



UNIVERSITÀ DEGLI STUDI DI SALERNO



UNIVERSITÀ DEGLI STUDI DI SALERNO
Dipartimento di Farmacia

PhD Program
in **Drug Discovery and Development**
XXXII Cycle — Academic Year 2019/2020

PhD Thesis in

***Development of a versatile screening platform
for the identification of small-molecule ligands
for reader proteins of the post-translational
methylation***

Candidate

Vincenzo Pisapia

Supervisor

Prof. *Gianluca Sbardella*

PhD Program Coordinator: Prof. Dr. *Gianluca Sbardella*

ABSTRACT	V
1. INTRODUCTION	1
1.1 Epigenetic proteins.....	1
1.2 Epigenetic readers.....	2
1.2.1 Bromodomains.....	3
1.2.2 Royal Superfamily.....	4
1.2.2.1 Tudor domains.....	5
1.2.2.1.1 Tudor domains of PHF20.....	5
1.2.2.1.2 Tudor domains of Spindlin1.....	8
1.2.2.2 Chromodomains.....	9
1.2.2.2.1 Chromodomain of MRG15.....	10
1.3 Chemical probes for epigenetic readers.....	12
1.4 Epigenetics and drug discovery.....	13
2. AIM OF THE WORK	15
2.1 Importance to use a screening platform.....	15
2.2 Aim of the work.....	16
2.3 Methods.....	18
2.3.1. nanoDSF.....	18
2.3.2 Surface Plasmon Resonance.....	19
2.3.3 Microscale Thermophoresis.....	21
2.3.4 AlphaLISA.....	23
3. RESULTS AND DISCUSSION	26
3.1. Tudor domains of PHF20.....	26
3.1.1. Background.....	26
3.1.1.1. Expression and purification of GST-Tudor1-2.....	29
3.1.1.2. GST cleavage and purification of Tudor1-2 domains.....	30
3.1.1.3. Expression and purification of His –Tudor2 sequence.....	31
3.1.1.4. Expression and purification of His –Tudor2 mutate sequence.....	32
3.1.2. Protein stability.....	33
3.1.3. nanoDSF as primary screening.....	34
3.1.4. Microscale Thermophoresis assay (MST).....	36
3.1.5. Surface plasmon resonance.....	38
3.1.6. AlphaLISA.....	43
3.1.7. Crosschecking.....	49

3.2	Spindlin1	50
3.2.1	Background	50
3.2.2	Expression and purification of Spindlin1	51
3.2.3	Design, optimization, and validation of MST assay for Spindlin1	52
3.2.4	Assay validation.....	55
3.2.5	Small library screening	55
3.3	MRG15.....	57
3.3.1	Background	57
3.3.2	Expression and purification of MRG15	57
3.3.3	Design, optimization, and validation of MST assay for MRG15.....	58
3.3.3.1	Protein labeling	58
3.3.3.2	Buffer optimization	60
3.3.3.3	Assay validation.....	61
3.3.4	Small library screening.....	62
3.4	GST tag interference check	64
4.	CONCLUSIONS.....	67
5.	MATERIALS AND METHODS.....	70
5.1	PHF20	70
5.1.1	Expression and purification of GST Tudor1-2.....	70
5.1.2	GST cleavage and purification of Tudor1-2.....	70
5.1.3	Expression and purification of His-Tudor2 wild type.....	71
5.1.4	Expression and purification of His-Tudor2 mutate	72
5.1.5	nanoDSF experiments	72
5.1.6	Microscale thermophoresis.....	73
5.1.6.1	GST-Tudor1-2 labeling.....	73
5.1.6.2	Tudor1-2 labeling.....	73
5.1.6.3	Tudor2 mutate labeling	74
5.1.6.4	Assay development for MST screening	74
5.1.7	Surface plasmon resonance	75
5.1.8	AlphaLISA assay optimization	76
5.2	Spindlin1	76
5.2.1	Protein expression and purification.....	76
5.2.2	Spindlin1 labeling for MST measurements.....	77
5.2.3	Protein thermal stability measurements	78

