P.; Grønbæk, K. The Role of Epigenetics in the Biology of Multiple Myeloma. *Blood Cancer J.* **2014**, *4* (5), e207–e207. https://doi.org/10.1038/bcj.2014.29.

(19) Saksouk, N.; Simboeck, E.; Déjardin, J. Constitutive Heterochromatin Formation and Transcription in Mammals. *Epigenetics Chromatin* **2015**, *8* (1), 3. https://doi.org/10.1186/1756-8935-8-3.

(20) Breiling, A.; Lyko, F. Epigenetic Regulatory Functions of DNA Modifications: 5-Methylcytosine and Beyond. *Epigenetics Chromatin* **2015**, *8* (1), 24. https://doi.org/10.1186/s13072-015-0016-6.

(21) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*.

(22) Barlow, D. P. Gametic Imprinting in Mammals. *Science* **1995**, *270* (5242), 1610–1613. https://doi.org/10.1126/science.270.5242.1610.

(23) Yen, P. H.; Patel, P.; Chinault, A. C.; Mohandas, T.; Shapiro, L. J. Differential Methylation of Hypoxanthine Phosphoribosyltransferase Genes on Active and Inactive Human X Chromosomes. *Proc. Natl. Acad. Sci.* **1984**, *81* (6), 1759–1763. https://doi.org/10.1073/pnas.81.6.1759.

(24) Choi, W.-Y.; Hwang, J.-H.; Cho, A.-N.; Lee, A. J.; Lee, J.; Jung, I.; Cho, S.-W.; Kim, L. K.; Kim, Y.-J. DNA Methylation of Intragenic CpG Islands Are Required for Differentiation from iPSC to NPC. *Stem Cell Rev. Rep.* **2020**, *16* (6), 1316–1327. https://doi.org/10.1007/s12015-020-10041-6.

(25) Biswas, S.; Rao, C. M. Epigenetic Tools (The Writers, The Readers and The Erasers) and Their Implications in Cancer Therapy. *Eur. J. Pharmacol.* **2018**, *837*, 8–24. https://doi.org/10.1016/j.ejphar.2018.08.021.

(26) Gyi, J. I.; Lane, A. N.; Conn, G. L.; Brown, T. The Orientation and Dynamics of the C2′-OH and Hydration of RNA and DNA·RNA Hybrids. *Nucleic Acids Res.* **1998**, *26* (13), 3104–3110. https://doi.org/10.1093/nar/26.13.3104.

(27) Rich, A.; Davies, D. R. A NEW TWO STRANDED HELICAL STRUCTURE: POLYADENYLIC ACID AND POLYURIDYLIC ACID. *J. Am. Chem. Soc.* **1956**, *78* (14), 3548–3549.

https://doi.org/10.1021/ja01595a086.

(28) Reymond, C.; Beaudoin, J.-D.; Perreault, J.-P. Modulating RNA Structure and Catalysis: Lessons from Small Cleaving Ribozymes. *Cell. Mol. Life Sci.* **2009**, *66* (24), 3937–3950. https://doi.org/10.1007/s00018-009-0124-1.

(29) Lam, J. K. W.; Chow, M. Y. T.; Zhang, Y.; Leung, S. W. S. siRNA Versus miRNA as Therapeutics for Gene Silencing. *Mol. Ther. - Nucleic Acids* **2015**, *4*. https://doi.org/10.1038/mtna.2015.23.

(30) Grummt, I. Regulation of Mammalian Ribosomal Gene Transcription by RNA Polymerase I. In *Progress in Nucleic Acid Research and Molecular Biology*; Moldave, K., Ed.; Academic Press, 1998; Vol. 62, pp 109–154. https://doi.org/10.1016/S0079-6603(08)60506-1.

(31) Willis, I. M. RNA Polymerase III. *Eur. J. Biochem.* **1993**, *212* (1), 1–11. https://doi.org/10.1111/j.1432-1033.1993.tb17626.x.

(32) Lee, Y.; Kim, M.; Han, J.; Yeom, K.-H.; Lee, S.; Baek, S. H.; Kim, V. N. MicroRNA Genes Are Transcribed by RNA Polymerase II. *EMBO J.* **2004**, *23* (20), 4051–4060. https://doi.org/10.1038/sj.emboj.7600385.

(33) Rodriguez, A.; Griffiths-Jones, S.; Ashurst, J. L.; Bradley, A. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Res.* **2004**, *14* (10a), 1902–1910. https://doi.org/10.1101/gr.2722704.

(34) Rana, T. M. Illuminating the Silence: Understanding the Structure and Function of Small RNAs. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (1), 23–36. https://doi.org/10.1038/nrm2085.

(35) Carthew, R. W.; Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **2009**, *136* (4), 642–655. https://doi.org/10.1016/j.cell.2009.01.035.

(36) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. An RNA-Directed Nuclease Mediates Post-Transcriptional Gene Silencing in Drosophila Cells. *Nature* **2000**, *404* (6775), 293–296. https://doi.org/10.1038/35005107.

(37) Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M. C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; Kodzius, R.; Shimokawa, K.; Bajic, V. B.; Brenner, S. E.; Batalov, S.; Forrest, A. R. R.; Zavolan, M.; Davis, M. J.; Wilming, L. G.; Aidinis, V.; Allen, J. E.; Ambesi-Impiombato, A.; Apweiler, R.; Aturaliya, R. N.; Bailey, T. L.; Bansal, M.; Baxter, L.; Beisel, K. W.; Bersano, T.; Bono, H.; Chalk, A. M.; Chiu, K. P.; Choudhary, V.; Christoffels, A.; Clutterbuck, D. R.; Crowe, M. L.; Dalla, E.; Dalrymple, B. P.; de Bono, B.; Gatta, G. D.; di Bernardo, D.; Down, T.; Engstrom, P.; Fagiolini, M.; Faulkner, G.; Fletcher, C. F.; Fukushima, T.; Furuno, M.; Futaki, S.; Gariboldi, M.; Georgii-Hemming, P.; Gingeras, T. R.; Gojobori, T.; Green, R. E.; Gustincich, S.; Harbers, M.; Hayashi, Y.; Hensch, T. K.; Hirokawa, N.; Hill, D.; Huminiecki, L.; Iacono, M.; Ikeo, K.; Iwama, A.; Ishikawa, T.; Jakt, M.; Kanapin, A.; Katoh, M.; Kawasawa, Y.; Kelso, J.; Kitamura, H.; Kitano, H.; Kollias, G.; Krishnan, S. P. T.; Kruger, A.; Kummerfeld, S. K.; Kurochkin, I. V.; Lareau, L. F.; Lazarevic, D.; Lipovich, L.; Liu, J.; Liuni, S.; McWilliam, S.; Babu, M. M.; Madera, M.; Marchionni, L.; Matsuda, H.; Matsuzawa, S.; Miki, H.; Mignone, F.; Miyake, S.; Morris, K.; Mottagui-Tabar, S.; Mulder, N.; Nakano, N.;

Nakauchi, H.; Ng, P.; Nilsson, R.; Nishiguchi, S.; Nishikawa, S.; Nori, F.; Ohara, O.; Okazaki, Y.; Orlando, V.; Pang, K. C.; Pavan, W. J.; Pavesi, G.; Pesole, G.; Petrovsky, N.; Piazza, S.; Reed, J.; Reid, J. F.; Ring, B. Z.; Ringwald, M.; Rost, B.; Ruan, Y.; Salzberg, S. L.; Sandelin, A.; Schneider, C.; Schönbach, C.; Sekiguchi, K.; Semple, C. A. M.; Seno, S.; Sessa, L.; Sheng, Y.; Shibata, Y.; Shimada, H.; Shimada, K.; Silva, D.; Sinclair, B.; Sperling, S.; Stupka, E.; Sugiura, K.; Sultana, R.; Takenaka, Y.; Taki, K.; Tammoja, K.; Tan, S. L.; Tang, S.; Taylor, M. S.; Tegner, J.; Teichmann, S. A.; Ueda, H. R.; van Nimwegen, E.; Verardo, R.; Wei, C. L.; Yagi, K.; Yamanishi, H.; Zabarovsky, E.; Zhu, S.; Zimmer, A.; Hide, W.; Bult, C.; Grimmond, S. M.; Teasdale, R. D.; Liu, E. T.; Brusic, V.; Quackenbush, J.; Wahlestedt, C.; Mattick, J. S.; Hume, D. A.; Kai, C.; Sasaki, D.; Tomaru, Y.; Fukuda, S.; Kanamori-Katayama, M.; Suzuki, M.; Aoki, J.; Arakawa, T.; Iida, J.; Imamura, K.; Itoh, M.; Kato, T.; Kawaji, H.; Kawagashira, N.; Kawashima, T.; Kojima, M.; Kondo, S.; Konno, H.; Nakano, K.; Ninomiya, N.; Nishio, T.; Okada, M.; Plessy, C.; Shibata, K.; Shiraki, T.; Suzuki, S.; Tagami, M.; Waki, K.; Watahiki, A.; Okamura-Oho, Y.; Suzuki, H.; Kawai, J.; Hayashizaki, Y. The Transcriptional Landscape of the Mammalian Genome. *Science* **2005**, *309* (5740), 1559–1563. https://doi.org/10.1126/science.1112014.

(38) Kapranov, P.; Cawley, S. E.; Drenkow, J.; Bekiranov, S.; Strausberg, R. L.; Fodor, S. P. A.; Gingeras, T. R. Large-Scale Transcriptional Activity in Chromosomes 21 and 22. *Science* **2002**, *296* (5569), 916–919. https://doi.org/10.1126/science.1068597.

(39) Flynn, R. A.; Chang, H. Y. Long Noncoding RNAs in Cell-Fate Programming and Reprogramming. *Cell Stem Cell* **2014**, *14* (6), 752–761. https://doi.org/10.1016/j.stem.2014.05.014.

(40) Cabili, M. N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J. L. Integrative Annotation of Human Large Intergenic Noncoding RNAs Reveals Global Properties and Specific Subclasses. *Genes Dev.* **2011**, *25* (18), 1915–1927. https://doi.org/10.1101/gad.17446611.

(41) Chen, J.; Wang, Y.; Wang, C.; Hu, J.-F.; Li, W. LncRNA Functions as a New Emerging Epigenetic Factor in Determining the Fate of Stem Cells. *Front. Genet.* **2020**, *11*.

(42) Sweta, S.; Dudnakova, T.; Sudheer, S.; Baker, A. H.; Bhushan, R. Importance of Long Non-Coding RNAs in the Development and Disease of Skeletal Muscle and Cardiovascular Lineages. *Front. Cell Dev. Biol.* **2019**, *7*.

(43) Brown, S. D. M. XIST and the Mapping of the X Chromosome Inactivation Centre. *BioEssays* **1991**, *13* (11), 607–612. https://doi.org/10.1002/bies.950131112.

(44) Loda, A.; Heard, E. Xist RNA in Action: Past, Present, and Future. *PLOS Genet.* **2019**, *15* (9), e1008333. https://doi.org/10.1371/journal.pgen.1008333.

(45) Marahrens, Y.; Panning, B.; Dausman, J.; Strauss, W.; Jaenisch, R. Xist-Deficient Mice Are Defective in Dosage Compensation but Not Spermatogenesis. *Genes Dev.* **1997**, *11* (2), 156–166. https://doi.org/10.1101/gad.11.2.156.

(46) Penny, G. D.; Kay, G. F.; Sheardown, S. A.; Rastan, S.; Brockdorff, N. Requirement for Xist in X Chromosome Inactivation. *Nature* **1996**, *379* (6561), 131–137. https://doi.org/10.1038/379131a0.

(47) Wutz, A. Xist Function: Bridging Chromatin and Stem Cells. *Trends Genet.* **2007**, *23* (9), 457–464. https://doi.org/10.1016/j.tig.2007.07.004.

(48) Morlando, M.; Fatica, A. Alteration of Epigenetic Regulation by Long Noncoding RNAs in Cancer. *Int. J. Mol. Sci.* **2018**, *19* (2), 570. https://doi.org/10.3390/ijms19020570.

(49) Ule, J.; Jensen, K. B.; Ruggiu, M.; Mele, A.; Ule, A.; Darnell, R. B. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **2003**, *302* (5648), 1212–1215. https://doi.org/10.1126/science.1090095.

(50) Li, X.; Song, J.; Yi, C. Genome-Wide Mapping of Cellular Protein–RNA Interactions Enabled by Chemical Crosslinking. *Genomics Proteomics Bioinformatics* **2014**, *12* (2), 72–78. https://doi.org/10.1016/j.gpb.2014.03.001.

(51) Ramanathan, M.; Porter, D. F.; Khavari, P. A. Methods to Study RNA–Protein Interactions. *Nat. Methods* **2019**, *16* (3), 225–234. https://doi.org/10.1038/s41592-019-0330-1.

(52) Yang, Y. W. Rna-Mediated Programming of Active Chromatin. PhD dissertation, STANFORD UNIVERSITY, 2012.

(53) Minuesa, G.; Albanese, S. K.; Xie, W.; Kazansky, Y.; Worroll, D.; Chow, A.; Schurer, A.; Park, S.-M.; Rotsides, C. Z.; Taggart, J.; Rizzi, A.; Naden, L. N.; Chou, T.; Gourkanti, S.; Cappel, D.; Passarelli, M. C.; Fairchild, L.; Adura, C.; Glickman, J. F.; Schulman, J.; Famulare, C.; Patel, M.; Eibl, J. K.; Ross, G. M.; Bhattacharya, S.; Tan, D. S.; Leslie, C. S.; Beuming, T.; Patel, D. J.; Goldgur, Y.; Chodera, J. D.; Kharas, M. G. Small-Molecule Targeting of MUSASHI RNA-Binding Activity in Acute Myeloid Leukemia. *Nat. Commun.* **2019**, *10* (1), 2691. https://doi.org/10.1038/s41467-019-10523-3.

(54) Wang, K. C.; Yang, Y. W.; Liu, B.; Sanyal, A.; Corces-Zimmerman, R.; Chen, Y.; Lajoie, B. R.; Protacio, A.; Flynn, R. A.; Gupta, R. A.; Wysocka, J.; Lei, M.; Dekker, J.; Helms, J. A.; Chang, H. Y. A Long Noncoding RNA Maintains Active Chromatin to Coordinate Homeotic Gene Expression. *Nature* **2011**, *472* (7341), 120–124. https://doi.org/10.1038/nature09819.

(55) Wang, S. S. W.; Wuputra, K.; Liu, C.-J.; Lin, Y.-C.; Chen, Y.-T.; Chai, C.-Y.; Lin, C.-L. S.; Kuo, K.-K.; Tsai, M.-H.; Wang, S.-W.; Chen, K.-K.; Miyoshi, H.; Nakamura, Y.; Saito, S.; Hanafusa, T.; Wu, D.-C.; Lin, C.-S.; Yokoyama, K. K. Oncogenic Function of the Homeobox A13-Long Noncoding RNA HOTTIP-Insulin Growth Factor-Binding Protein 3 Axis in Human Gastric Cancer. *Oncotarget* **2016**, *7* (24), 36049–36064. https://doi.org/10.18632/oncotarget.9102.

(56) Lin, C.; Wang, Y.; Wang, Y.; Zhang, S.; Yu, L.; Guo, C.; Xu, H. Transcriptional and Posttranscriptional Regulation of HOXA13 by lncRNA HOTTIP Facilitates Tumorigenesis and Metastasis in Esophageal Squamous Carcinoma Cells. *Oncogene* **2017**, *36* (38), 5392–5406. https://doi.org/10.1038/onc.2017.133.

(57) Guarnaccia, A. duPuy; Tansey, W. P. Moonlighting with WDR5: A Cellular Multitasker. *J. Clin. Med.* **2018**, *7* (2), 21. https://doi.org/10.3390/jcm7020021.

(58) Chen, X.; Xu, J.; Wang, X.; Long, G.; You, Q.; Guo, X. Targeting WD Repeat-Containing Protein 5 (WDR5):

A Medicinal Chemistry Perspective. *J. Med. Chem.* **2021**, *64* (15), 10537–10556. https://doi.org/10.1021/acs.jmedchem.1c00037.

(59) Dias, J.; Van Nguyen, N.; Georgiev, P.; Gaub, A.; Brettschneider, J.; Cusack, S.; Kadlec, J.; Akhtar, A. Structural Analysis of the KANSL1/WDR5/KANSL2 Complex Reveals That WDR5 Is Required for Efficient Assembly and Chromatin Targeting of the NSL Complex. *Genes Dev.* **2014**, *28* (9), 929–942. https://doi.org/10.1101/gad.240200.114.

(60) Thomas, L. R.; Wang, Q.; Grieb, B. C.; Phan, J.; Foshage, A. M.; Sun, Q.; Olejniczak, E. T.; Clark, T.; Dey, S.; Lorey, S.; Alicie, B.; Howard, G. C.; Cawthon, B.; Ess, K. C.; Eischen, C. M.; Zhao, Z.; Fesik, S. W.; Tansey, W. P. Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Mol. Cell* **2015**, *58* (3), 440–452. https://doi.org/10.1016/j.molcel.2015.02.028.

(61) Ee, L.-S.; McCannell, K. N.; Tang, Y.; Fernandes, N.; Hardy, W. R.; Green, M. R.; Chu, F.; Fazzio, T. G. An Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. *Stem Cell Rep.* **2017**, *8* (6), 1488–1496. https://doi.org/10.1016/j.stemcr.2017.04.020.

(62) Couture, J.-F.; Collazo, E.; Trievel, R. C. Molecular Recognition of Histone H3 by the WD40 Protein WDR5. *Nat. Struct. Mol. Biol.* **2006**, *13* (8), 698–703. https://doi.org/10.1038/nsmb1116.

(63) Migliori, V.; Müller, J.; Phalke, S.; Low, D.; Bezzi, M.; Mok, W. C.; Sahu, S. K.; Gunaratne, J.; Capasso, P.; Bassi, C.; Cecatiello, V.; De Marco, A.; Blackstock, W.; Kuznetsov, V.; Amati, B.; Mapelli, M.; Guccione, E. Symmetric Dimethylation of H3R2 Is a Newly Identified Histone Mark That Supports Euchromatin Maintenance. *Nat. Struct. Mol. Biol.* **2012**, *19* (2), 136–144. https://doi.org/10.1038/nsmb.2209.

(64) Chang, J.-Y.; Neugebauer, C.; Schmeing, S.; Amrahova, G.; 't Hart, P. Macrocyclic Peptides as Inhibitors of WDR5 - lncRNA Interactions. *Chem. Commun.* **2023**. https://doi.org/10.1039/D3CC03221C.

(65) Yang, Y. W.; Flynn, R. A.; Chen, Y.; Qu, K.; Wan, B.; Wang, K. C.; Lei, M.; Chang, H. Y. Essential Role of lncRNA Binding for WDR5 Maintenance of Active Chromatin and Embryonic Stem Cell Pluripotency. *eLife* **2014**, *3*, e02046. https://doi.org/10.7554/eLife.02046.

(66) Yuan, X.; Sun, Z.; Cui, C. Knockdown of lncRNA HOTTIP Inhibits Retinoblastoma Progression by Modulating the miR-101-3p/STC1 Axis. *Technol. Cancer Res. Treat.* **2021**, *20*, 1533033821997831. https://doi.org/10.1177/1533033821997831.

(67) Luo, H.; Zhu, G.; Xu, J.; Lai, Q.; Yan, B.; Guo, Y.; Fung, T. K.; Zeisig, B. B.; Cui, Y.; Zha, J.; Cogle, C.; Wang, F.; Xu, B.; Yang, F.-C.; Li, W.; So, C. W. E.; Qiu, Y.; Xu, M.; Huang, S. HOTTIP lncRNA Promotes Hematopoietic Stem Cell Self-Renewal Leading to AML-like Disease in Mice. *Cancer Cell* **2019**, *36* (6), 645-659.e8. https://doi.org/10.1016/j.ccell.2019.10.011.

(68) Han, S.; Jin, X.; Liu, Z.; Xing, F.; Han, Y.; Yu, X.; He, G.; Qiu, F. The Long Noncoding RNA HOTTIP Promotes Breast Cancer Cell Migration, Invasiveness, and Epithelial–Mesenchymal Transition via the Wnt–β-Catenin Signaling Pathway. *Biochem. Cell Biol.* **2019**, *97* (5), 655–664. https://doi.org/10.1139/bcb-2018-0313.

(69) Tang, Y.; Ji, F. lncRNA HOTTIP Facilitates Osteosarcoma Cell Migration, Invasion and Epithelial-mesenchymal Transition by Forming a Positive Feedback Loop with c-Myc. *Oncol. Lett.* **2019**, *18* (2), 1649–1656. https://doi.org/10.3892/ol.2019.10463.

(70) Yao, X.-Y.; Liu, J.-F.; Luo, Y.; Xu, X.-Z.; Bu, J. LncRNA HOTTIP Facilitates Cell Proliferation, Invasion, and Migration in Osteosarcoma by Interaction with PTBP1 to Promote KHSRP Level. *Cell Cycle* **2021**, *20* (3), 283–297. https://doi.org/10.1080/15384101.2020.1870820.

(71) Hu, B.; Zhong, L.; Weng, Y.; Peng, L.; Huang, Y.; Zhao, Y.; Liang, X.-J. Therapeutic siRNA: State of the Art. *Signal Transduct. Target. Ther.* **2020**, *5* (1), 1–25. https://doi.org/10.1038/s41392-020-0207-x.

(72) Avdic, V.; Zhang, P.; Lanouette, S.; Groulx, A.; Tremblay, V.; Brunzelle, J.; Couture, J.-F. Structural and Biochemical Insights into MLL1 Core Complex Assembly. *Structure* **2011**, *19* (1), 101–108. https://doi.org/10.1016/j.str.2010.09.022.

(73) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. https://doi.org/10.1021/ja00897a025.

(74) Vinogradov, A. A.; Yin, Y.; Suga, H. Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J. Am. Chem. Soc.* **2019**, *141* (10), 4167–4181. https://doi.org/10.1021/jacs.8b13178.

(75) Byk, G.; Halle, D.; Zeltser, I.; Bitan, G.; Selinger, Z.; Gilon, C. Synthesis and Biological Activity of NK-1 Selective, N-Backbone Cyclic Analogs of the C-Terminal Hexapeptide of Substance P. *J. Med. Chem.* **1996**, *39* (16), 3174–3178. https://doi.org/10.1021/jm960154i.

(76) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Signal Transduct. Target. Ther.* **2022**, *7* (1), 1–27. https://doi.org/10.1038/s41392-022-00904-4.

(77) Al Musaimi, O.; Al Shaer, D.; Albericio, F.; de la Torre, B. G. 2022 FDA TIDES (Peptides and Oligonucleotides) Harvest. *Pharmaceuticals* **2023**, *16* (3), 336. https://doi.org/10.3390/ph16030336.

(78) Raitano, L.; Masoni, D. Novo Nordisk Briefly Overtakes LVMH as Europe’s Most Valuable Company. *Reuters*. September 1, 2023. https://www.reuters.com/markets/europe/drugmaker-novo-nordisk-briefly-overtakes-lvmh-europes-most-valuable-company-2023-09-01/ (accessed 2023-09-24).

(79) Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent Advances in the Development of Protein–Protein Interactions Modulators: Mechanisms and Clinical Trials. *Signal Transduct. Target. Ther.* **2020**, *5* (1), 1–23. https://doi.org/10.1038/s41392-020-00315-3.

(80) Smith, M. C.; Gestwicki, J. E. Features of Protein–Protein Interactions That Translate into Potent Inhibitors: Topology, Surface Area and Affinity. *Expert Rev. Mol. Med.* **2012**, *14*, e16. https://doi.org/10.1017/erm.2012.10.

(81) Teuscher, K. B.; Chowdhury, S.; Meyers, K. M.; Tian, J.; Sai, J.; Van Meveren, M.; South, T. M.; Sensintaffar, J. L.; Rietz, T. A.; Goswami, S.; Wang, J.; Grieb, B. C.; Lorey, S. L.; Howard, G. C.; Liu, Q.; Moore,

W. J.; Stott, G. M.; Tansey, W. P.; Lee, T.; Fesik, S. W. Structure-Based Discovery of Potent WD Repeat Domain 5 Inhibitors That Demonstrate Efficacy and Safety in Preclinical Animal Models. *Proc. Natl. Acad. Sci.* **2023**, *120* (1), e2211297120. https://doi.org/10.1073/pnas.2211297120.

(82) Geppert, T.; Hoy, B.; Wessler, S.; Schneider, G. Context-Based Identification of Protein-Protein Interfaces and "Hot-Spot" Residues. *Chem. Biol.* **2011**, *18* (3), 344–353. https://doi.org/10.1016/j.chembiol.2011.01.005.

(83) Chang, S.; Liu, J.; Guo, S.; He, S.; Qiu, G.; Lu, J.; Wang, J.; Fan, L.; Zhao, W.; Che, X. HOTTIP and HOXA13 Are Oncogenes Associated with Gastric Cancer Progression. *Oncol. Rep.* **2016**, *35* (6), 3577–3585. https://doi.org/10.3892/or.2016.4743.

(84) Hendrickson, O. D.; Taranova, N. A.; Zherdev, A. V.; Dzantiev, B. B.; Eremin, S. A. Fluorescence Polarization-Based Bioassays: New Horizons. *Sensors* **2020**, *20* (24), 7132. https://doi.org/10.3390/s20247132.

(85) Ladbury, J. E. Application of Isothermal Titration Calorimetry in the Biological Sciences: Things Are Heating Up! *BioTechniques* **2004**, *37* (6), 885–887. https://doi.org/10.2144/04376TE01.

(86) Prozeller, D.; Morsbach, S.; Landfester, K. Isothermal Titration Calorimetry as a Complementary Method for Investigating Nanoparticle–Protein Interactions. *Nanoscale* **2019**, *11* (41), 19265–19273. https://doi.org/10.1039/C9NR05790K.

(87) Karatas, H.; Townsend, E. C.; Bernard, D.; Dou, Y.; Wang, S. Analysis of the Binding of Mixed Lineage Leukemia 1 (MLL1) and Histone 3 Peptides to WD Repeat Domain 5 (WDR5) for the Design of Inhibitors of the MLL1−WDR5 Interaction. *J. Med. Chem.* **2010**, *53* (14), 5179–5185. https://doi.org/10.1021/jm100139b.

(88) Chatterjee, J.; Laufer, B.; Kessler, H. Synthesis of N-Methylated Cyclic Peptides. *Nat. Protoc.* **2012**, *7* (3), 432–444. https://doi.org/10.1038/nprot.2011.450.

(89) Nikolovska-Coleska, Z.; Wang, R.; Fang, X.; Pan, H.; Tomita, Y.; Li, P.; Roller, P. P.; Krajewski, K.; Saito, N. G.; Stuckey, J. A.; Wang, S. Development and Optimization of a Binding Assay for the XIAP BIR3 Domain Using Fluorescence Polarization. *Anal. Biochem.* **2004**, *332* (2), 261–273. https://doi.org/10.1016/j.ab.2004.05.055.

(90) Slyusareva, E. A.; Gerasimova, M. A. pH-Dependence of the Absorption and Fluorescent Properties of Fluorone Dyes in Aqueous Solutions. *Russ. Phys. J.* **2014**, *56* (12), 1370–1377. https://doi.org/10.1007/s11182-014-0188-8.

(91) Fuentes-Iglesias, A.; Garcia-Outeiral, V.; Pardavila, J. A.; Wang, J.; Fidalgo, M.; Guallar, D. An Optimized Immunoprecipitation Protocol for Assessing Protein-RNA Interactions In Vitro. *STAR Protoc.* **2020**, *1* (2), 100093. https://doi.org/10.1016/j.xpro.2020.100093.

(92) Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* **2009**, *55* (4), 611–622. https://doi.org/10.1373/clinchem.2008.112797.

(93) Rogers-Broadway, K.-R.; Karteris, E. Amplification Efficiency and Thermal Stability of qPCR Instrumentation: Current Landscape and Future Perspectives. *Exp. Ther. Med.* **2015**, *10* (4), 1261–1264. https://doi.org/10.3892/etm.2015.2712.

(94) Li, Z.; Zhao, X.; Zhou, Y.; Liu, Y.; Zhou, Q.; Ye, H.; Wang, Y.; Zeng, J.; Song, Y.; Gao, W.; Zheng, S.; Zhuang, B.; Chen, H.; Li, W.; Li, H.; Li, H.; Fu, Z.; Chen, R. The Long Non-Coding RNA HOTTIP Promotes Progression and Gemcitabine Resistance by Regulating HOXA13 in Pancreatic Cancer. *J. Transl. Med.* **2015**, *13* (1), 84. https://doi.org/10.1186/s12967-015-0442-z.

(95) Grebien, F.; Vedadi, M.; Getlik, M.; Giambruno, R.; Grover, A.; Avellino, R.; Skucha, A.; Vittori, S.; Kuznetsova, E.; Smil, D.; Barsyte-Lovejoy, D.; Li, F.; Poda, G.; Schapira, M.; Wu, H.; Dong, A.; Senisterra, G.; Stukalov, A.; Huber, K. V. M.; Schönegger, A.; Marcellus, R.; Bilban, M.; Bock, C.; Brown, P. J.; Zuber, J.; Bennett, K. L.; Al-awar, R.; Delwel, R.; Nerlov, C.; Arrowsmith, C. H.; Superti-Furga, G. Pharmacological Targeting of the Wdr5-MLL Interaction in C/EBPα N-Terminal Leukemia. *Nat. Chem. Biol.* **2015**, *11* (8), 571–578. https://doi.org/10.1038/nchembio.1859.

(96) Macdonald, J. D.; Chacón Simon, S.; Han, C.; Wang, F.; Shaw, J. G.; Howes, J. E.; Sai, J.; Yuh, J. P.; Camper, D.; Alicie, B. M.; Alvarado, J.; Nikhar, S.; Payne, W.; Aho, E. R.; Bauer, J. A.; Zhao, B.; Phan, J.; Thomas, L. R.; Rossanese, O. W.; Tansey, W. P.; Waterson, A. G.; Stauffer, S. R.; Fesik, S. W. Discovery and Optimization of Salicylic Acid-Derived Sulfonamide Inhibitors of the WD Repeat-Containing Protein 5–MYC Protein–Protein Interaction. *J. Med. Chem.* **2019**, *62* (24), 11232–11259. https://doi.org/10.1021/acs.jmedchem.9b01411.

(97) Getlik, M.; Smil, D.; Zepeda-Velázquez, C.; Bolshan, Y.; Poda, G.; Wu, H.; Dong, A.; Kuznetsova, E.; Marcellus, R.; Senisterra, G.; Dombrovski, L.; Hajian, T.; Kiyota, T.; Schapira, M.; Arrowsmith, C. H.; Brown, P. J.; Vedadi, M.; Al-Awar, R. Structure-Based Optimization of a Small Molecule Antagonist of the Interaction Between WD Repeat-Containing Protein 5 (WDR5) and Mixed-Lineage Leukemia 1 (MLL1). *J. Med. Chem.* **2016**, *59* (6), 2478–2496. https://doi.org/10.1021/acs.jmedchem.5b01630.

(98) Aho, E. R.; Wang, J.; Gogliotti, R. D.; Howard, G. C.; Phan, J.; Acharya, P.; Macdonald, J. D.; Cheng, K.; Lorey, S. L.; Lu, B.; Wenzel, S.; Foshage, A. M.; Alvarado, J.; Wang, F.; Shaw, J. G.; Zhao, B.; Weissmiller, A. M.; Thomas, L. R.; Vakoc, C. R.; Hall, M. D.; Hiebert, S. W.; Liu, Q.; Stauffer, S. R.; Fesik, S. W.; Tansey, W. P. Displacement of WDR5 from Chromatin by a WIN Site Inhibitor with Picomolar Affinity. *Cell Rep.* **2019**, *26* (11), 2916-2928.e13. https://doi.org/10.1016/j.celrep.2019.02.047.

(99) Arafiles, J. V. V.; Hirose, H.; Hirai, Y.; Kuriyama, M.; Sakyiamah, M. M.; Nomura, W.; Sonomura, K.; Imanishi, M.; Otaka, A.; Tamamura, H.; Futaki, S. Discovery of a Macropinocytosis-Inducing Peptide Potentiated by Medium-Mediated Intramolecular Disulfide Formation. *Angew. Chem. Int. Ed.* **2021**, *60* (21), 11928–11936. https://doi.org/10.1002/anie.202016754.

(100) Ueda, T.; Konishi, H.; Manabe, K. Palladium-Catalyzed Carbonylation of Aryl, Alkenyl, and Allyl Halides with Phenyl Formate. *Org. Lett.* **2012**, *14* (12), 3100–3103. https://doi.org/10.1021/ol301192s.

(101) Kim, S. M.; Chae, M. K.; Lee, C.; Yim, M. S.; Bang, J. K.; Ryu, E. K. Enhanced Cellular Uptake of a TAT-Conjugated Peptide Inhibitor Targeting the Polo-Box Domain of Polo-like Kinase 1. *Amino Acids* **2014**, *46* (11), 2595–2603. https://doi.org/10.1007/s00726-014-1798-8.

(102) Sun, Y.; Lau, S. Y.; Lim, Z. W.; Chang, S. C.; Ghadessy, F.; Partridge, A.; Miserez, A. Phase-Separating Peptides for Direct Cytosolic Delivery and Redox-Activated Release of Macromolecular Therapeutics. *Nat. Chem.* **2022**, *14* (3), 274–283. https://doi.org/10.1038/s41557-021-00854-4.

(103) Qiu, Z.; Liu, W.; Zhu, Q.; Ke, K.; Zhu, Q.; Jin, W.; Yu, S.; Yang, Z.; Li, L.; Sun, X.; Ren, S.; Liu, Y.; Zhu, Z.; Zeng, J.; Huang, X.; Huang, Y.; Wei, L.; Ma, M.; Lu, J.; Chen, X.; Mou, Y.; Xie, T.; Sui, X. The Role and Therapeutic Potential of Macropinocytosis in Cancer. *Front. Pharmacol.* **2022**, *13*.

(104) Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R. Membrane Permeation of a Peptide: It Is Better to Be Positive. *J. Phys. Chem. B* **2015**, *119* (21), 6412–6420. https://doi.org/10.1021/acs.jpcb.5b02122.

(105) Trabulo, S.; Cardoso, A. L.; Mano, M.; De Lima, M. C. P. Cell-Penetrating Peptides—Mechanisms of Cellular Uptake and Generation of Delivery Systems. *Pharmaceuticals* **2010**, *3* (4), 961–993. https://doi.org/10.3390/ph3040961.

(106) Delling, U.; Roy, S.; Sumner-Smith, M.; Barnett, R.; Reid, L.; Rosen, C. A.; Sonenberg, N. The Number of Positively Charged Amino Acids in the Basic Domain of Tat Is Critical for Trans-Activation and Complex Formation with TAR RNA. *Proc. Natl. Acad. Sci.* **1991**, *88* (14), 6234–6238. https://doi.org/10.1073/pnas.88.14.6234.

(107) Bechara, C.; Sagan, S. Cell-Penetrating Peptides: 20 Years Later, Where Do We Stand? *FEBS Lett.* **2013**, *587* (12), 1693–1702. https://doi.org/10.1016/j.febslet.2013.04.031.

(108) Rong, G.; Wang, C.; Chen, L.; Yan, Y.; Cheng, Y. Fluoroalkylation Promotes Cytosolic Peptide Delivery. *Sci. Adv.* **2020**, *6* (33), eaaz1774. https://doi.org/10.1126/sciadv.aaz1774.

(109) Northrop, B. H.; Frayne, S. H.; Choudhary, U. Thiol–Maleimide "Click" Chemistry: Evaluating the Influence of Solvent, Initiator, and Thiol on the Reaction Mechanism, Kinetics, and Selectivity. *Polym. Chem.* **2015**, *6* (18), 3415–3430. https://doi.org/10.1039/C5PY00168D.

(110) Kauffman, W. B.; Fuselier, T.; He, J.; Wimley, W. C. Mechanism Matters: A Taxonomy of Cell Penetrating Peptides. *Trends Biochem. Sci.* **2015**, *40* (12), 749–764. https://doi.org/10.1016/j.tibs.2015.10.004.

(111) Illien, F.; Rodriguez, N.; Amoura, M.; Joliot, A.; Pallerla, M.; Cribier, S.; Burlina, F.; Sagan, S. Quantitative Fluorescence Spectroscopy and Flow Cytometry Analyses of Cell-Penetrating Peptides Internalization Pathways: Optimization, Pitfalls, Comparison with Mass Spectrometry Quantification. *Sci. Rep.* **2016**, *6* (1), 36938. https://doi.org/10.1038/srep36938.

(112) Lim, Z. W.; Varma, V. B.; Ramanujan, R. V.; Miserez, A. Magnetically Responsive Peptide Coacervates for Dual Hyperthermia and Chemotherapy Treatments of Liver Cancer. *Acta Biomater.* **2020**, *110*, 221–230. https://doi.org/10.1016/j.actbio.2020.04.024.

(113) Jayashankar, V.; Edinger, A. L. Macropinocytosis Confers Resistance to Therapies Targeting Cancer Anabolism. *Nat. Commun.* **2020**, *11* (1), 1121. https://doi.org/10.1038/s41467-020-14928-3.

(114) Thieriet, N.; Alsina, J.; Giralt, E.; Guibé, F.; Albericio, F. Use of Alloc-Amino Acids in Solid-Phase Peptide Synthesis. Tandem Deprotection-Coupling Reactions Using Neutral Conditions. *Tetrahedron Lett.* **1997**, *38* (41), 7275–7278. https://doi.org/10.1016/S0040-4039(97)01690-0.

(115) Fukuyama, T.; Cheung, M.; Kan, T. N-Carboalkoxy-2-Nitrobenzenesulfonamides: A Practical Preparation of N-Boc-, N-Alloc-, and N-Cbz-Protected Primary Amines. *Synlett* **1999**, *1999* (8), 1301–1303. https://doi.org/10.1055/s-1999-2827.

(116) Winter, G. *Xia2* : An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr.* **2010**, *43* (1), 186–190. https://doi.org/10.1107/S0021889809045701.

(117)Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the *CCP* 4 Suite and Current Developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (4), 235–242. https://doi.org/10.1107/S0907444910045749.

(118) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. It Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.

(119) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* **2019**, *75* (10), 861–877. https://doi.org/10.1107/S2059798319011471.

(120) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (4), 486–501. https://doi.org/10.1107/S0907444910007493.

(121)Couture, J.-F.; Collazo, E.; Trievel, R. C. Molecular Recognition of Histone H3 by the WD40 Protein WDR5. *Nat. Struct. Mol. Biol.* **2006**, *13* (8), 698–703. https://doi.org/10.1038/nsmb1116.

(122) Patel, A.; Dharmarajan, V.; Cosgrove, M. S. Structure of WDR5 Bound to Mixed Lineage Leukemia Protein-1 Peptide. *J. Biol. Chem.* **2008**, *283* (47), 32158–32161. https://doi.org/10.1074/jbc.C800164200.

(123) Vandesompele, J. *Eurogentec qPCR Guide*.

(124) Livak, K. J.; Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **2001**, *25* (4), 402–408. https://doi.org/10.1006/meth.2001.1262.

(125) Holm, T.; Johansson, H.; Lundberg, P.; Pooga, M.; Lindgren, M.; Langel, Ü. Studying the Uptake of Cell-

Penetrating Peptides. *Nat. Protoc.* **2006**, *1* (2), 1001–1005. https://doi.org/10.1038/nprot.2006.174.

(126) Jakic, B.; Buszko, M.; Cappellano, G.; Wick, G. Elevated Sodium Leads to the Increased Expression of HSP60 and Induces Apoptosis in HUVECs. *PLOS ONE* **2017**, *12* (6), e0179383. https://doi.org/10.1371/journal.pone.0179383.

(127) Chou, P.-H.; Wang, S.-T.; Ma, H.-L.; Liu, C.-L.; Chang, M.-C.; Lee, O. K.-S. Development of a Two-Step Protocol for Culture Expansion of Human Annulus Fibrosus Cells with TGF-B1 and FGF-2. *Stem Cell Res. Ther.* **2016**, *7* (1), 89. https://doi.org/10.1186/s13287-016-0332-1.

(128) Iempridee, T.; Wiwithaphon, S.; Piboonprai, K.; Pratedrat, P.; Khumkhrong, P.; Japrung, D.; Temisak, S.; Laiwejpithaya, S.; Chaopotong, P.; Dharakul, T. Identification of Reference Genes for Circulating Long Noncoding RNA Analysis in Serum of Cervical Cancer Patients. *FEBS Open Bio* **2018**, *8* (11), 1844–1854. https://doi.org/10.1002/2211-5463.12523.

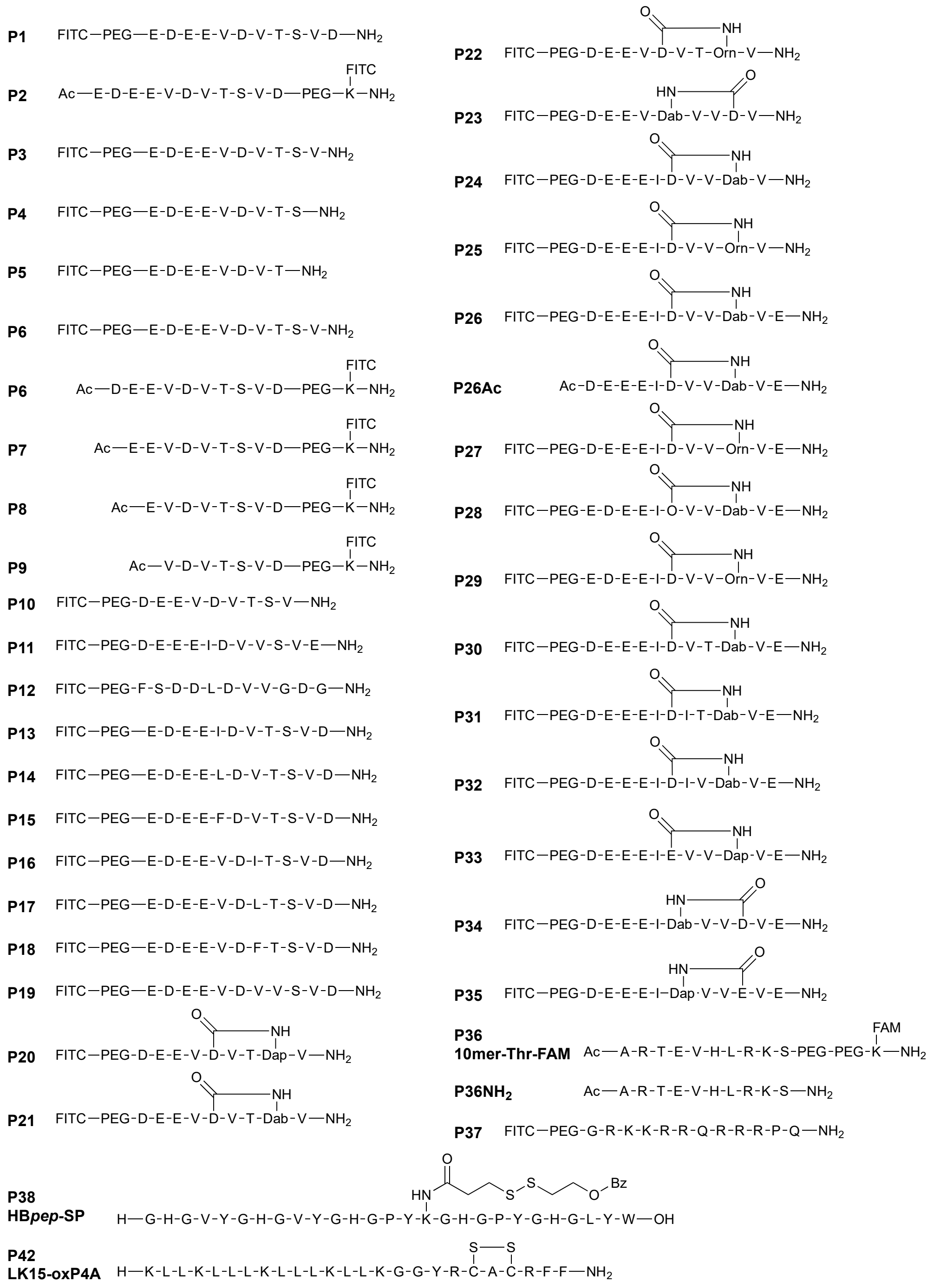
(129) Li, Y.; Liu, Y.; Chen, G.; Liu, H.; Wu, Y.; Liu, J.; Zhang, Z. HOTTIP Is Upregulated in Esophageal Cancer and Triggers the Drug Resistance.