5.2.4	Assay development for MST screening	78
5.3	MRG15.....	79
5.3.1	Protein expression and purification.....	79
5.3.2	Protein labeling for MST measurements.....	79
5.3.3	Protein thermal stability measurements	80
5.3.4	Assay development for MST screening	80

ABSTRACT

Covalent Post Translational Modifications (PTMs) of histone proteins are introduced and removed by the ‘writer’ and ‘eraser’ enzymes, respectively. These modifications are recognized by highly specific binding domains, the so-called ‘reader’ proteins, which are thought to mediate epigenetic signalling. Misregulation of these proteins often leads to abnormal gene expression patterns that are more frequently linked to human diseases.

In comparison to writers and erasers, readers have been less intensively pursued so far as therapeutic targets, especially with regard to methylation. To date, different biochemical and biophysical techniques are reported to assess the binding of new putative modulators of methyl readers. Nevertheless, given the intrinsic limitations each technique encompasses, reliable data can be obtained using a combination of methods.

Thus, this project is focused on the development of a versatile screening platform for the identification of small-molecule ligands of methylation reader proteins. Among different reader domains, the attention was focused on the Tudor domains of PHF20, Tudor domains of Spindlin1 and the Chromodomain of MRG15. After optimizing the conditions of expression and purification for these proteins, the screening platform was applied to all proteins in order to screen a library of compounds synthesized in Epigenetic Medicinal Chemistry Laboratory (EMCL).

This platform is composed by biophysical techniques such as nanoDSF, MST and SPR while the only biochemical technique used was the AlphaLISA. Comparing the results obtained from each technique it was possible to identify true hits.

1. INTRODUCTION

1.1 Epigenetic proteins

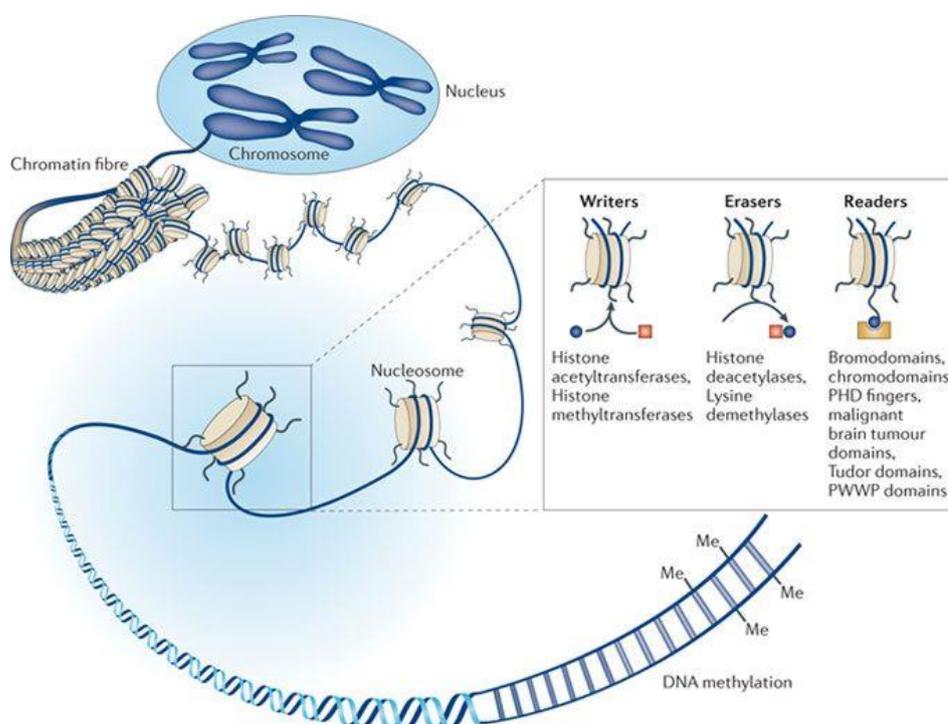
The definition of Epigenetics has been changed several times since the first time it was used by Conrad H. Waddington. Nowadays, the more exhaustive definition was given by Adrian Bird, which defines Epigenetics as “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity state.”¹ In other words, it is the study of all the heritable changes that alter gene activity without changing the DNA sequence but modifying the chromatin structure. Chromatin is a highly organized complex of DNA and proteins and its fundamental unit is the nucleosome. This latter consists of an octamer of histone proteins (dimers of H2A, H2B, H3, and H4) wrapped by DNA and the histone tails that protrude from the core are subjected to various chemical modifications. Also called post-translational modifications (PTMs), these changes along with the direct modification of the DNA, being reversible, determine the continuous transition between the opened and closed state of chromatin, and depending on their nature, they allow the genes to be expressed or unexpressed.