(130) Subhash, S.; Mishra, K.; Akhade, V. S.; Kanduri, M.; Mondal, T.; Kanduri, C. H3K4me2 and WDR5 Enriched Chromatin Interacting Long Non-Coding RNAs Maintain Transcriptionally Competent Chromatin at Divergent Transcriptional Units. *Nucleic Acids Res.* **2018**, *46* (18), 9384–9400. https://doi.org/10.1093/nar/gky635.

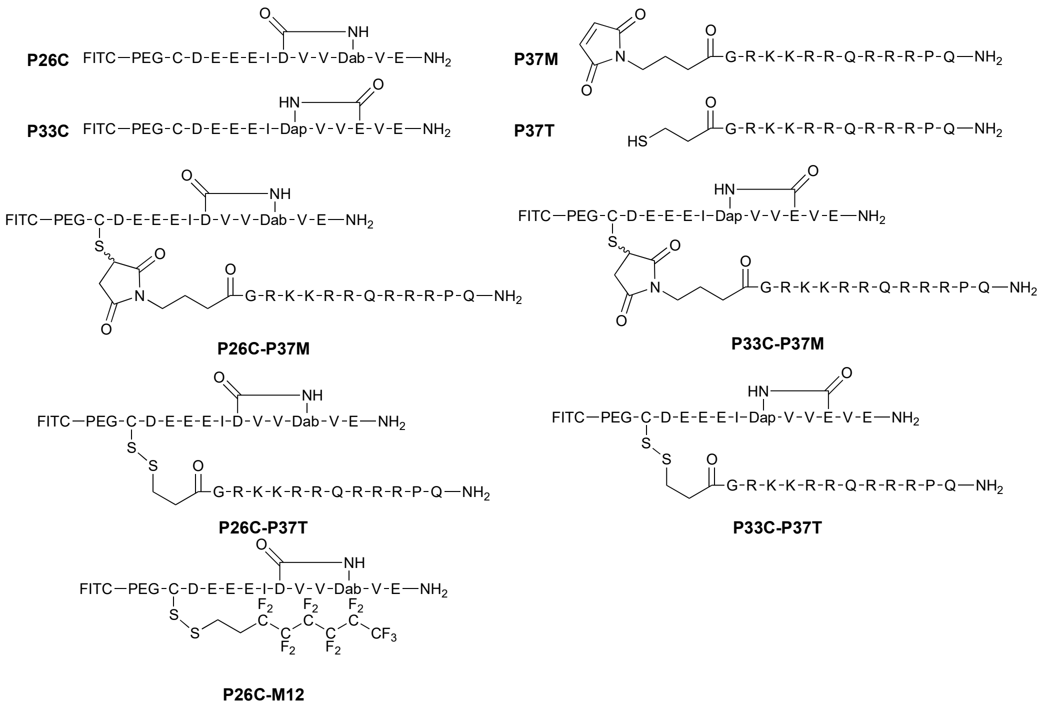
(131) Scavello, F.; Zeni, F.; Milano, G.; Macrì, F.; Castiglione, S.; Zuccolo, E.; Scopece, A.; Pezone, G.; Tedesco, C. C.; Nigro, P.; Degani, G.; Gambini, E.; Veglia, F.; Popolo, L.; Pompilio, G.; Colombo, G. I.; Bianchi, M. E.; Raucci, A. Soluble Receptor for Advanced Glycation End-Products Regulates Age-Associated Cardiac Fibrosis. *Int. J. Biol. Sci.* **2021**, *17* (10), 2399–2416. https://doi.org/10.7150/ijbs.56379.

Chapter 6. Appendix

6.1 Characterization of Peptides



Supplemental Figure 1: Structure of peptides **P1** – **P42**.



Supplemental Figure 2: Structure of peptides involved in conjugation chemistry.

Supplemental Table 1: List of peptides' synthesis methods, purification conditions, purity, and retention time.

Nr	Elongation	Cyclization	Purification	Purity	Retention time (min)
P1	B	-	Acidic	97.2%	9.598
P2	B	-	Acidic	96.8%	9.456
P3	B	-	Acidic	98.6%	9.728
P4	B	-	Acidic	98.5%	9.393
P5	B	-	Acidic	99.7%	9.562
P6	B	-	Acidic	98.1%	9.529
P7	B	-	Acidic	97.2%	9.592
P8	B	-	Acidic	96.8%	9.701
P9	B	-	Acidic	97.5%	9.969
P10	B	-	Acidic	97.0%	9.911
P11	C	-	Basic	96.9%	10.29
P12	B	-	Acidic	99.4%	11.09
P13	B	-	Acidic	99.4%	9.832
P14	B	-	Acidic	99.2%	9.932
P15	B	-	Acidic	99.4%	10.09
P16	B	-	Acidic	99.4%	9.857
P17	B	-	Acidic	99.1%	9.947

P18	B	-	Acidic	99.1%	10.08
P19	B	-	Acidic	98.5%	9.966
P20	A	D	Acidic	97.6%	10.17
P21	A	D	Acidic	95.0%	10.16
P22	A	D	Acidic	97.8%	10.23
P23	A	D	Acidic	95.4%	10.24
P24	A	D	Acidic	95.6%	10.92
P25	A	D	Acidic	90.4%	10.90
P26	C	E	Basic	95.8%	10.70
P26Ac	B	E	Basic	98.3%[a]	8.746[a]
P26C	C	E	Basic	n.d.	n.d.
P27	A	D	Acidic	85.0%	10.64
P28	A	D	Acidic	97.8%	10.72
P29	A	D	Acidic	89.0%	10.64
P30	A	D	Acidic	98.0%	10.07
P31	A	D	Acidic	93.5%	10.35
P32	A	D	Acidic	98.1%	11.01
P33	C	E	Basic	83.2%	10.05
P33C	C	E	Basic	n.d.	n.d.
P34	B	E	Basic	96.1%	10.82
P35	B	E	Basic	78.9%	10.83
P36	B	-	Acidic	>99.5%	8.607
P36NH2	B	-	Acidic	>99.5%[a]	6.322[a]
P37	A	-	Acidic	>99.5%	10.03
P37M	C	-	Acidic	n.d.	n.d.
P37T	C	-	Acidic	n.d.	n.d.
P26-P37M	B	-	Acidic	99.1%	9.429
P26-P37T	B	-	Acidic	98.0%	9.372
P26-M12	B	-	Acidic	97.6%	14.77
P33-P37M	B	-	Acidic	99.1%	9.518
P33-P37T	B	-	Acidic	99.4%	8.970
P38	C	-	Acidic	95.9%[a]	10.93[a]
P42	C	-	Acidic	95.6%[a]	16.27[a]

[a]: Recorded in 210 nm channel. n.d.: not determined.

Supplemental Table 2: List of peptides' HRMS.

Nr	Calc. m/z		Obs. m/z	ppm
P1	[M+H]$^{+}$	1769.63707	1769.63639	-0.38
P2	[M+H]$^{+}$	1939.74260	1939.74319	0.31
P3	[M+2H]$^{2+}$	1653.61013	1653.61042	0.18
P4	[M+H]$^{+}$	1555.54171	1555.54223	0.33
P5	[M+H]$^{+}$	1468.50968	1468.51076	0.74
P6	[M+2H]$^{2+}$	905.85364	905.85489	1.38
P7	[M+H]$^{+}$	1695.67306	1695.67275	-0.18
P8	[M+H]$^{+}$	1566.63047	1566.63068	0.14
P9	[M+H]$^{+}$	1437.58787	1437.58913	0.87
P10	[M+H]$^{+}$	763.28740	763.28791	0.66
P11	[M+Na+H]$^{2+}$	909.33916	909.33947	0.34
P12	[M+H]$^{+}$	1671.61554	1671.61547	-0.05
P13	[M+H]$^{+}$	1783.65272	1783.65174	-0.55
P14	[M+H]$^{+}$	1783.65272	1783.65172	-0.56
P15	[M+H]$^{+}$	1817.63707	1817.63552	-0.85
P16	[M+H]$^{+}$	1783.65272	1783.65162	-0.62
P17	[M+H]$^{+}$	1783.65272	1783.65166	-0.59
P18	[M+H]$^{+}$	1817.63707	1817.63568	-0.77
P19	[M+H]$^{+}$	1767.65780	1767.65688	-0.52
P20	[M+H]$^{+}$	1506.57295	1506.57368	1.31
P21	[M+H]$^{+}$	1520.58860	1520.58958	0.64
P22	[M+H]$^{+}$	1534.60425	1534.60512	0.57
P23	[M+H]$^{+}$	1520.58860	1520.59305	2.93
P24	[M+H]$^{+}$	1661.66758	1661.66733	0.15
P25	[M+H]$^{+}$	1675.68323	1675.68282	0.24
P26	[M+H]$^{+}$	1790.71017	1790.70992	-0.14
P26Ac	[M+2H]$^{2+}$	649.80916	649.80998	1.25
P26C	[M+H]$^{+}$	1893.71936	1893.72082	0.77
P27	[M+2H]$^{2+}$	902.86655	902.86739	0.93
P28	[M+2H]$^{2+}$	895.85873	895.8597	1.08
P29	[M+2H]$^{2+}$	902.86655	902.8671	0.61
P30	[M+2H]$^{2+}$	896.84836	896.85091	2.85
P31	[M+2H]$^{2+}$	903.85618	903.85858	2.65
P32	[M+2H]$^{2+}$	902.86655	902.86907	2.79
P33	[M+H]$^{+}$	1790.71017	1790.71363	1.93

P33C	$[M+2H]^{2+}$	947.36332	947.36394	0.66
P34	$[M+H]^{+}$	1790.71017	1790.7128	1.47
P35	$[M+H]^{+}$	1790.71017	1790.71242	1.25
P36	$[M+H]^{+}$	975.51524	975.51661	1.40
P36NH2	$[M+2H]^{2+}$	619.35983	619.35941	0.67
P37	$[M+3H]^{3+}$	719.0492	719.0485	0.97
P37M	$[M+H]^{+}$	596.0[a]	596.2[a]	LRMS
P37T	$[M+H]^{+}$	855.01368	855.01363	0.06
P26-P37M	$[M+6H]^{6+}$	613.30121	613.30004	1.91
P26-P37T	$[M+4H]^{4+}$	900.68459	900.68513	0.60
P26-M12	$[M+2H]^{2+}$	1136.35071	1136.35179	0.95
P33-P37M	$[M+4H]^{4+}$	920.95125	920.95086	0.42
P33-P37T	$[M+H]^{+}$	900.93654	900.93542	1.24
P38	$[M+5H]^{5+}$	639.87958	639.87821	2.14
P42	$[M+3H]^{3+}$	984.28932	984.28906	0.26

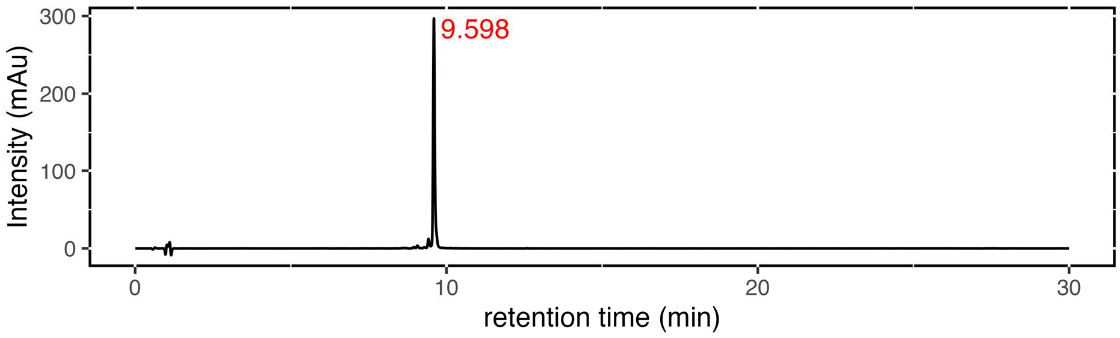
[a]: Recorded LRMS.

6.1.1 HPLC Analysis of Compounds

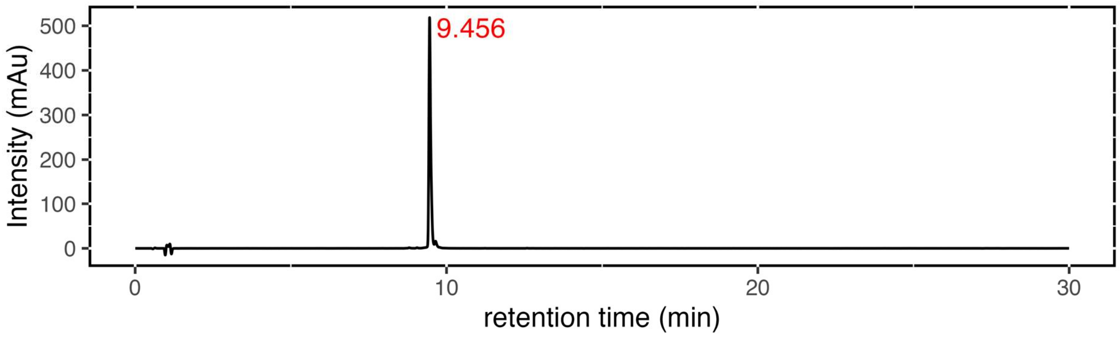
Analytical HPLC-MS was performed by using Agilent 1200 Infinity II LC system equipped with a C18 column (Agilent Poroshell 120, 2.7 µm, 3 x 100 mm, flow rate: 0.4 mL/min, temperature: 33ºC) and Agilent InfinityLab LC/MSD G6125C. Compounds were eluted with a gradient showing in the supplementary table 4. In short, a linear gradient from 5% to 95% (percentage of B. A: 0.1% TFA in H2O; B: 0.1% TFA in ACN) over 20 min. The whole UV-VIS spectrum was recorded during the run, and the selected channel (210 / 254 nm), which was indicated in Supplemental Table 3 for each compound, was used to plot and calculate the purity of the sample. Isolating yield was calculated from dividing the amount of purified product by the amount of resin loading.

Supplemental Table 3. Gradient program for the analytical HPLC-MS analysis.

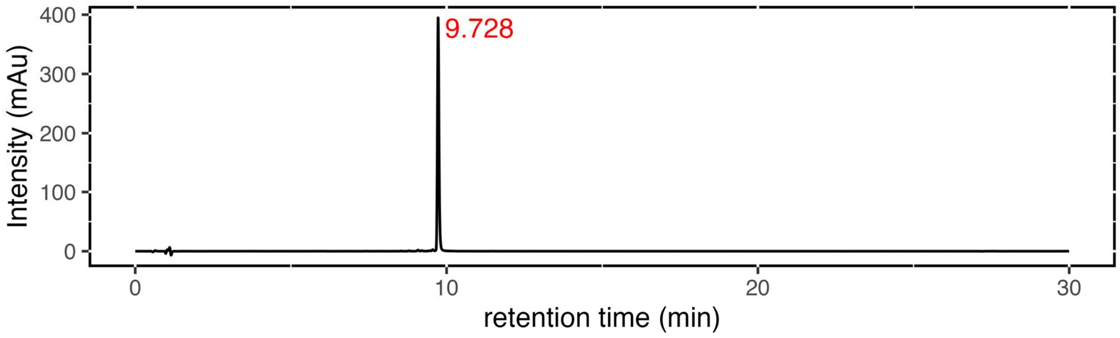
Time points (min)	0 min	1 min	21 min	25 min	27 min	30 min
B (%)	5%	5%	95%	95%	5%	5%



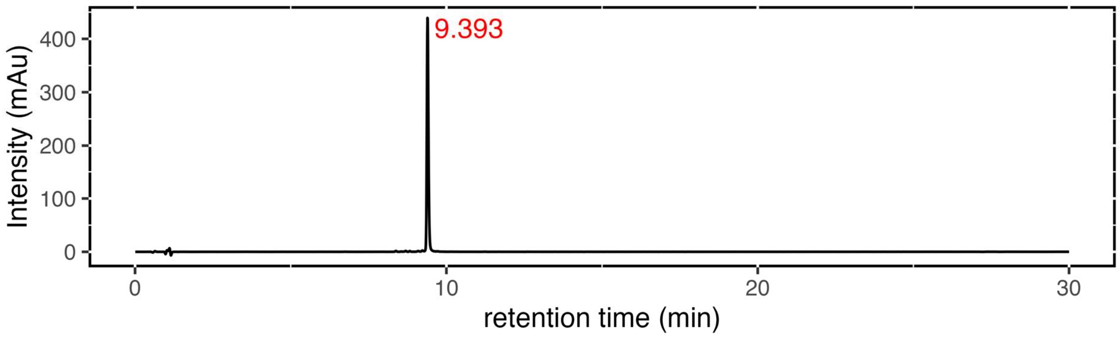
Supplemental Figure 3: HPLC analysis of the compound **P1**.



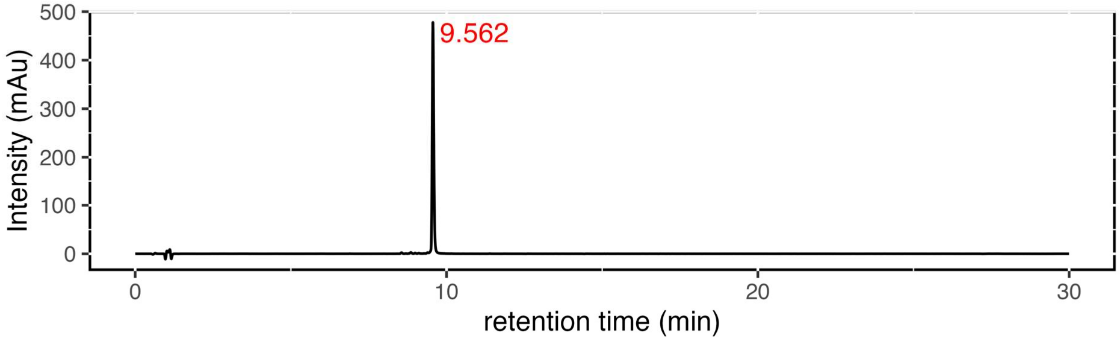
Supplemental Figure 4: HPLC analysis of the compound **P2**.



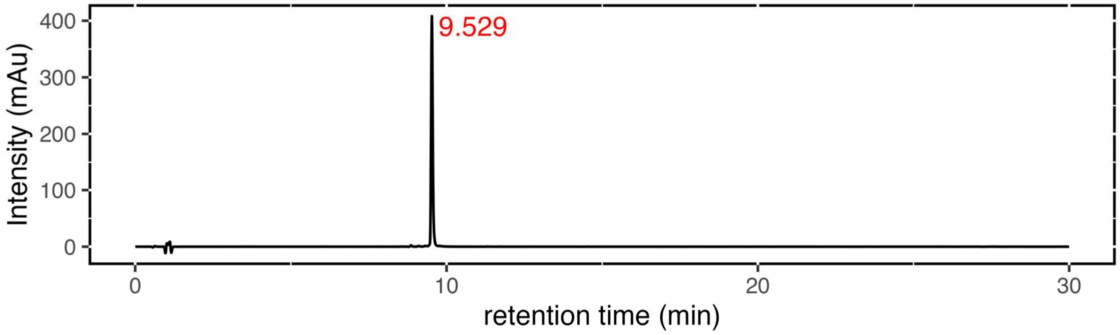
Supplemental Figure 5: HPLC analysis of the compound **P3**.



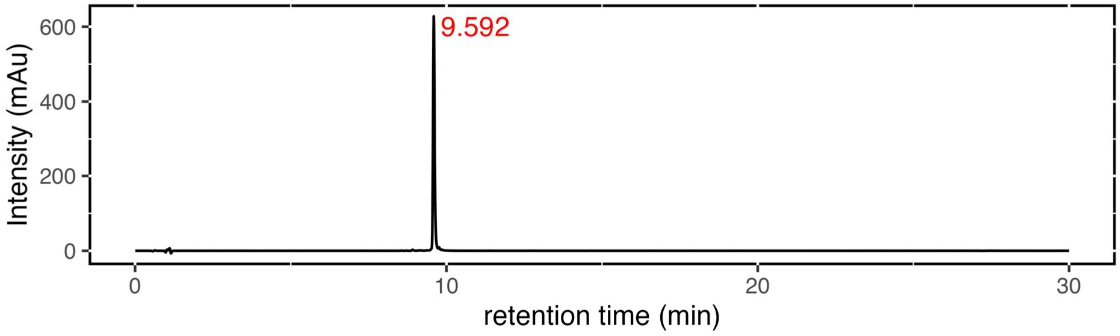
Supplemental Figure 6: HPLC analysis of the compound **P4**.