These PTMs are catalyzed by enzymes that are well known as ‘writers’ and result in the addition to DNA or histones of entities ranging in size from a single methyl group to proteins such as ubiquitin. Such molecular changes not only directly influence the affinity between DNA and histone proteins but also recruit partner macromolecules such as non-coding RNAs (ncRNAs) and chromatin remodelers. The binding interactions are controlled through so-called

‘reader’ domains that recognize specific features within the chemically modified nucleic acids and proteins.

Finally, to ensure the process is reversible, a series of ‘eraser’ enzymes catalyze the removal of the written information making all the epigenetic modifications highly dynamic.²

Playing a crucial role in the maintenance of cellular homeostasis and cell differentiation, these proteins when expressed in uncontrolled manner often leads to abnormal gene expression patterns that are more frequently linked to human diseases.³



Nature Reviews | Drug Discovery

Figure 1.1 Schematic representation of main epigenetic modifications

1.2 Epigenetic readers

The “reader” domains are considered the interpreters of the epigenetic code. They carry out different functions, for example, they can recruit enzymes or transcription factors with the aim to modify chromatin, or they can possess an intrinsic chromatin-modifying activity themselves,

also they can stabilize chromatin complexes. Some readers even contain an assembly of different binding domains in one protein and they are able to interact with more than one PTM at the same time.⁴

The reader domains of histone PTMs histone can be split into different categories depending on the mark recognized. There are domains able to recognize the acetylation (bromodomains) and methylation (PHD, WD40, Ankyrin, BAH and the Royal superfamily that is composed by Tudor, Chromo, PWWP, plant agenet, MBT domains).

1.2.1 Bromodomains

Identified for the first time in the early 1990s in *Drosophila melanogaster*, bromodomains are a family of conserved motifs capable to bind the acetylated lysines on the histone tails.

Founded in 46 human proteins, these domains can be classified into eight subfamilies.

Among these subfamilies, certainly, the most investigated is the bromodomain and extra terminal (BET) domain, which is made up of BRD2, BRD3, BRD4, and BRDT. In turn, they share the presence of two bromodomains in tandem (BD1 and BD2) in the N-terminal and a C-terminal extra-terminal (ET) domain.⁵ Despite their function is poorly understood, these proteins seem to be crucial in the regulation of transcription and chromatin remodeling. Furthermore, they have been implicated in human diseases that include neurological and inflammatory diseases, as well as cancer.⁶

In general, the acetylation is an epigenetic change that occurs on the ϵ -ammonium group of lysine residues that lead to a loss of positive charge under physiological conditions. As a consequence, the interaction between the negatively charged DNA and histone proteins is weaker, increasing the accessibility of chromatin and promoting the gene expression.⁷ The affinity of bromodomains for isolated acetylated peptides is surprisingly low (K_D values from 10 μ M to low millimolar range), thus many bromodomains could bind small molecule

compounds with higher affinity. Indeed, for this reason, bromodomains represent one of the great successes in the discovery of chemical probes.⁸

1.2.2 Royal Superfamily

The Royal Superfamily is composed of tandem Tudor domains (TTD), chromodomains, malignant brain tumor (MBT) domains, Pro-Trp-Trp-Pro (PWWP) domains, and plant agenet domains and it is considered the largest superfamily of methyllysine readers,

Differently from the acetylation, methylation is an epigenetic change that leaves unaltered the charge of the amino acid side chain but it modifies the size and lipophilicity of substrates. Methylation can occur on the lysine and arginine residues. The lysine can be mono- di- and trimethylated, whereas arginines can be mono, symmetrically and or asymmetrically dimethylated. With the incremental addition of methyl groups, the hydrophobicity and the cation radius of the lysine methylammonium group increases, and its ability to donate hydrogen bonds concomitantly decreases. Depending on the substrate and the degree of methylation, histone methylation can be correlated with gene activation or silencing.⁹

All these domains share a highly conserved motif characterized by a barrel-like structure made up of a 3-5 antiparallel sheets. The loop regions, which originate from the β sheets, constitute the aromatic binding cage built up of 2 to 4 aromatic amino acid residues able to host the methylated substrate. In many complexes the aromatic rings are positioned perpendicular to each other, surrounding the fully extended side chain of the methylated lysine.

The aromatic rings and the methyl ammonium group form a complex that is driven by cation- π interactions as well as hydrophobic and van der Waals contacts. Specificity for a particular methylated lysine is imparted by interaction with surrounding residues.¹⁰

Considering that this work is focused on three reader proteins (PHF20, SPIN1, MRG15) containing two different domains of the Royal Superfamily (Tudor and Chromo- domains), in the following paragraphs there will be a deeper description of these domains and then a detailed discussion about the three specific proteins.

1.2.2.1 Tudor domains

The Tudor domain was named after the *Drosophila tudor (tud)* gene identified in a screen for maternal-effect recessive lethality or sterility. Found in 41 different proteins, Tudor domains can occur as single domain alone, multiple tandem Tudor domain or one or more Tudor domains in conjunction with other types of domains.

Considering their capability to interact with methylated residues of arginine and lysine, Tudor containing proteins can be categorized into two groups. Regarding the lysine methylated Tudor domains, they interact with both, higher and lower lysine methylation states and they are involved in histone modifications and chromatin remodeling. On the other side, arginine methylated Tudor domains are involved in RNA regulation, alternative splicing, small RNA pathways, and germ cell development. Among all the Tudor-containing proteins identified, 53BP1 and JMJD2A were the first ones that were shown to harbor histone methylation-binding capacities *via* Tudor. In addition, recent works also demonstrated that JMJD2A and 53BP1 binding to dimethylated histone H4 lysine 20 (H4K20me2) is critical for regulation of cellular response to DNA damage.¹¹⁻¹²

1.2.2.1.1 Tudor domains of PHF20

Plant homeodomain finger protein 20 (PHF20/GLEA2/HCA58) was initially identified in glioblastoma and adenocarcinoma patients as an immunogenic antigen.¹³

Previous studies reported that PHF20 is abundantly expressed in various cancers like hepatocellular carcinoma, medulloblastoma, and also in Non-Small Cell Lung Cancer (NSCLC), suggesting that this protein could have a role in cancer development.¹⁴ Further studies showed that PHF20 is expressed in 48.6% of glioblastomas and the patients that express this protein showed a survival increase.¹⁵ Recent studies have revealed that PHF20 is a component of the 'male absent on the first' (MOF), a lysine acetyltransferase, and together with the isoform 1 of O-linked- β -N-acetylglucosamine transferase constitute the Nonspecific Lethal complex (NSL). Moreover, PHF20-deficient mice demonstrated defective transcriptional activation of H4K16 target genes, suggesting that PHF20 is a potent transcriptional activator by an epigenetic-based mechanism.¹³

PHF20 protein has been isolated from MLL methyltransferase complexes involved in the deposition of methyl marks along histone tails. The ability of PHF20-Tud2 to recognize dimethylated histone substrates may be an important contributor to the regulation and/or assembly of MLL complexes.¹⁶