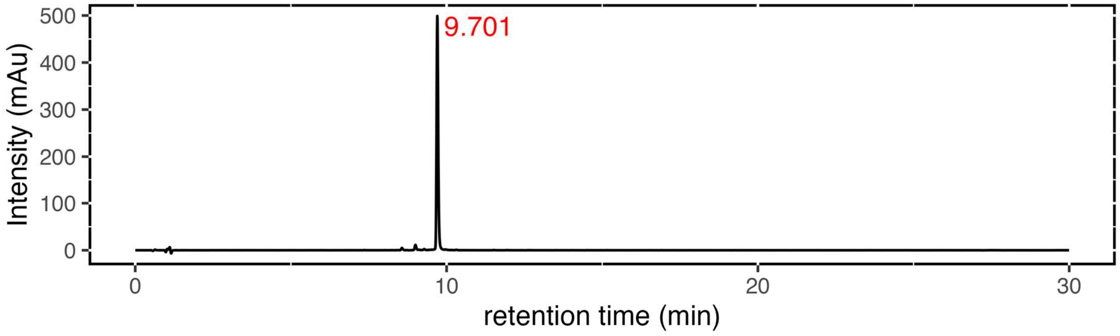
Supplemental Figure 7: HPLC analysis of the compound **P5**.



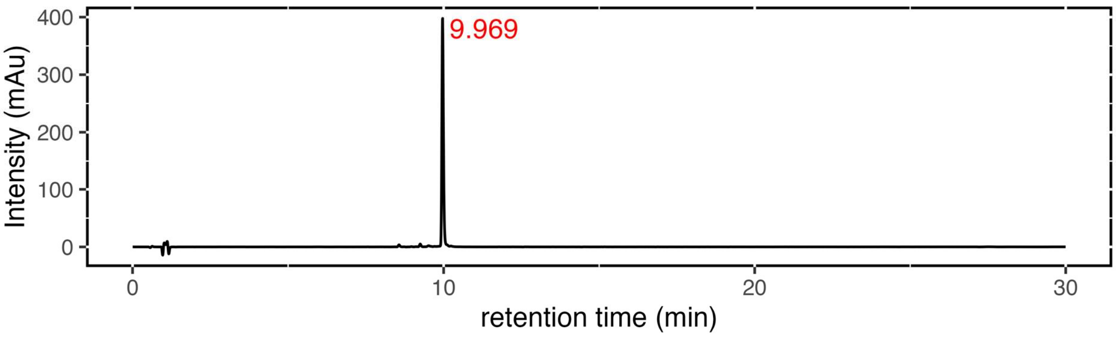
Supplemental Figure 8: HPLC analysis of the compound **P6**.



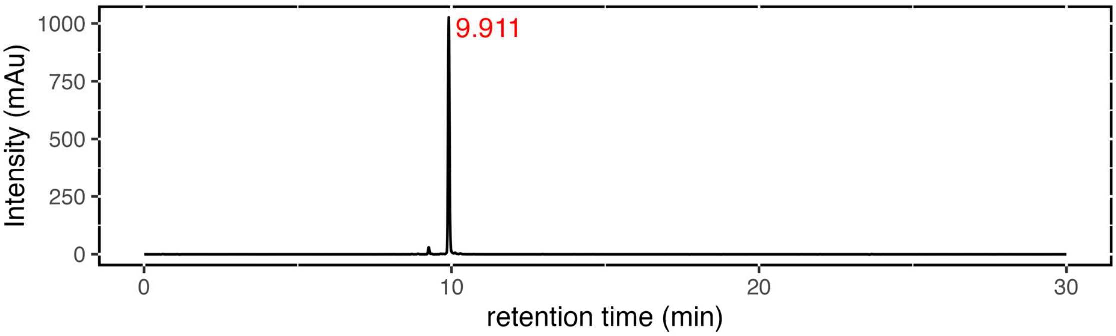
Supplemental Figure 9: HPLC analysis of the compound **P7**.



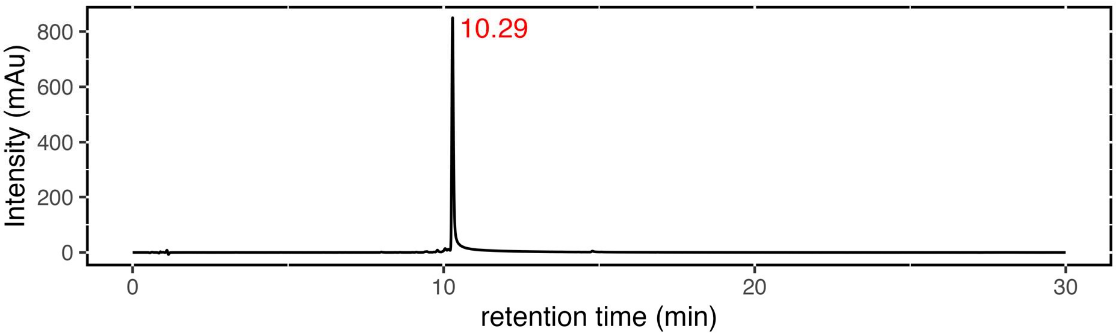
Supplemental Figure 10: HPLC analysis of the compound **P8**.



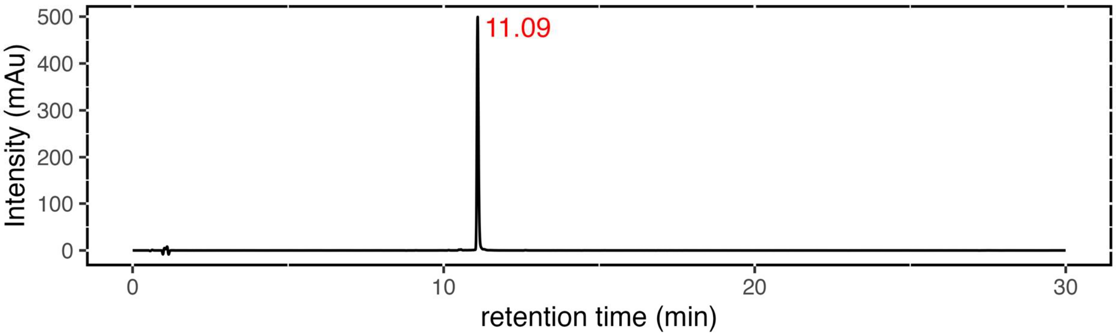
Supplemental Figure 11: HPLC analysis of the compound **P9.**



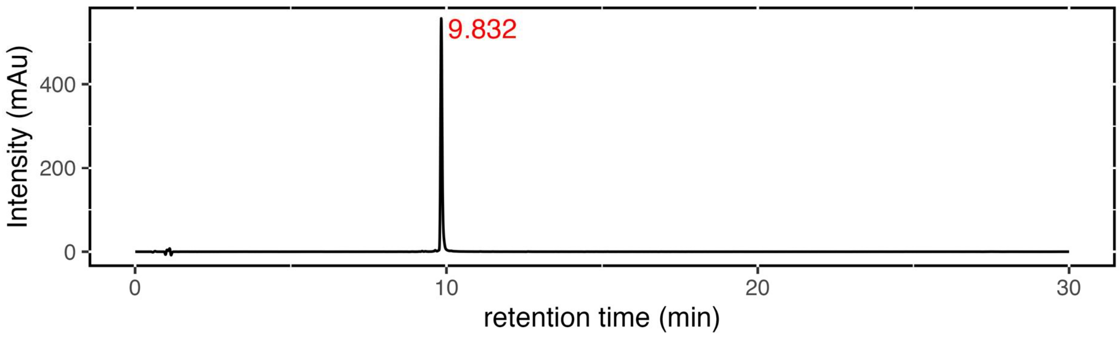
Supplemental Figure 12: HPLC analysis of the compound **P10**.



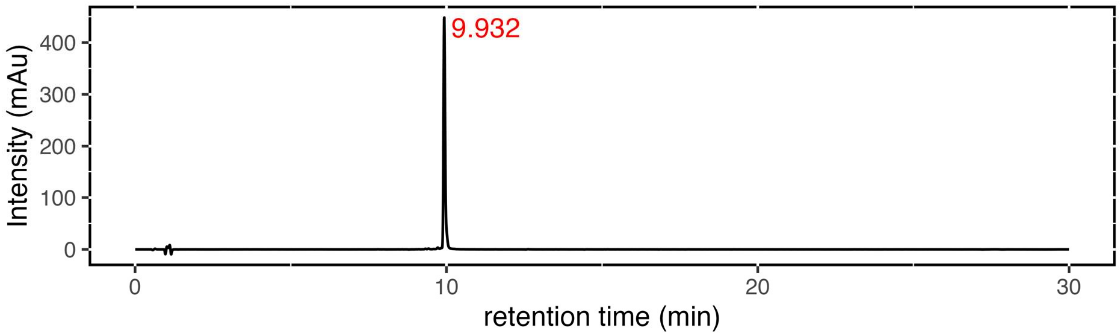
Supplemental Figure 13: HPLC analysis of the compound **P11**.



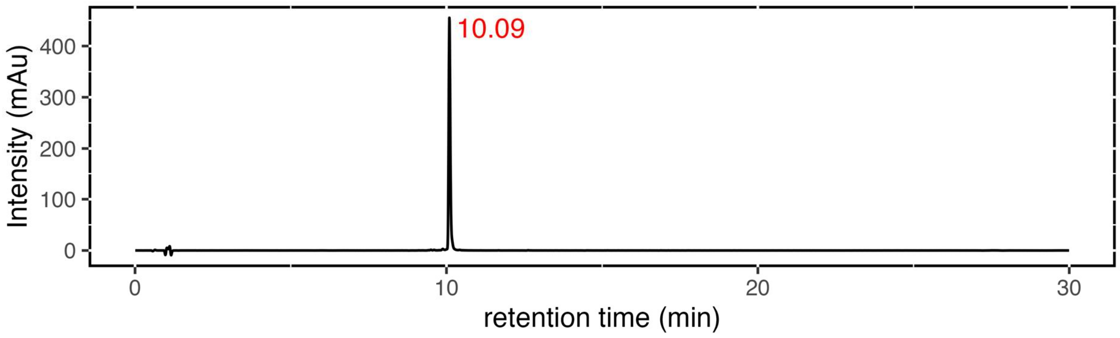
Supplemental Figure 14: HPLC analysis of the compound **P12**.



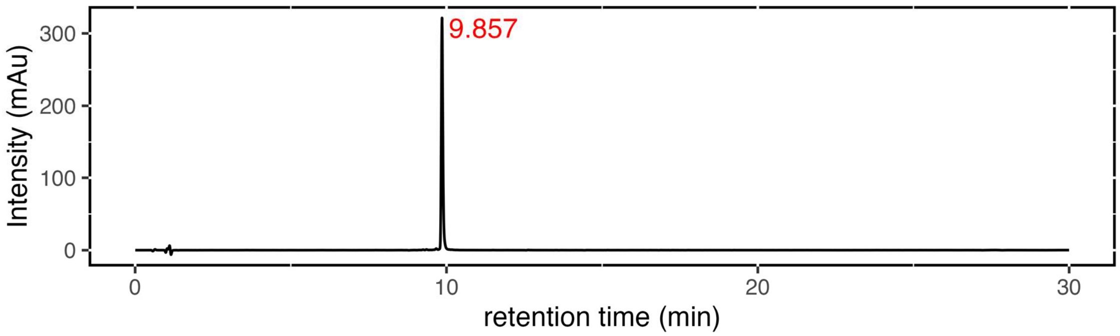
Supplemental Figure 15: HPLC analysis of the compound **P13**.



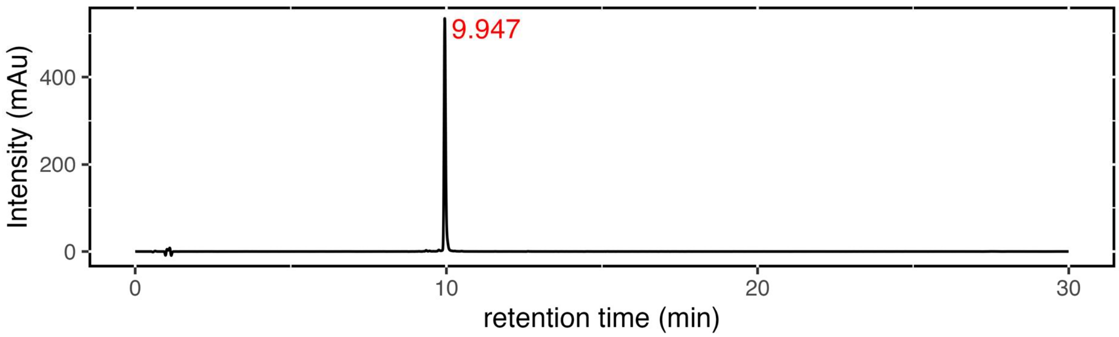
Supplemental Figure 16: HPLC analysis of the compound **P14**.



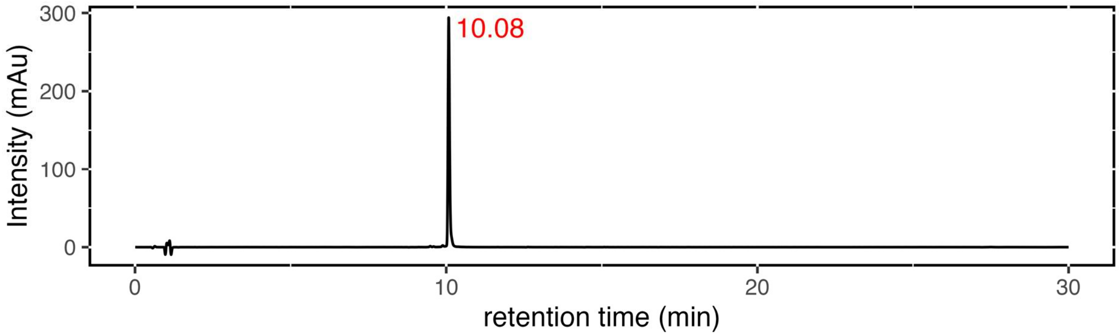
Supplemental Figure 17: HPLC analysis of the compound **P15**.



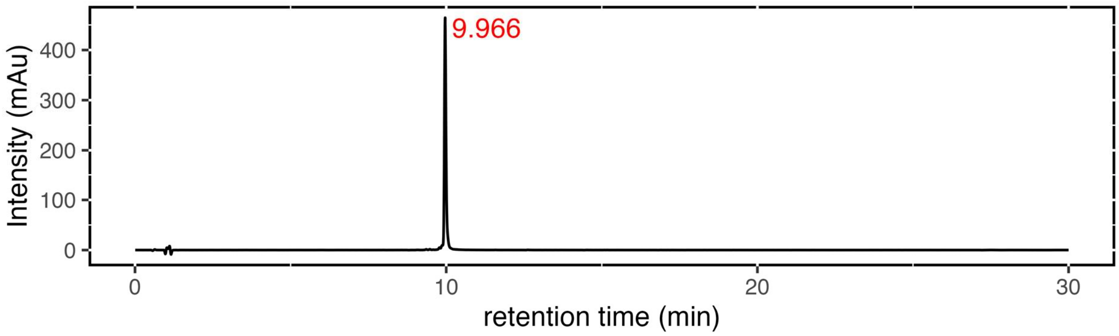
Supplemental Figure 18: HPLC analysis of the compound **P16**.



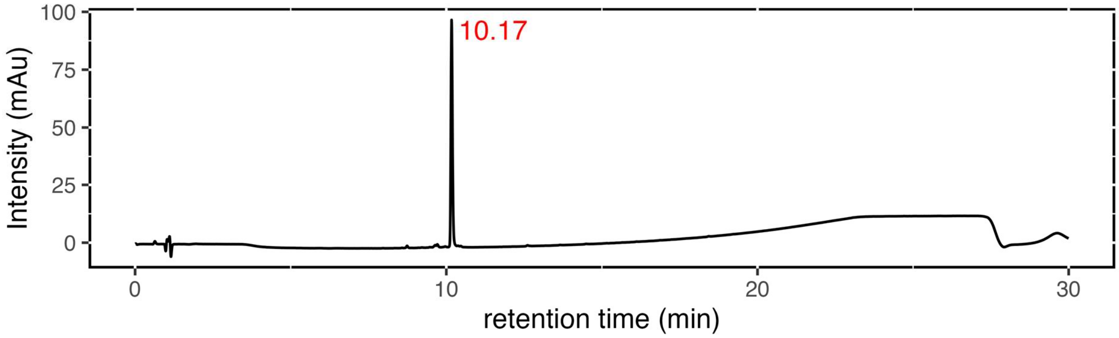
Supplemental Figure 19: HPLC analysis of the compound **P17**.



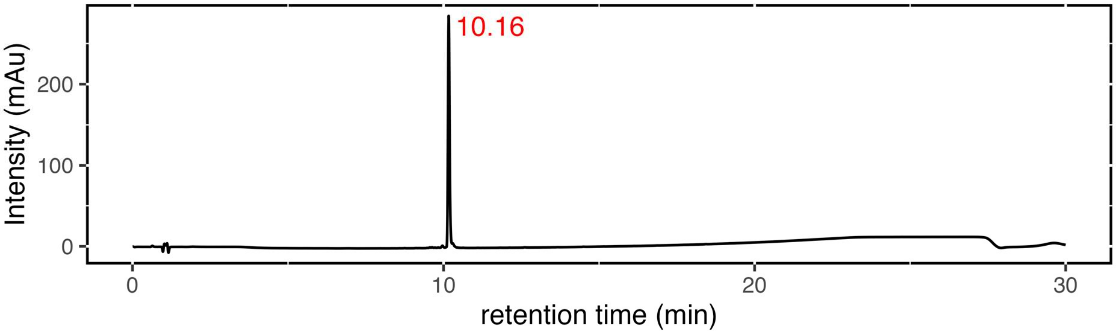
Supplemental Figure 20: HPLC analysis of the compound **P18**.



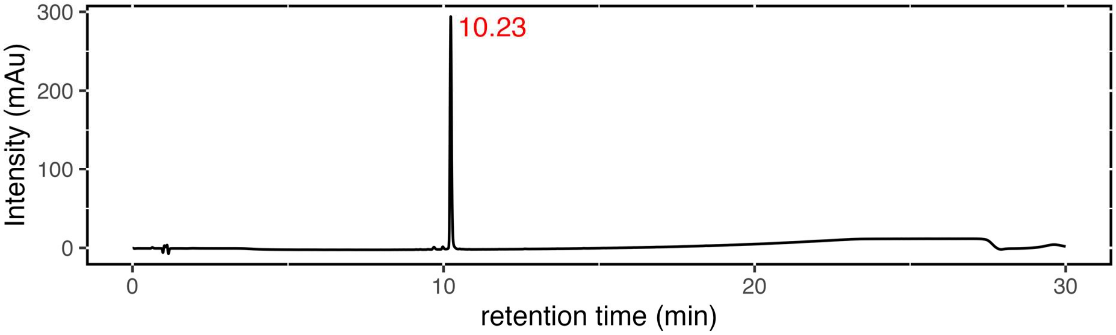
Supplemental Figure 21: HPLC analysis of the compound **P19**.



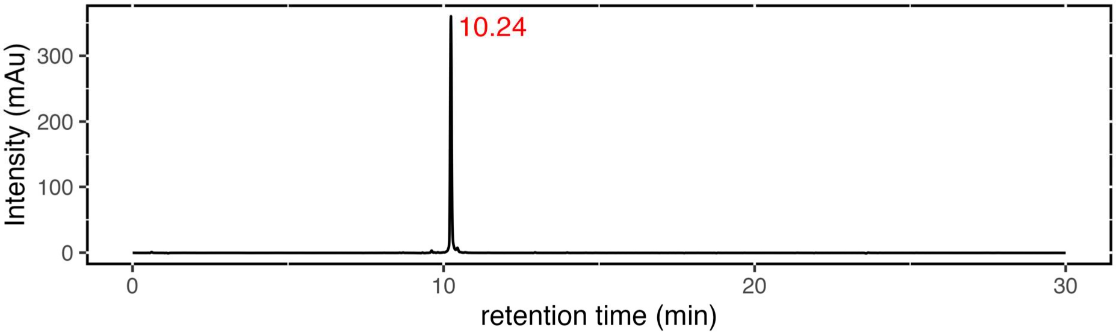
Supplemental Figure 22: HPLC analysis of the compound **P20**.



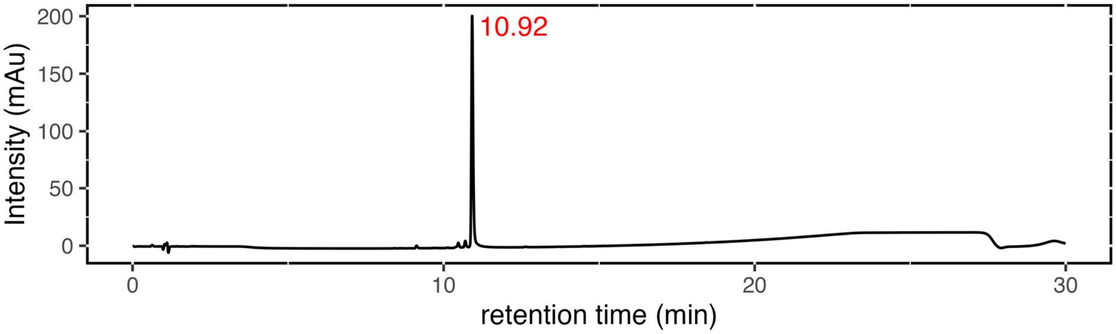
Supplemental Figure 23: HPLC analysis of the compound **P21**.



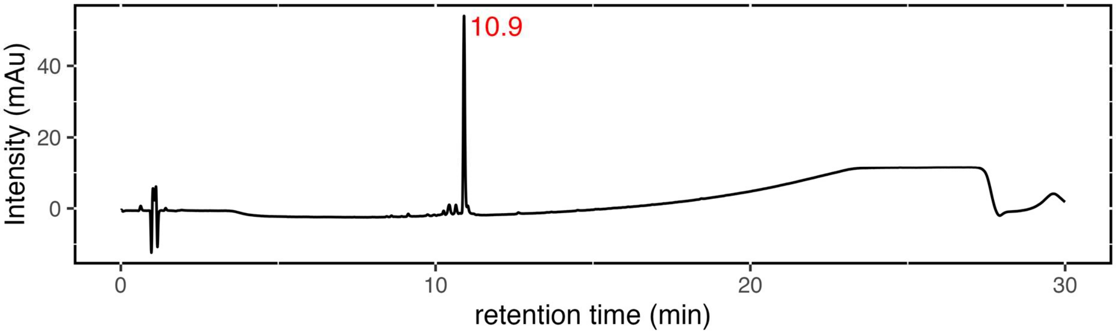
Supplemental Figure 24: HPLC analysis of the compound **P22**.