Interestingly, a previous report showed that a Tudor domain in PHF20 can also associate with p53 through dimethylated Lys residues, leading to the stabilization of p53, and then suggesting a tumor-suppressor activity.¹³ It was also reported that PHF20 can promote the transcription of NF- κ B, indeed interacting with p65 in a methylated-dependent manner, PHF20 contribute to the phosphorylation of p65 and interrupting, subsequently, the recruitment of the phosphatase PP2A.¹⁷ Nowadays, it is not clear how PHF20 is involved in the growth and proliferation of the tumor. These contradictory theses suggest that PHF20 can have a dual role, oncogenic and oncosuppressor, depending on the tissue involved and the context.¹⁸

Structurally, PHF20 is a multidomain protein that contains on the C-terminal region a PHD domain, in the center a C2H2 Zinc finger and a tandem Tudor domain on the N-terminal portion (Figure 1.2).

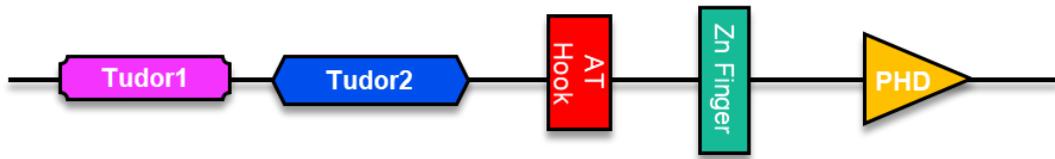


Figure 1.2 *Domain architecture of PHF20*

The second Tudor domain is capable to bind methylated residues on histone tails, but regarding the first, its function is still not clear. The second Tudor domain possesses the canonical structure of Tudor domains, which is constituted by five antiparallel β -sheets positioned to create a barrel structure. The aromatic cage is built up Trp97, Tyr103, Phe120, Asp122, and V124 (Figure 1.3). Comparing the structures of the binding cage of both Tudor domains, Tudor1 domain possesses an atypical binding cage because the presence of the residue Trp50 toward the center of the binding site, prevents the recognition of methylated residues by this domain.

In addition, another important feature of Tudor2 domain is the capability to homodimerize, indeed the presence of two cysteines (Cys96 and Cys100) allows the formation of two disulfide bonds.¹⁶

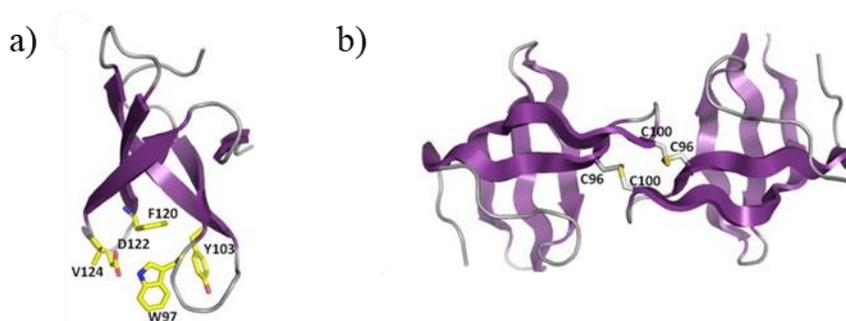


Figure 1.3 *a) Aromatic cage structure of the Tudor2 domain, b) structure of homodimer constituted by two Tudor domains.*

1.2.2.1.2 *Tudor domains of Spindlin1*

Spindlin1 (SPIN1) was initially described as an abundant maternal transcript deposited in the unfertilized mouse egg. The name derived from its association and co-migration with the meiotic spindle in the first meiotic cell cycle caused by periodical phosphorylation during the meiosis. Together with Y-linked spermiogenesis specific transcript (SSTY), Spindlin1 forms a new SPIN/SSTY gene family implicated in cell cycle regulation during gametogenesis and the transition between gamete and embryo. These two proteins share more than 50% identity in amino acid sequence.¹⁹

Spindlin1 is abundantly expressed in human early embryonic tissues, including ovary and kidney. Furthermore, SPIN1 was reported to be highly expressed in several types of tumors, like ovarian tumors, Non-small cell lung cancers, and some hepatic carcinoma.¹⁹

It was observed the ectopic expression of Spindlin1 in NIH3T3 induce a cell-cycle delay in metaphase and causes chromosome instability. Furthermore, the transfected cells, which are able to grow like cancer cells in nude mice, showed a complete morphological change.²⁰

Spindlin1 was also found to function as an activator of the WNT/TCF-4 signaling pathway that regulates cell fate in a large number of developmental processes and differentiation during embryogenesis.²⁰ All these aspects suggest and confirm the importance of this protein in the tumor breakout.

Regarding the structure, Spindlin1 is a histone code reader composed of three Tudor-like domains. The second Tudor domain possesses an aromatic cage that is built up to Phe141, Trp151, Tyr170, and Tyr177.²¹ It is able to bind with high affinity the histone H3 trimethylated at lysine 4 (H3K4me3), a chromatin mark typically located at promoters and associated with active or poised genes. This association was recently shown to be further enhanced by the presence of asymmetrically dimethylated arginine 8 (H3R8me2a), a mark implicated in the triggering of organizer gene expression, through the first Tudor domain (Figure 1.4).^{22 23}

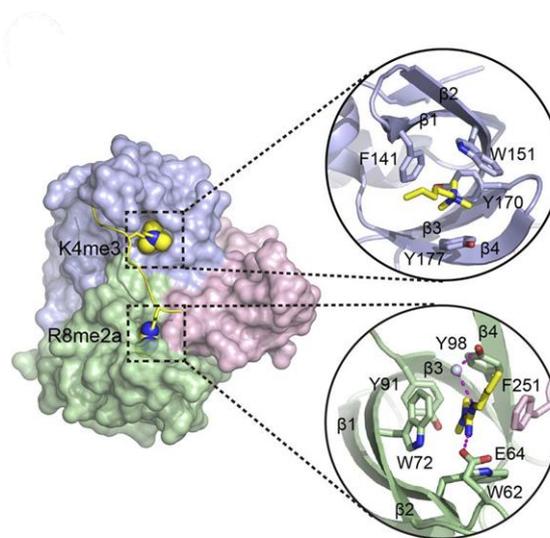


Figure 1.4 *Encapsulation of K4me3 and R8me2a in Spindlin1 pockets 2 and 1. Spindlin1 is shown in surface mode with Tudor domain 1, 2, and 3 colored green, blue, and pink, respectively.*

1.2.2.2 Chromodomains

The name of this family of readers is an acronym for Chromatin Organization Modifier and derived from a significant structural homology with the protein Polycomb and Heterochromatin protein 1 (HP1) implicated in gene silencing by a chromatin-based mechanism. Several proteins containing one or more chromodomains were identified in yeast, protozoa, animals, and plants.²⁴ The canonical chromodomain structure is a monomeric three-stranded antiparallel β -sheets flanked by a C-terminal alpha helix.

There are three classes of chromodomain containing proteins: some of them contain an N-terminal chromodomain and a structurally similar C-terminal chromo shadow domain; in addition, there are proteins with a single chromodomain and proteins with paired tandem chromodomains.²⁴ More in the details, the histone peptides reside in a surface groove, where

three aromatic residues that constitute the binding cage are able to interact with various degrees of methylated lysine, with a preference for trimethylated over dimethylated lysines.