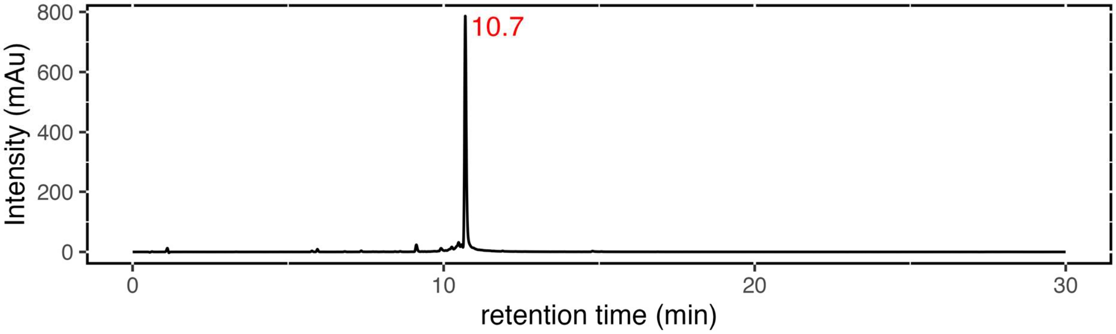
Supplemental Figure 25: HPLC analysis of the compound **P23**.



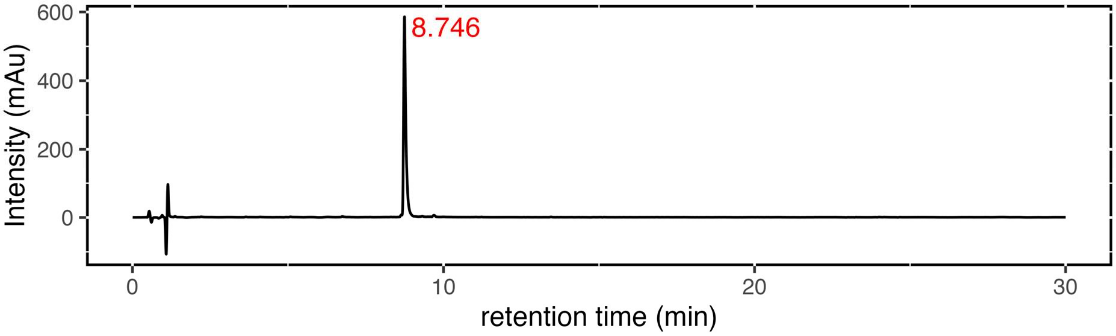
Supplemental Figure 26: HPLC analysis of the compound **P24**.



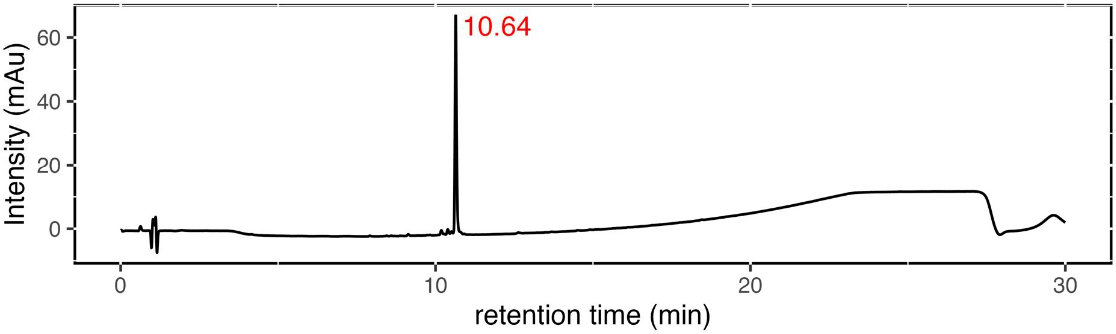
Supplemental Figure 27: HPLC analysis of the compound **P25**.



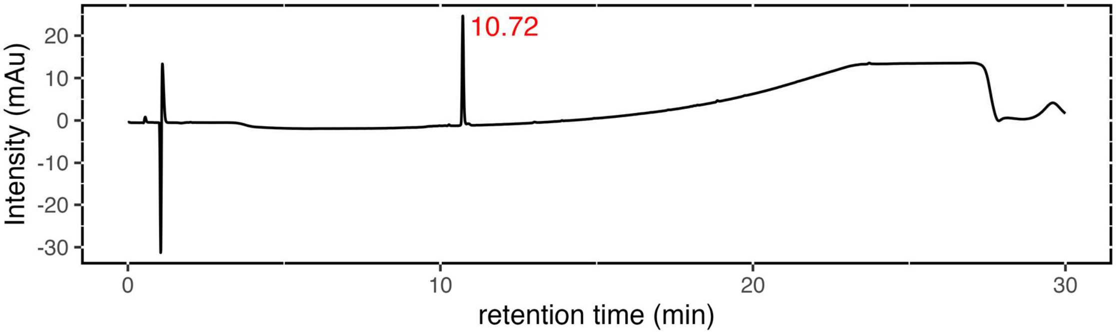
Supplemental Figure 28: HPLC analysis of the compound **P26**.



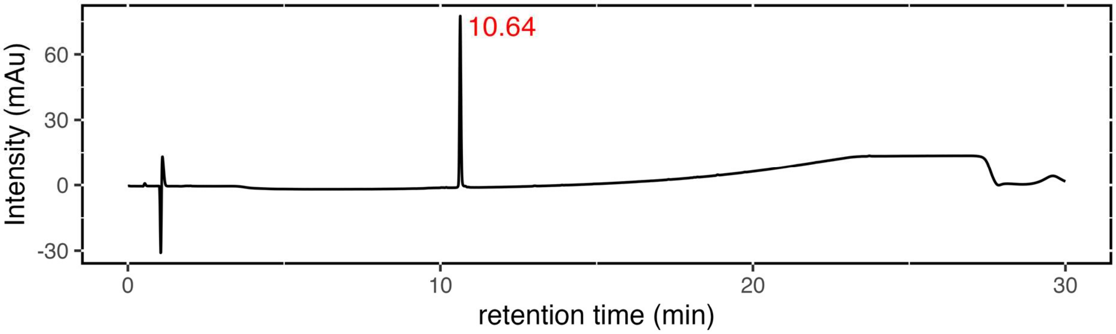
Supplemental Figure 29: HPLC analysis of the compound **P26Ac** in the 210 nm channel.



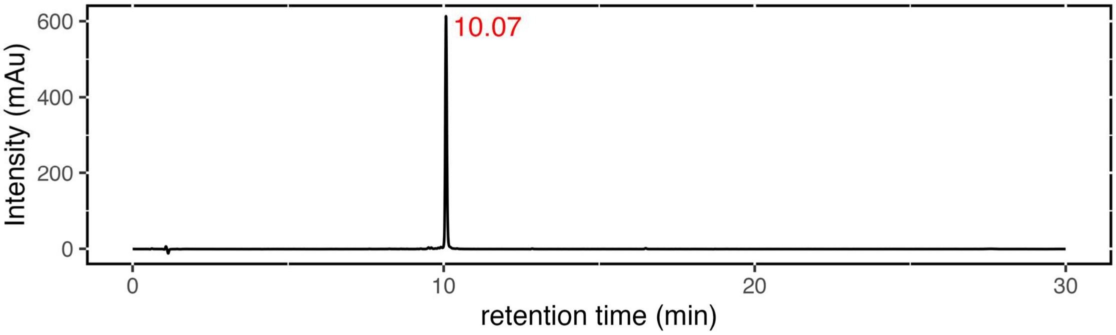
Supplemental Figure 30: HPLC analysis of the compound **P27**.



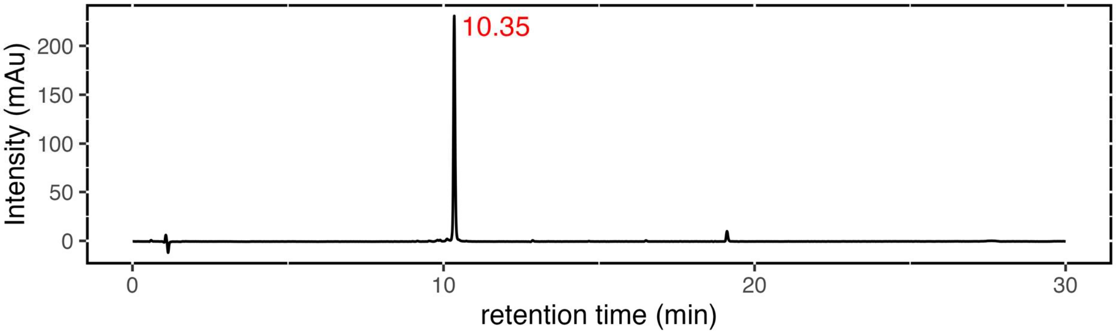
Supplemental Figure 31: HPLC analysis of the compound **P28**.



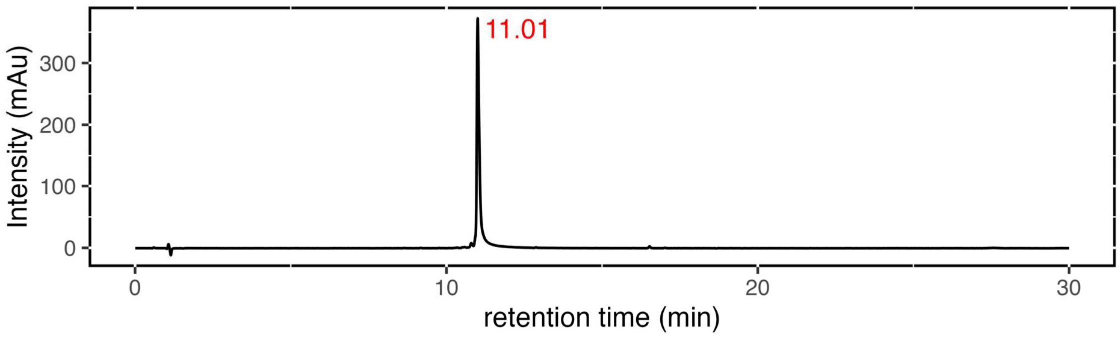
Supplemental Figure 32: HPLC analysis of the compound **P29**.



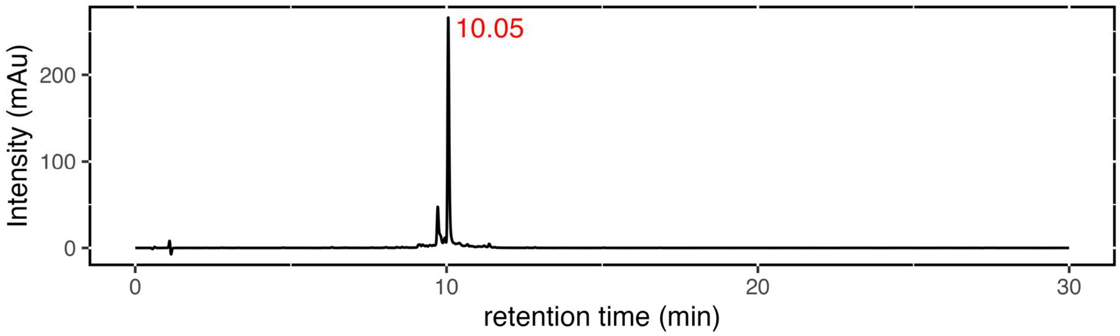
Supplemental Figure 33: HPLC analysis of the compound **P30**.



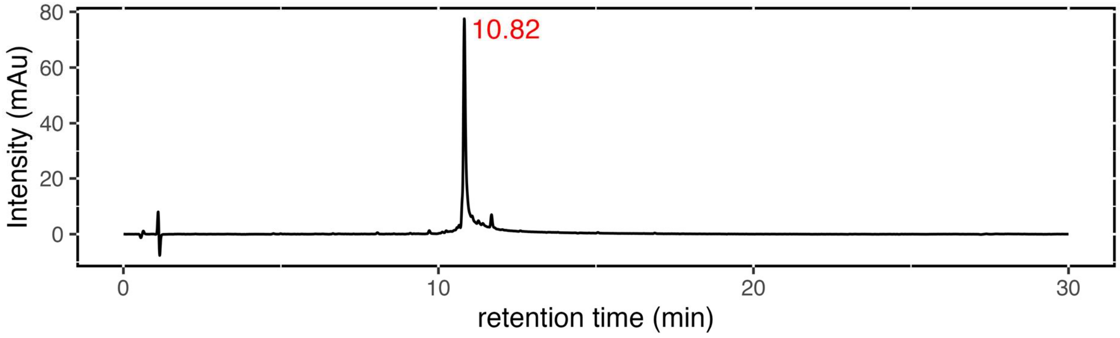
Supplemental Figure 34: HPLC analysis of the compound **P31**.



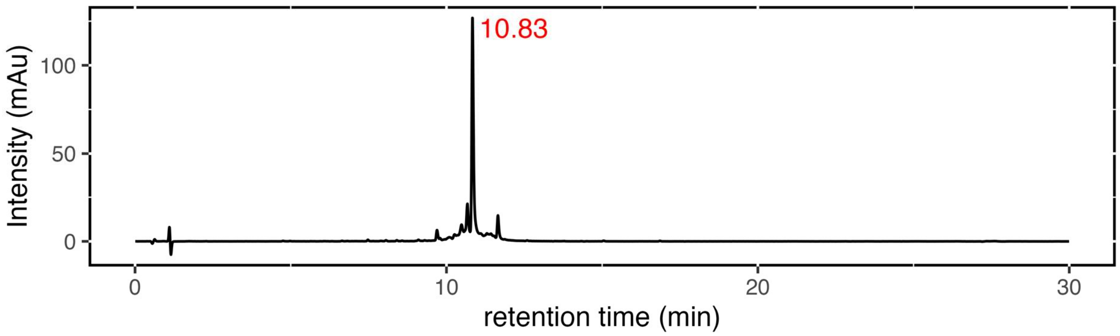
Supplemental Figure 35: HPLC analysis of the compound **P32**.



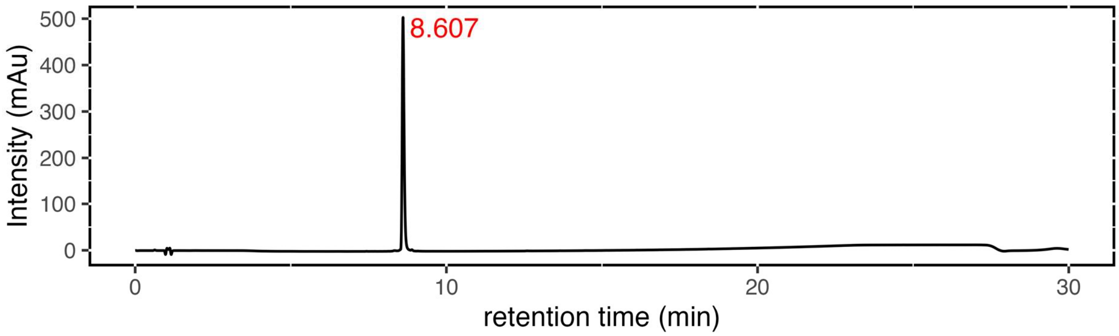
Supplemental Figure 36: HPLC analysis of the compound **P33**.



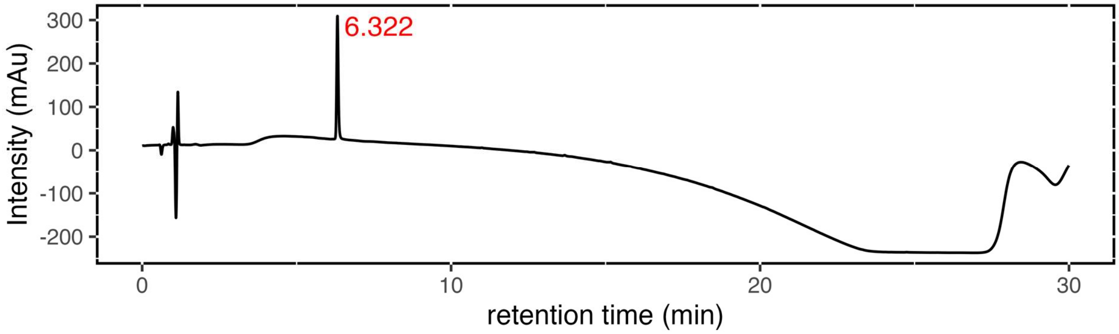
Supplemental Figure 37: HPLC analysis of the compound **P34**.



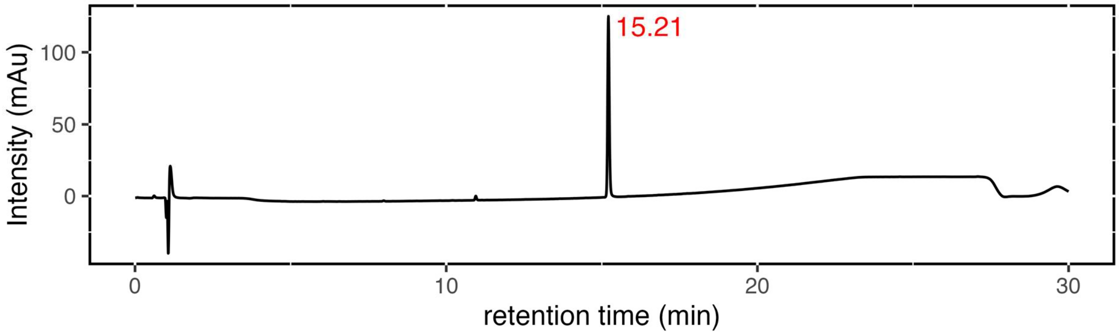
Supplemental Figure 38: HPLC analysis of the compound **P35**.



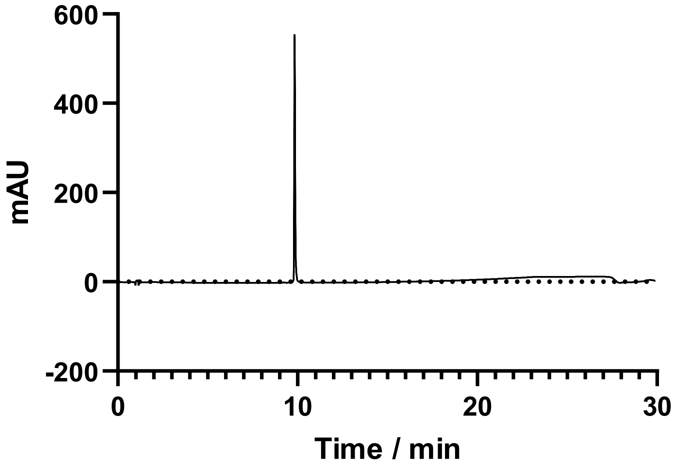
Supplemental Figure 39: HPLC analysis of the compound **P36**.



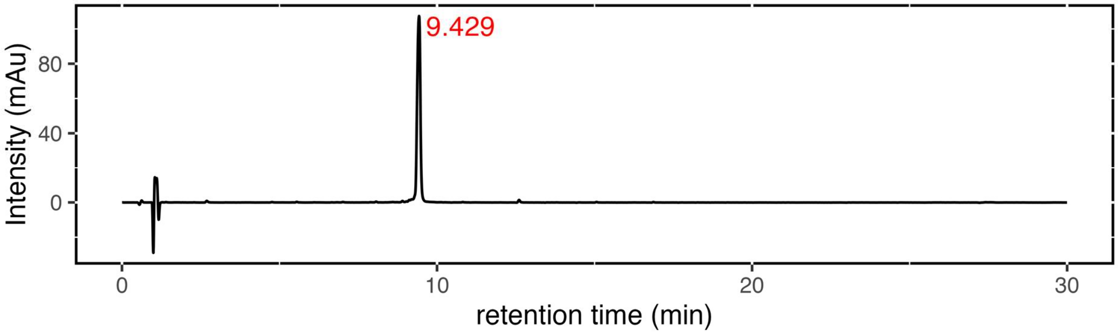
Supplemental Figure 40: HPLC analysis of the compound **P36NH$_2$** in the 210 nm channel.



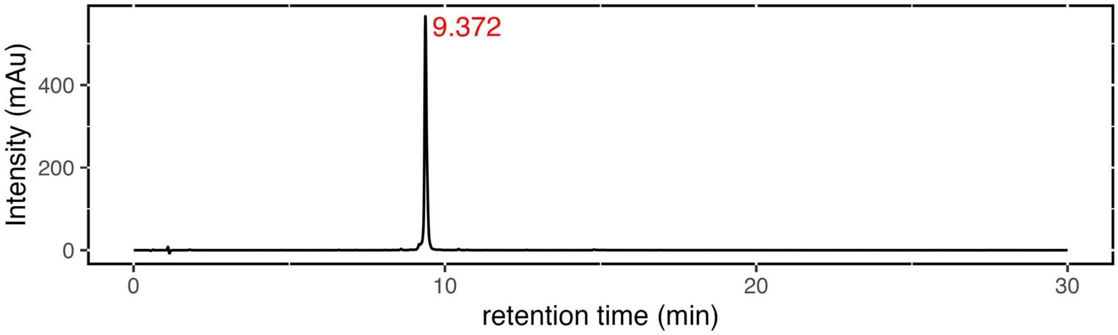
Supplemental Figure 41: HPLC analysis of the compound **M2**.



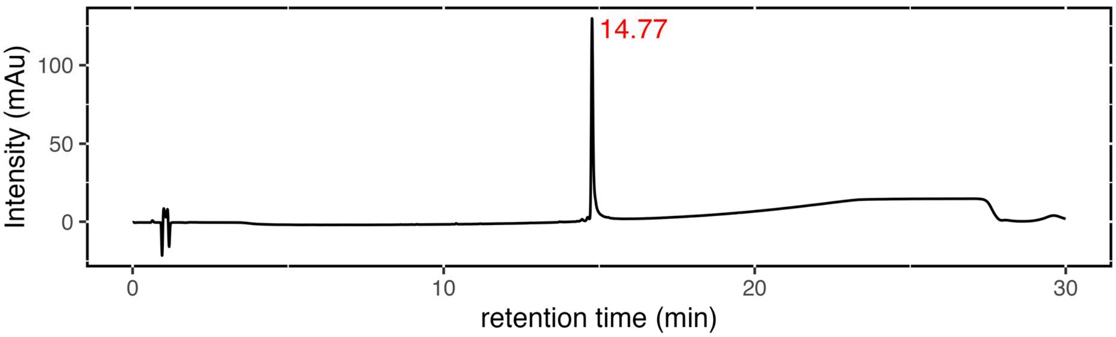
Supplemental Figure 42: HPLC analysis of the compound **P37**.



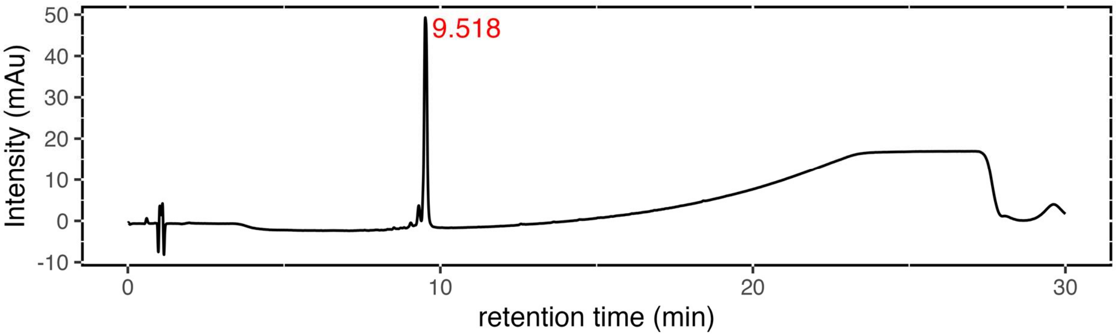
Supplemental Figure 43: HPLC analysis of the compound **P26-P37M**.



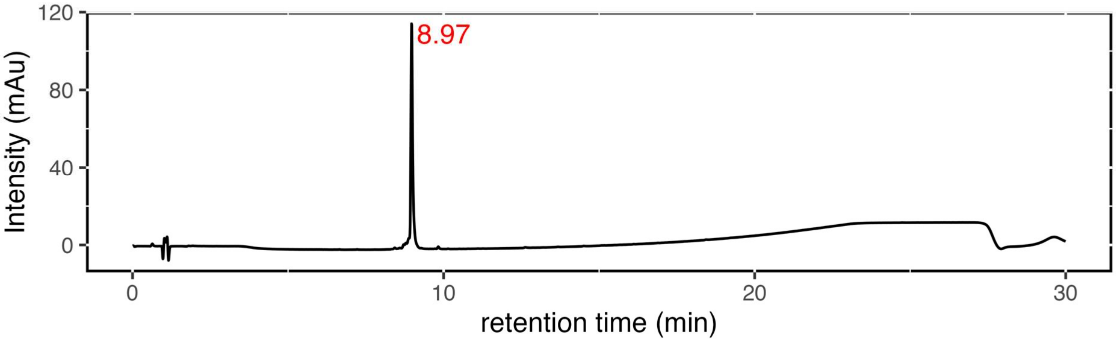
Supplemental Figure 44: HPLC analysis of the compound **P26-P37T**.



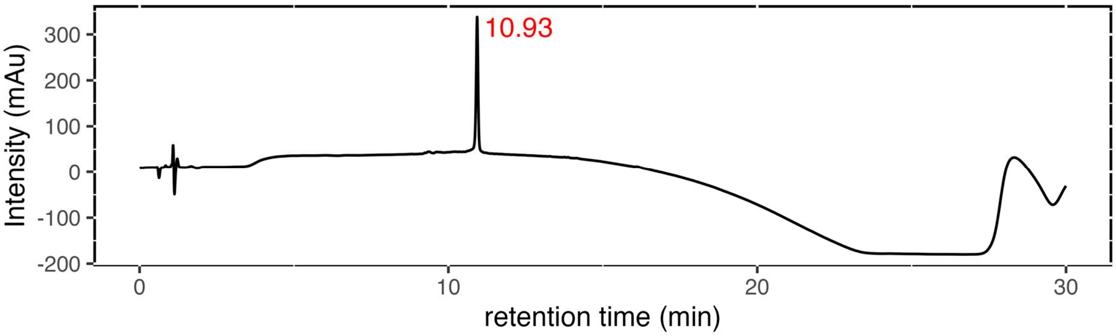
Supplemental Figure 45: HPLC analysis of the compound **P26-M12**.



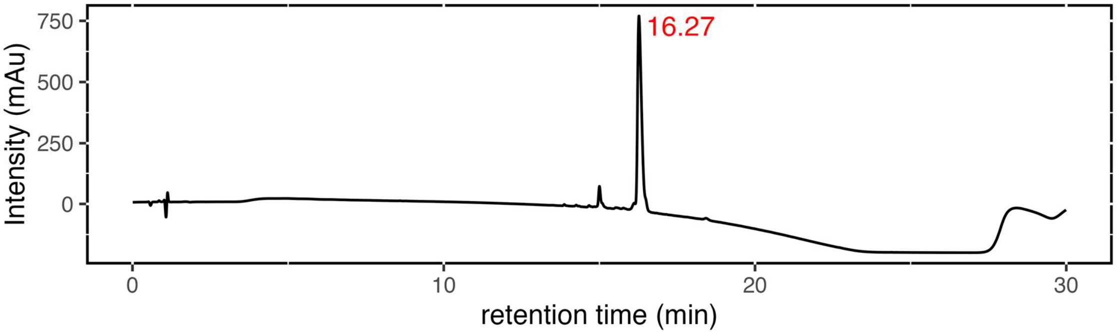
Supplemental Figure 46: HPLC analysis of the compound **P33-P37M**.



Supplemental Figure 47: HPLC analysis of the compound **P33-P37T**.

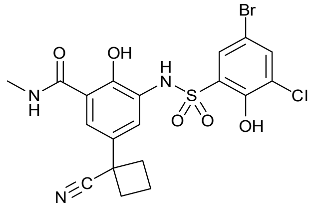


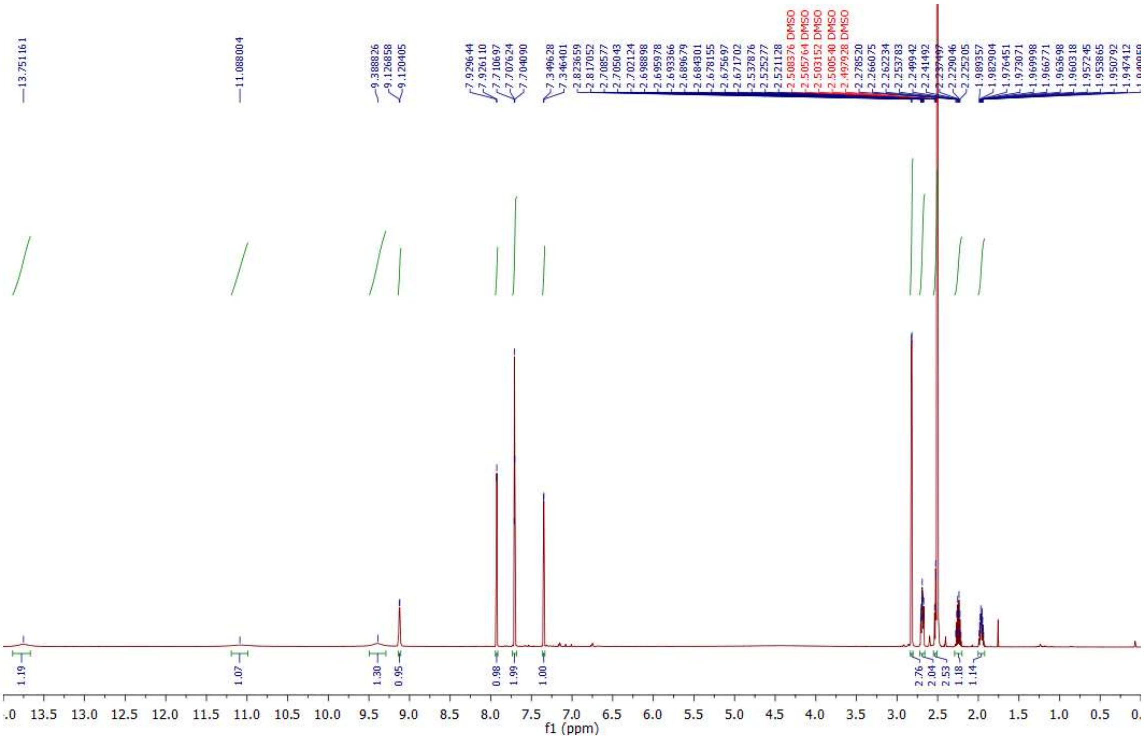
Supplemental Figure 48: HPLC analysis of the compound **P38** in the 210 nm channel.



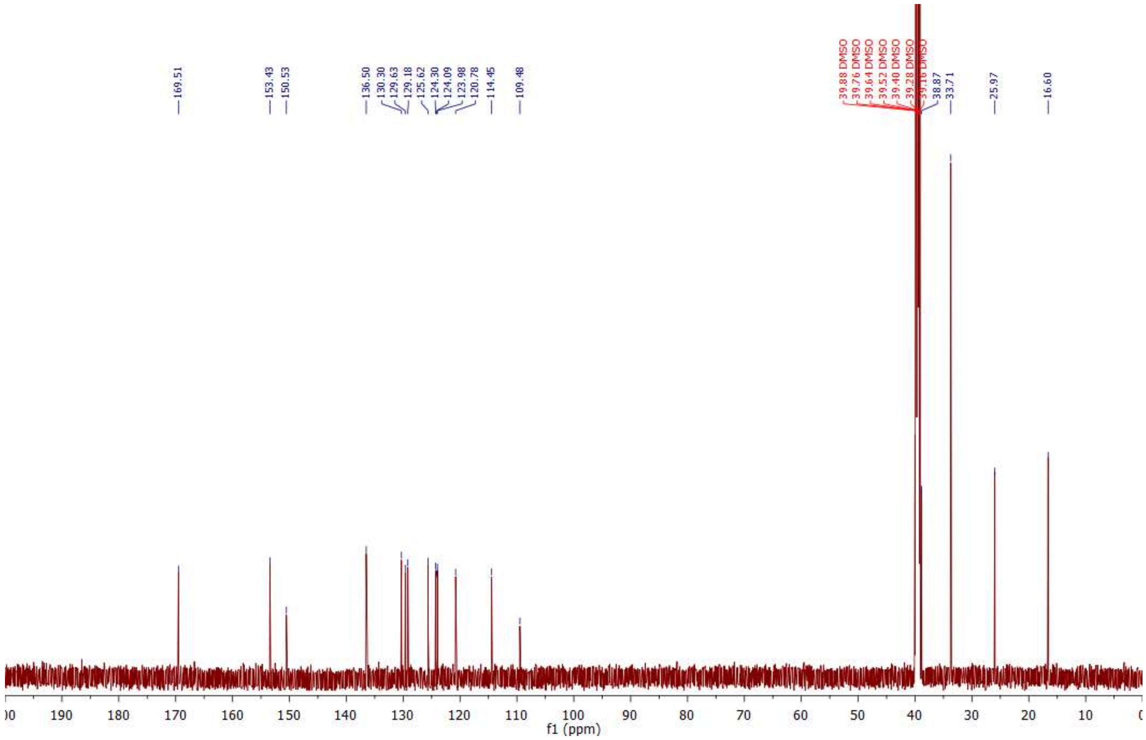
Supplemental Figure 49: HPLC analysis of the compound **P42** in the 210 nm channel.

6.1.2 Characterization of Compound 7k

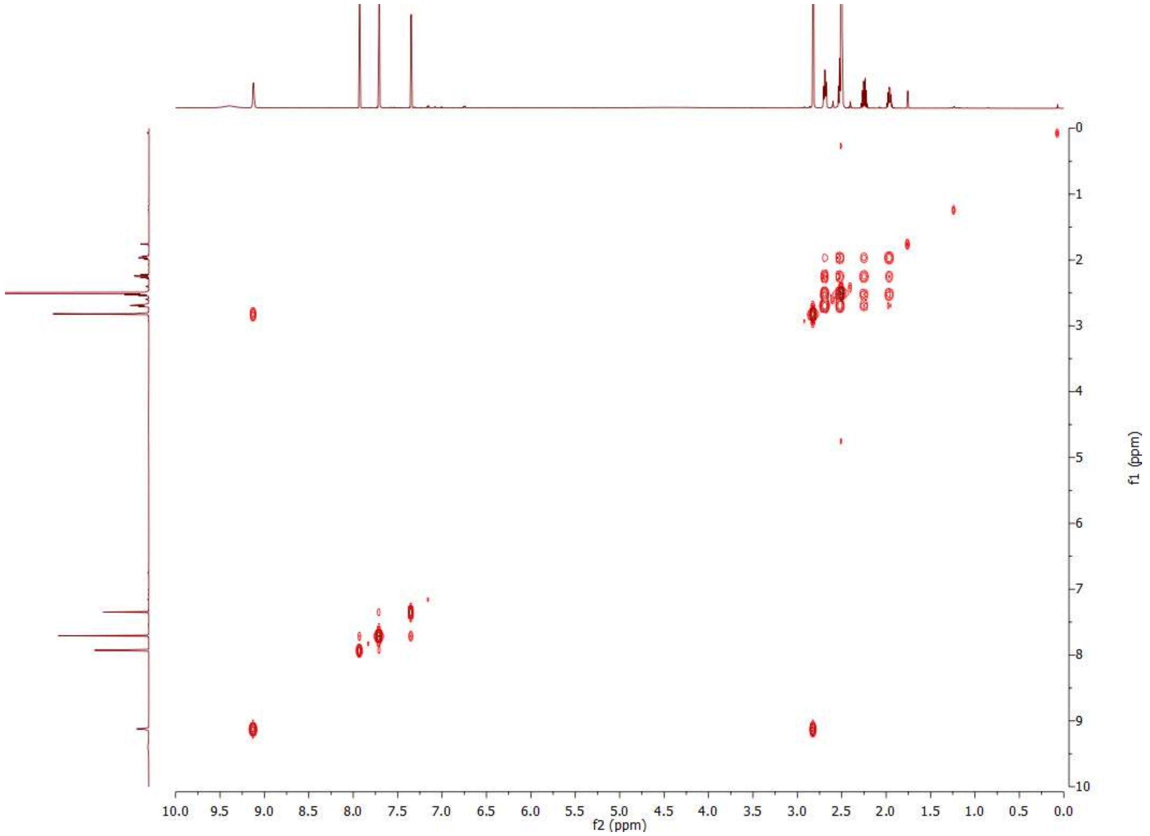




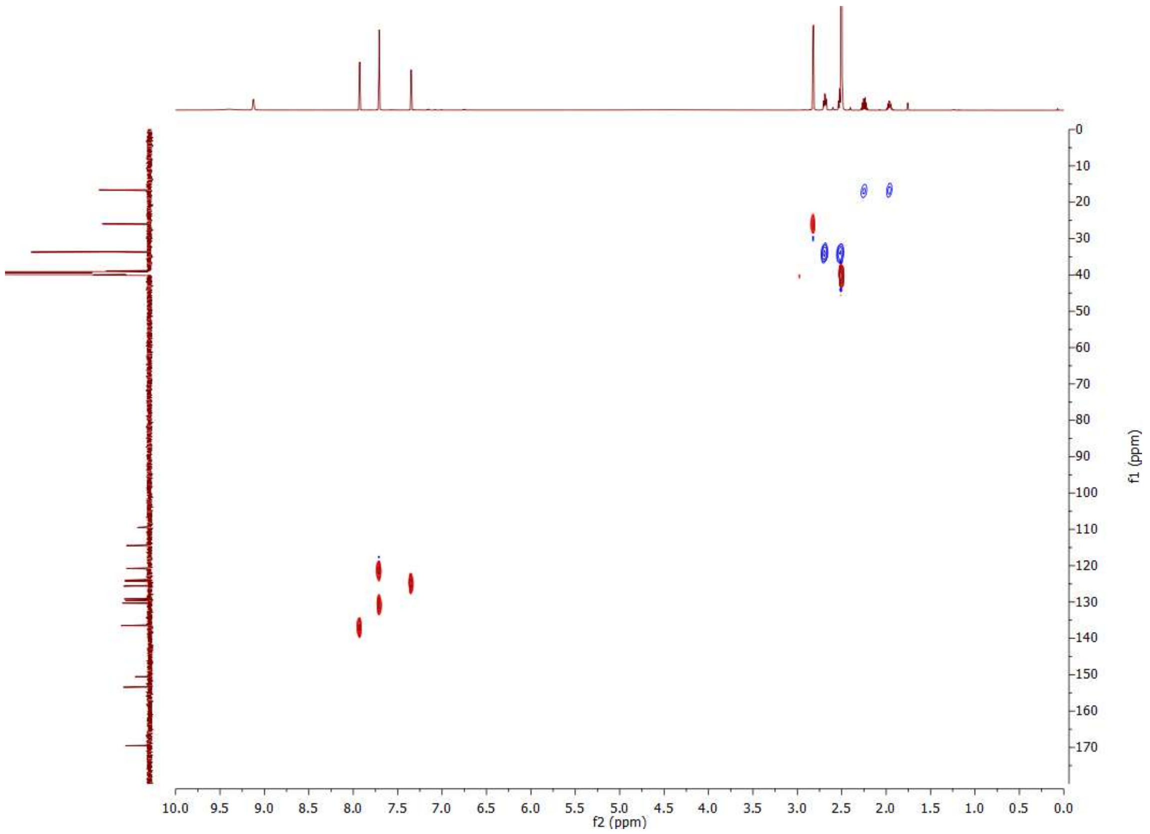
Supplemental Figure 50: 1H NMR spectrum of **7k** (700 MHz, DMSO-d_6).



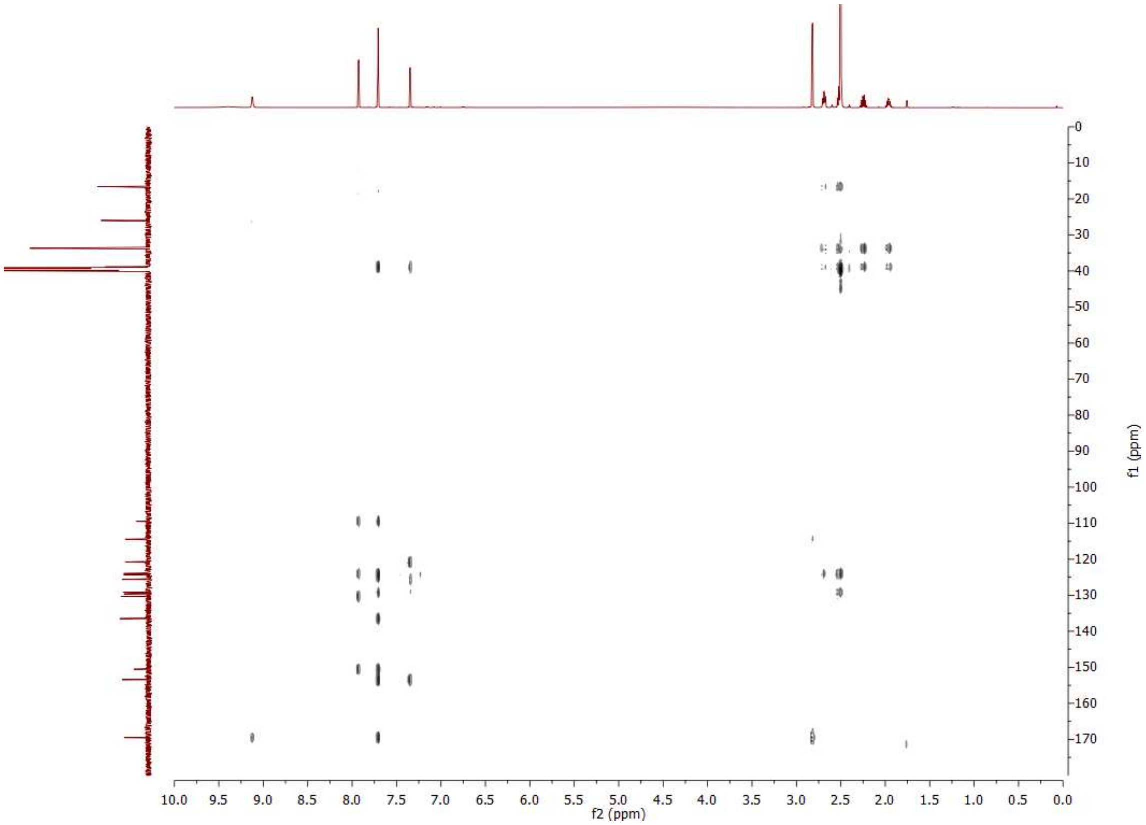
Supplemental Figure 51: 13C NMR spectrum of **7k** (176 MHz, DMSO-d_6).



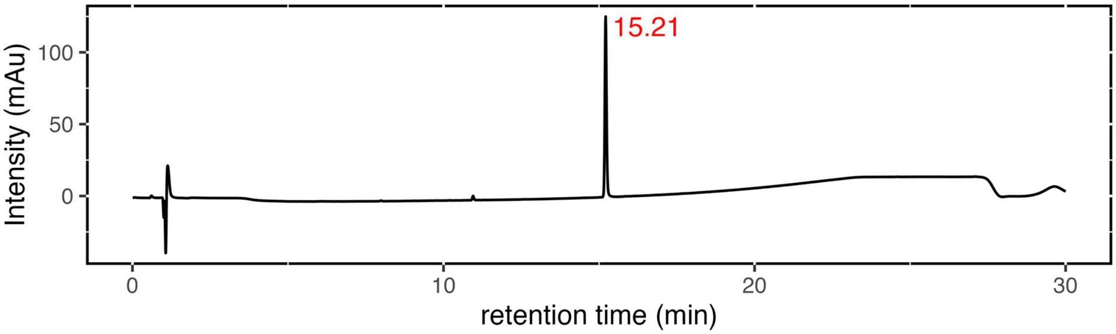
Supplemental Figure 52: COSY NMR spectrum of **7k** (**M2**).



Supplemental Figure 53: HSQC NMR spectrum of **7k** (**M2**).

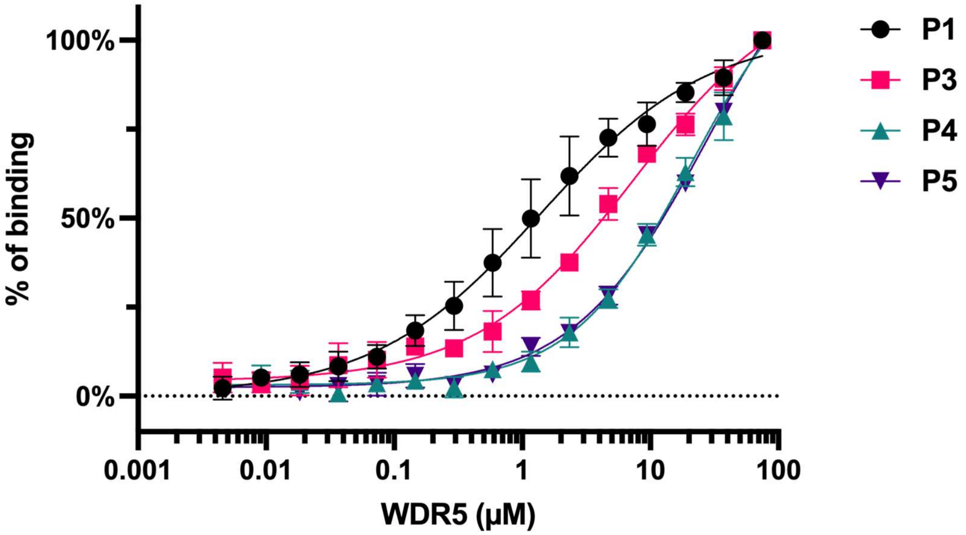


Supplemental Figure 54: HMBC NMR spectrum of **7k** (**M2**).

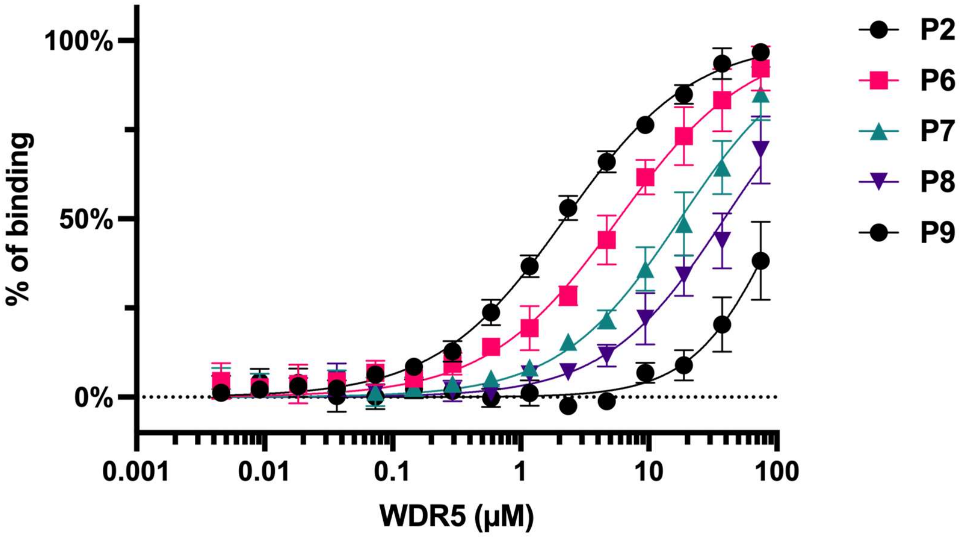


Supplemental Figure 55: HPLC analysis of the compound **7k** (**M2**).

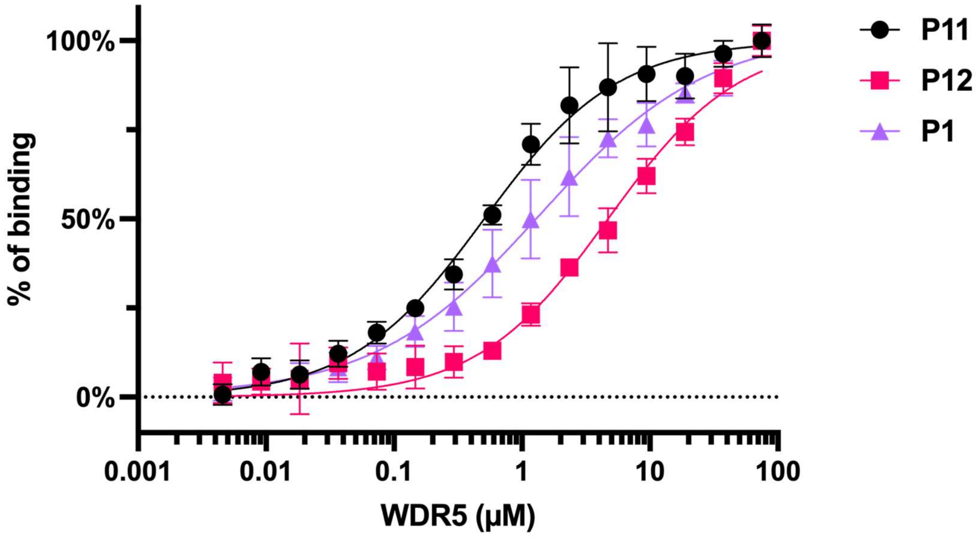
6.2 Results of Fluorescence Polarization



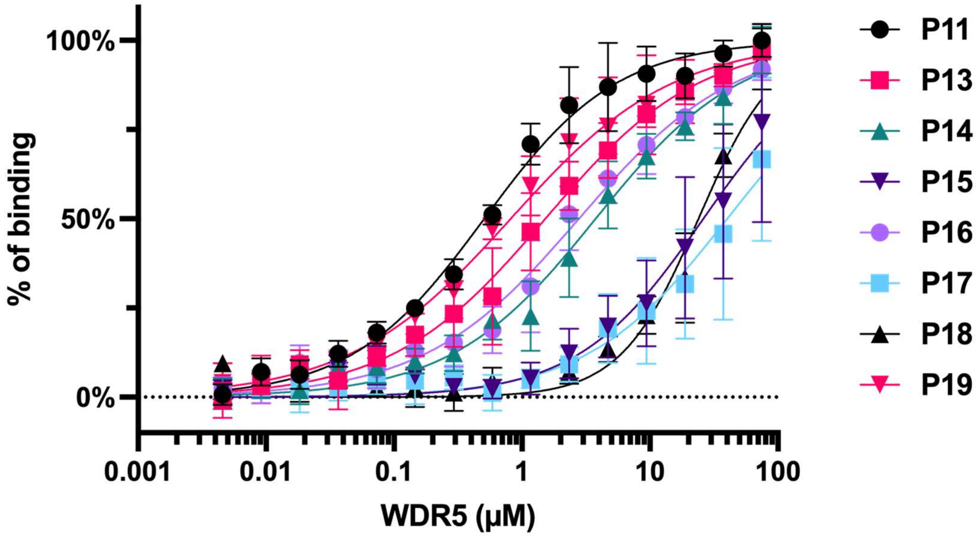
Supplemental Figure 56: Direct fluorescence polarization binding curve for peptide **P1** and **P3** – **P5** (Normalized). Assay performed in FP Buffer A.



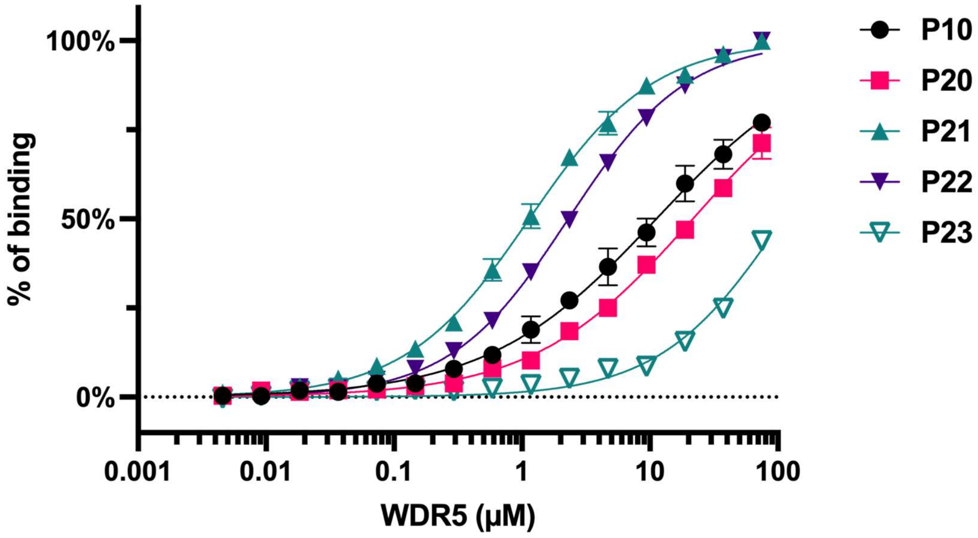
Supplemental Figure 57: Direct fluorescence polarization binding curve for peptide **P2** and **P6** – **P9** (Normalized). Assay performed in FP Buffer A.



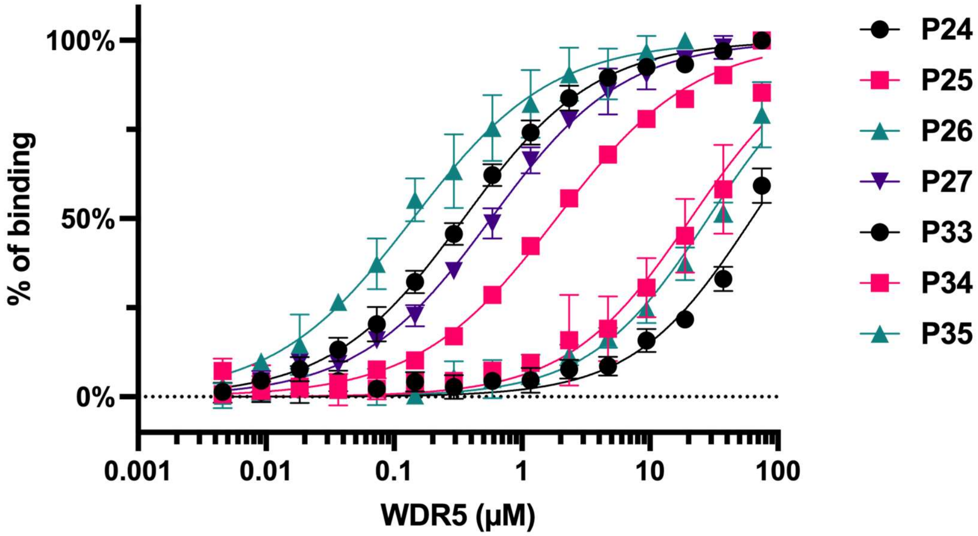
Supplemental Figure 58: Direct fluorescence polarization binding curve for peptide **P1, P11** and **P12** (Normalized). Assay performed in FP Buffer A.



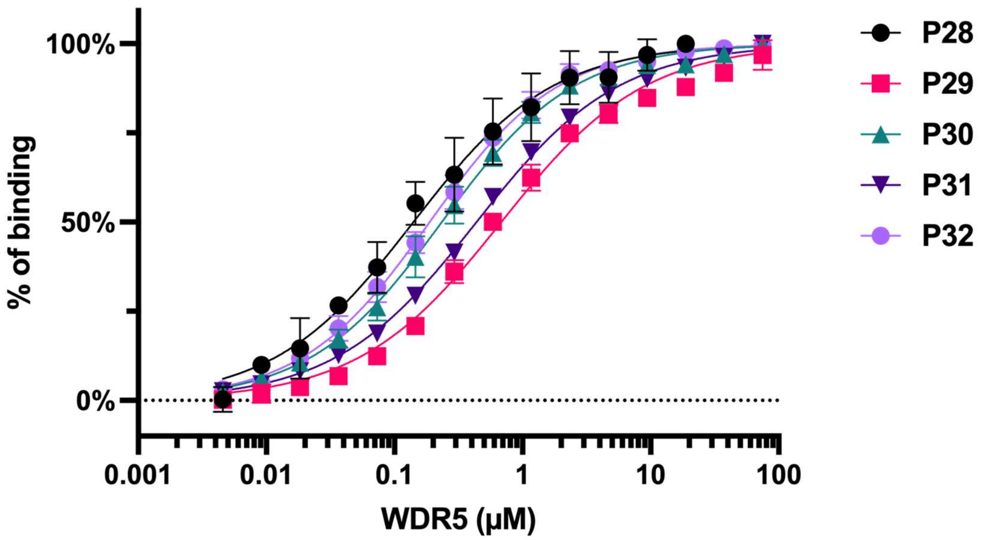
Supplemental Figure 59: Direct fluorescence polarization binding curve for peptide **P11** and **P13** – **P19** (Normalized). Assay performed in FP Buffer A.



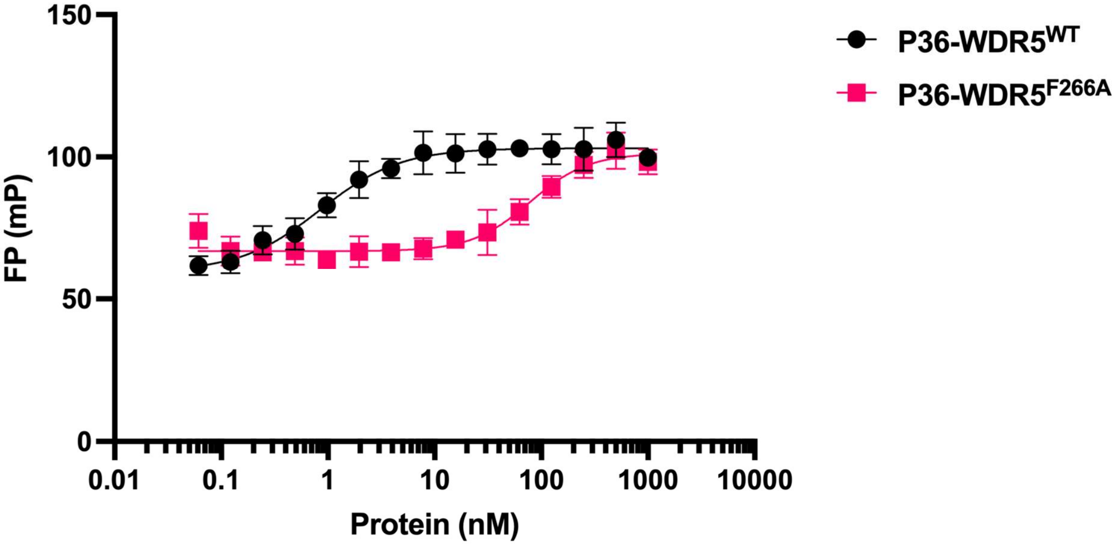
Supplemental Figure 60: Direct fluorescence polarization binding curve for peptide **P10** and **P20** – **P23** (Normalized). Assay performed in FP Buffer A.



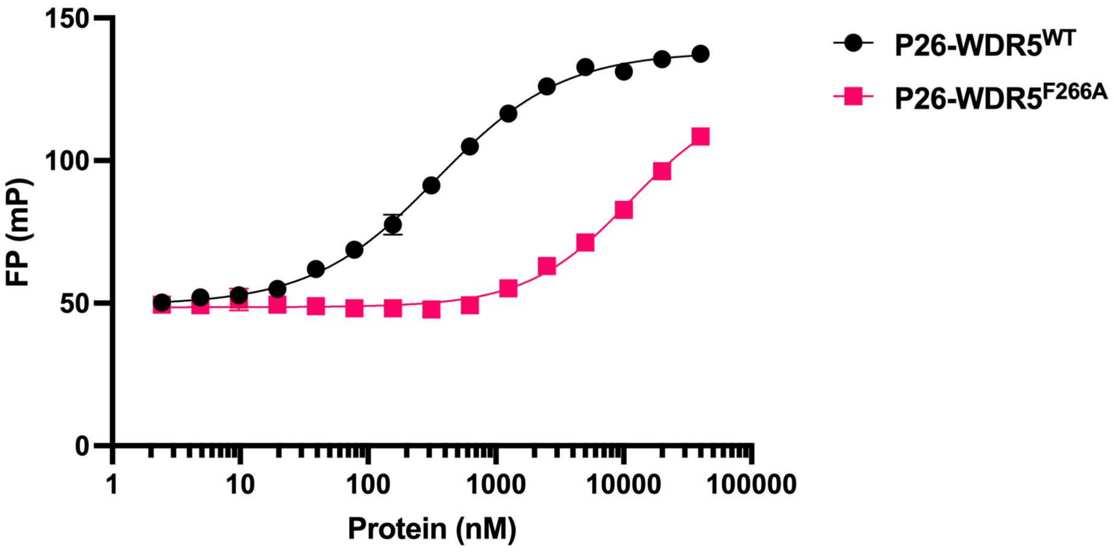
Supplemental Figure 61: Direct fluorescence polarization binding curve for peptide **P24** – **P35** (Normalized). Assay performed in FP Buffer A.



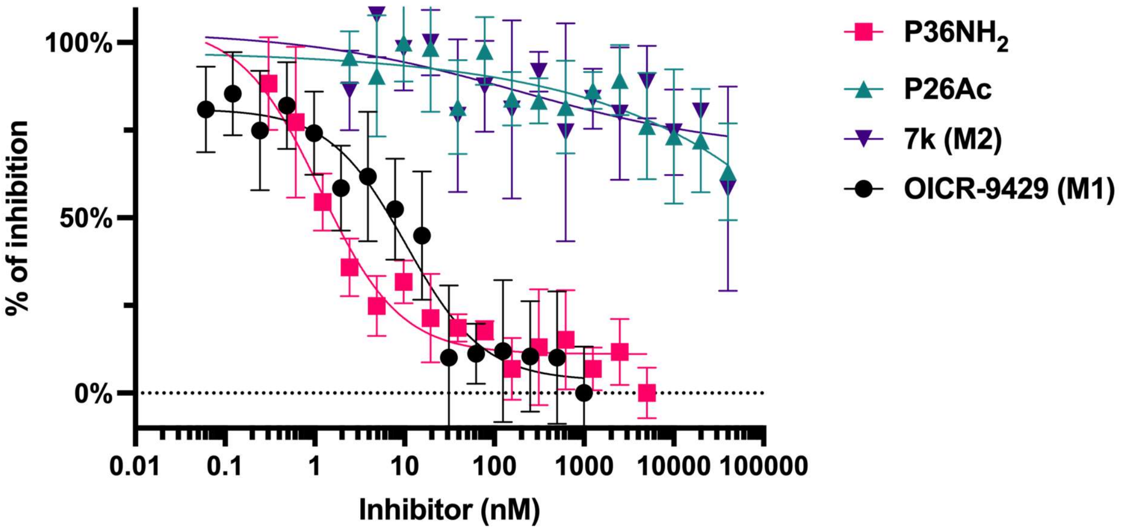
Supplemental Figure 62: Direct fluorescence polarization binding curve for peptide **P28** – **P32** (Normalized). Assay performed in FP Buffer A.



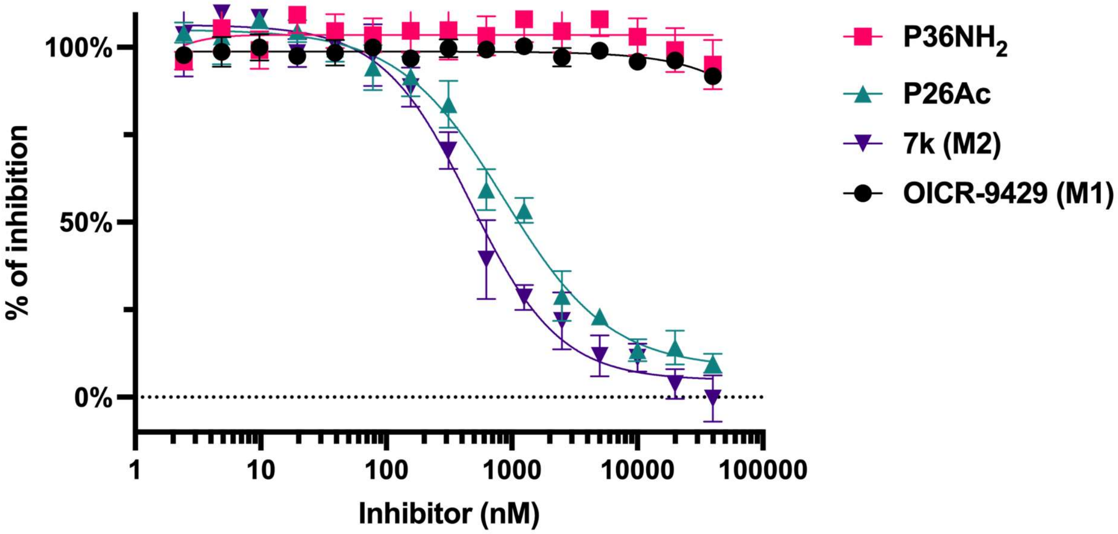
Supplemental Figure 63: Direct fluorescence polarization binding curve for peptide **P36** against WDR5^{WT} or WDR5^{F266A}. Assay performed in FP Buffer B.



Supplemental Figure 64: Direct fluorescence polarization binding curve for peptide **P26** against WDR5^{WT} or WDR5^{F266A}. Assay performed in FP Buffer B.



Supplemental Figure 65: Competitive fluorescence polarization binding curve for inhibitors against fix concentration of **P36**-WDR5^{WT} (Normalized). Assay performed in FP Buffer B.



Supplemental Figure 66: Competitive fluorescence polarization binding curve for inhibitors against fix concentration of **P26**-WDR5^{WT} (Normalized). Assay performed in FP Buffer B.

6.3 Buffer Recipes

Buffer A:

50 mM HEPES, 300 mM NaCl, 30 mM Imidazole, 0.1 mM PMSF, 1 mM TCEP, pH = 8.0

Buffer B:

50 mM HEPES, 300 mM NaCl, 500 mM Imidazole, 1 mM TCEP, pH = 8.0

Buffer C:

25 mM HEPES, 150 mM NaCl, pH = 7.0

FP Buffer A:

0.1 M $K_nH_{(3-n)}PO_4$, 25 mM KCl, and 0.01% Triton-X at pH 6.5[87]

FP Buffer B:

25 mM HEPES, 150 mM NaCl, 1% DMSO and 0.01% Triton-X, pH 7.0

FITC concentration Buffer:

0.1 M $NaHCO_{3(aq)}$, pH 9.0

ITC buffer:

20 mM Tris-HCl, 200 mM NaCl, pH = 7.0

RIP buffer[91]:

25 mM Tris-HCl, 150 mM KCl, 5 mM EDTA, 0.1% IGEPAL, 1 mg/mL BSA, 10 μg/mL yeast tRNA, 80 U/mL Recombinant RNasin Ribonuclease Inhibitor (freshly added, Promega, #N251A), pH 7.5

SEC Buffer:

50 mM HEPES, 300 mM NaCl, 1 mM TCEP, pH = 8.0

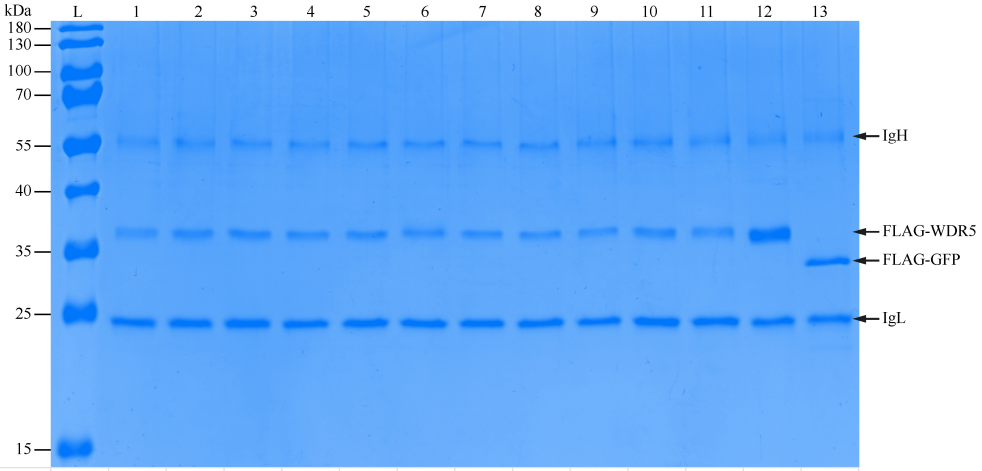
Coacervate Formatting Buffer:

10 mM NaH_2PO_4, 84 mM NaCl, pH 6.5

6.4 Primer List

*GAPDH*_Fw	AGCCACATCGCTCAGACAC[127]
*GAPDH*_Rev	GCCCAATACGACCAAATCC[127]
*U6*_Fw	CTCGCTTCGGCAGCACATATAC[128]
*U6*_Rev	GGAACGCTTCACGAATTTGC[128]
*HOTTIP*_Fw	TACCGGAATAGTGCTGGGGA[129]
*HOTTIP*_Rev	TGCGTGCTGCTCTGAGTTTA[129]
*HOXC13-AS*_Fw	GAAACTGCATTTCCTGGGGC[130]
*HOXC13-AS*_Rev	GGCTGGAGTCTTTGTCCTCC[130]
*FOXD3-AS1*_Fw	GCGAGGATGTGTGGCCAAT[130]
*FOXD3-AS1*_Rev	AGACAGGGATTGGGTTCCGT[130]
*TGF-β1*_Fw	CGTGGAGCTGTACCAGAAATAC[131]
*TGF-β1*_Rev	CACAACTCCGGTGACATCAA[131]

6.5 Gel Image



Nr	L	1	2	3	4	5	6	7	8	9	10	11	12	13
Label	Ladder	Blank	test 1 µM	test 5 µM	test 10 µM	neg 10 µM	Irrelevant*	Irrelevant*	win test	**M1** 10 µM	**7k (M2)** 10 µM	No RNA	F266A	GFP
Protein	-	WDR5^{WT}	WDR5^{WT}	WDR5^{WT}	WDR5^{WT}	WDR5^{WT}			WDR5^{WT}	WDR5^{WT}	WDR5^{WT}	WDR5^{WT}	WDR5^{F266A}	FLAG-GFP
RNA	-	+	+	+	+	+			+	+	+	-	+	
Inhibitors	-	-	**P26** (1 µM)	**P26** (5 µM)	**P26** (10 µM)	**P33** (10 µM)			**P36NH$_2$** (10 µM)	**OICR-9429** (10 µM)	**7k (M2)** (10 µM)	-	-	-

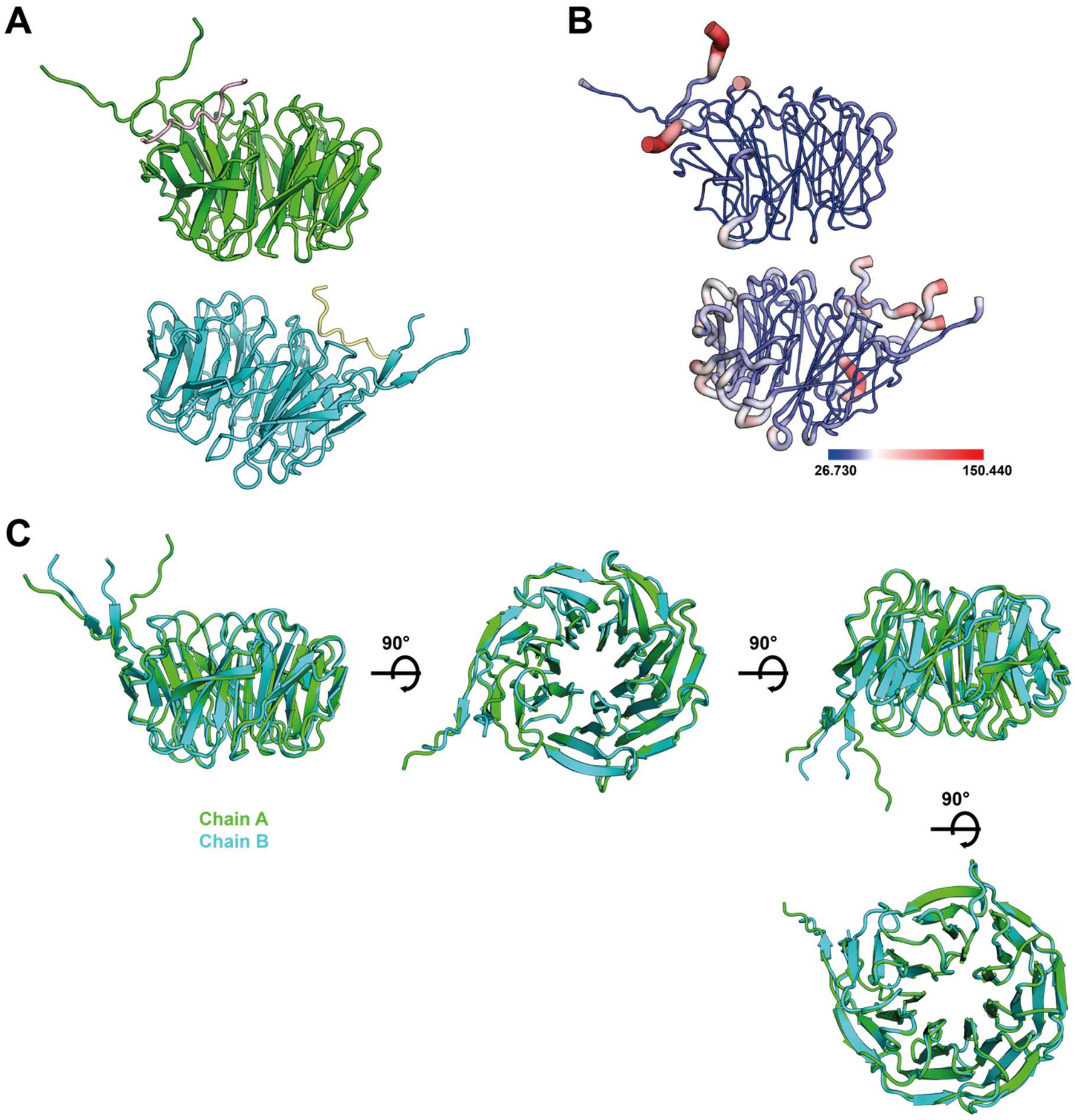
Supplemental Figure 67: Protein loading control of *iv*-RIP experiment. * Data was not discussed in this thesis.

6.6 Result of X-ray cocrystal structure

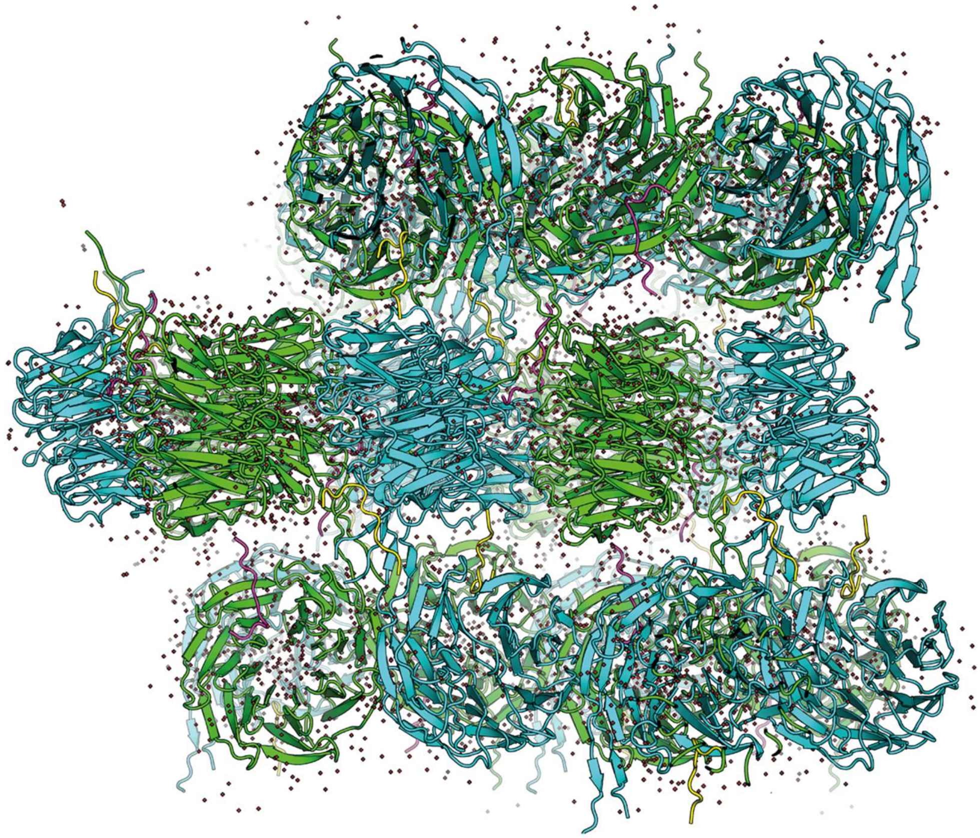
Supplemental Table 4. Data collection and refinement statistics of the WDR5-**P26Ac** crystal.

	WDR5-P26Ac
Data collection	
Space group	P 43 21 2
Cell dimensions	
a, *b*, *c* (Å)	82.1737 82.1737 201.707
α, β, γ (°)	90, 90 ,90
Resolution (Å)	42.98 - 1.843 (1.909 - 1.843)
R_{merge}	0.1688 (3.632)
I / s*I*	8.05 (0.26)
Completeness (%)	96.59 (66.37)
Redundancy (Multiplicity)	26.7 (27.4)
CC1/2	0.995 (0.361)
Refinement	
Resolution (Å)	1.84
No. Reflections	58517 (3960)
R_{work} / R_{free}	0.1975 / 0.2429
No. Atoms	5614
Protein	5179
Ligand/ion	0
Water	435
***B*-factors**	
Protein	56.22
Water	57.80
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.25
Ramachandran favored (%)	95.19
Ramachandran allowed (%)	4.65
Ramachandran outliers (%)	0.16
Rotamer outliers (%)	2.36
Clashscore	8.32

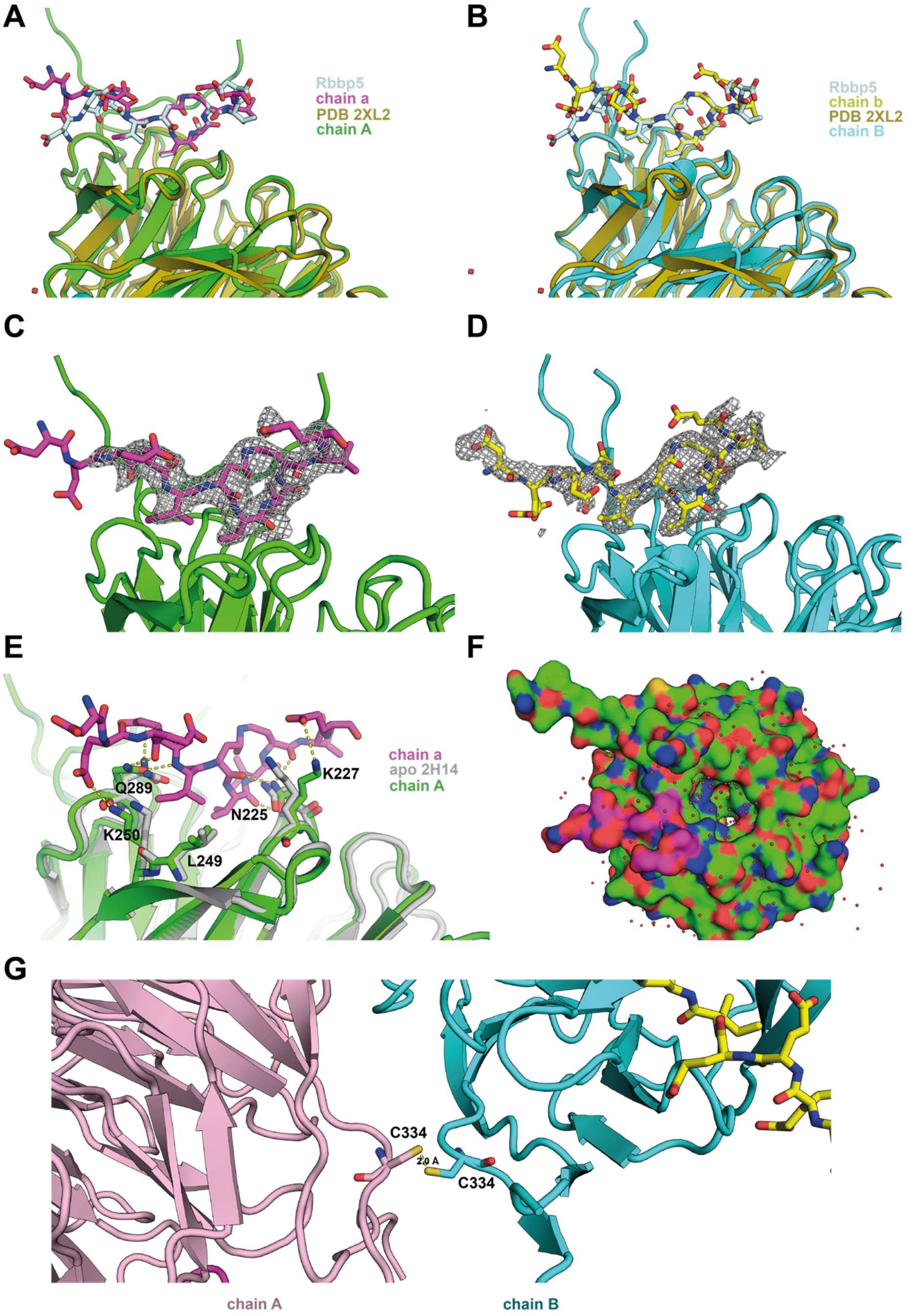
Statistics for the highest-resolution shell are shown in parentheses.



Supplemental Figure 68. **(A)** Asymmetric unit (ASU) of the WDR5-**P26Ac** co-crystal containing two protein-peptide dimers. **(B)** Depiction of the B-factors of both dimers A/a (top) and B/b (bottom) shows that the termini of the peptide ligands and chain B have higher B-factors than chain A which corresponds to the quality of electron density. **(C)** Overlay of the protein chains A and B indicate a high similarity of both chains except the termini.



Supplemental Figure 69. Symmetry mates in 20 Å distance show that the crystal packing of the structure is high.



Supplemental Figure 70. **(A/B)** The WBM binding site with **P26Ac** bound overlays well with a peptide with similar binding mode of Rbbp5 (PDB 2XL2) in the core regions of the peptide. **(C/D)** $2F_O$-F_C-maps of ligands a and b respectively indicate correct posing of the ligand in the WBM site. **(E)** The ligand **P26Ac** binds to the WBM interface without obstructing the apo-fold compared to a published apo-structure (PDB 2H14). **(F)** Like in previous structures, the channel in the WD-fold is filled with many structured waters. **(G)** Chain A forms a disulfide bridge with a symmetry related chain B.

6.7 Credit and Copyright licenses

Several contend in Chapter 2 and the corresponding figures in Chapter 6 have been published in (Chang. et al., 2023)[64] and licensed under Creative Commons Attribution 3.0 Unported Licence. The modification of individual reprinted figures is listed below:

[1] Figure 21 - Figure 23, Figure 25, Figure 26, Figure 31, Supplemental Figure 56 - Supplemental Figure 58, Supplemental Table 4 and Supplemental Figure 67 - Supplemental Figure 70 are reprinted from (Chang. et al., 2023) and the following figures are modified:

Figure 21: Add two white blocks behind the labels.
Figure 22: Add two white blocks behind the labels. Add blue arrow and label.
Figure 23: Change the compound number. Add a table to show the numbers.
Figure 25: Change the compound number. Add ITC data.
Figure 26: Remove C and D from the original pictures.
Figure 31: Change the compound number. Change the color scheme. Remove A – C from the original pictures.
Supplemental Figure 56 – 58: Change the compound number. Change the color scheme.
Supplemental Figure 67: Change the compound number. Add two more data.
Supplemental Table 4, Supplemental Figure 68 - Supplemental Figure 70: Change the compound number.

I am grateful for the permission to use the copyrighted material as follow:

[2] Figure 2 is reprinted from (Brander et al., 2017) with permission (5663310306189). Copyright © 2017 SETAC

[3] Figure 3 is reprinted from (Dimopoulos et al., 2014). It is licensed under CC BY-NC-SA 3.0. Copyright © 2014, K Dimopoulos, P Gimsing & K Grønbæk.

[4] Figure 4 is reprinted from (Biswas et al., 2018) with permission (5663320433427). Copyright © 2018 Elsevier B.V.

[5] Figure 6 is reprinted from (Lam et al., 2015). It is licensed under CC BY-NC-ND 4.0. Copyright © 2023 Elsevier Inc.

[6] Figure 7 is reprinted from (Flynn and Chang, 2014) with permission (5663321306895).

Copyright © 2014 Elsevier Inc.

[7] Figure 8 is reprinted from (Sweta et al., 2019) and licensed under CC BY 4.0. Copyright © 2019 Sweta, Dudnakova, Sudheer, Baker and Bhushan.

[8] Figure 9 is reprinted from (Wutz, 2007) with permission (5663330653007). Copyright © 2007 Elsevier Ltd.

[9] Figure 10 is reprinted from (Ramanathan et al., 2019) with permission (5663331035091). Copyright © 2019, Springer Nature America, Inc.

[10] Figure 11 is adapted from (Guarnaccia and Tansey, 2018) by combining Figure 4 and Figure 6. It is licensed under CC BY 4.0. Copyright © 2018 Alissa DuPuy Guarnaccia and William Patrick Tansey.

[11] Figure 18 is reprint from (Hendrickson et al. 2020). It is licensed under CC BY 4.0. Copyright © 2020 Olga D. Hendrickson, Nadezhda A. Taranova, Anatoly V. Zherdev, Boris B. Dzantiev, and Sergei A. Eremin

[12] Figure 19 is reprinted from (Prozeller et al., 2004). It is licensed under CC BY 3.0. Copyright © The Royal Society of Chemistry 2019.

[13] Table 1 is adapted from (Avdic et al., 2011) with permission (5663350406999). The original Figure 2E was rearranged without changing the content. Copyright © 2011 Elsevier Ltd.

I am grateful for the permission to use the copyrighted material from BioRender.com as follow:

[14] Figure 1 is created with BioRender.com with permission (TB25WAXTE6).

[15] Figure 5 is created with BioRender.com with permission (AO25WBAEP2).

[16] Figure 13 is created with BioRender.com with permission (TP25WL5ZYW).

[17] Figure 14 is created with BioRender.com with permission (CX25WL7WSQ).

[18] Figure 16 is created with BioRender.com with permission (UR25WVHC1N).

[19] Figure 17 is created with BioRender.com with permission (YP25X1HJR8).

[20] Figure 20 is created with BioRender.com with permission (ER261ZN6E7).

[21] Figure 27 is created with BioRender.com with permission (LS262ABVRQ).