Some chromodomain proteins appear to function as adapters to recruit protein complexes that implement genetic regulatory programs in transcription, replication and/or repair. In several cases, though, chromodomains are structural components of enzymes and the chromodomain may function targeting intrinsic catalytic activity to local chromosomal substrates as well as recruiting protein complexes. Some chromodomains appear to have significant nucleic acid binding activity, although the nature of binding specificity and the contribution of this property to activity *in vivo* remain to be unraveled.⁷

1.2.2.2.1 Chromodomain of MRG15

MORF4-related gene on chromosome 15 (MRG15) is highly expressed in a wide variety human tissues and identified in many other eukaryotes, and together with MORF4 (mortality factor on chromosome 4) and MRGX (MORF4-related gene on chromosome X) constitute the MRG protein family that was first identified as transcription factor involved in cellular senescence.²⁵

In addition to its involvement in cellular senescence, MRG15 is found to be crucial in embryonic development and cell proliferation. The knockout of MRG15 in mice is embryonic lethal and exhibits a developmental delay.²⁶

Cell biological and biochemical studies showed that MRG15 is most likely to function in chromatin remodeling and transcriptional regulation through participation in two nucleoprotein complexes, MAF1 and MAF2 (MRG15-associated factors 1 and 2, respectively). MRG15 interacts through the C-terminus with the tumor suppressor protein retinoblastoma (Rb) and a novel nuclear protein PAM14, a protein associated with MRG15 in MAF1.²⁷ It is also involved in interactions with HDAC (histone deacetylase) containing transcriptional corepressor

mSin3A and the plant homeodomain zinc finger protein Pf1. The N-terminal part of MRG15 interacts with hMOF (human male absent on first) in MAF2.²⁷ In addition, MRG15 stimulates the catalytic activity of the protein Ash1, which is a SET domain histone methyltransferase that mono- and di-methylates lysine 36 of histone H3 (H3K36), its natural substrate.

Human MRG15 (37 kDa protein composed by 323 amino acid residues) consists of a chromodomain and a conserved MRG domain, which are linked together by a flexible region (Figure 1.5).

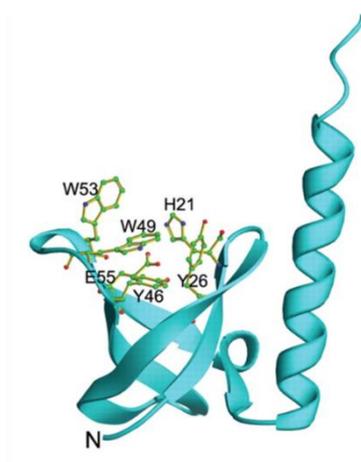


Figure 1.5 Representation of aromatic cage of MRG15.

The MRG domain is a long α -helix highly conserved and among all MRG proteins and it appears to function as an adaptor module to interact with other proteins in nuclear protein complexes.

The chromodomain of MRG15 is a twisted β -barrel that consists of five β strands. Highly conserved residues Tyr26, Tyr46 and Trp49 form the hydrophobic pocket which appears to be a potential binding site for a modified lysine of histones. Through this binding cage, MRG15 is able to bind di- and tri-methylated lysine 36 of histone H3 (H3K36) but its exact function is not

yet well understood. However, the conservation of the chromodomain in many MRG15 homologs underscores its functional importance.

The structural comparison indicates that the overall structure of the MRG15 chromodomain is more similar to that of the dMOF chromo barrel domain than that of the prototype HP1/Pc chromodomains and the HP1 chromo shadow domain.

The structural and biochemical data together suggest that the MRG15 chromodomain is able to interact with the modified histone in a mode different from that of the HP1/Pc.²⁸

1.3 Chemical probes for epigenetic readers

So far, several studies are in progress with the aim to identify small molecule chemical probes for epigenetic readers. Chemical probes represent important tools to explain the biological function of epigenetic regulators, allowing the study of specific domain that compose the proteins of interest. In this context, it is well known that many epigenetic regulators are multidomain proteins mediating writer, eraser and reader functions and molecular biology approaches resulting in reduction or loss of multi-domain proteins do not allow to evaluate functions of individual domains.²⁹

Epigenetic readers represent an important group of targets for the development of chemical probes or drugs. In particular, regarding the methylation readers, druggability is more variable compared to bromodomains, considering aspects as the volume and hydrophobicity of the binding pocket. In addition, docking approaches based on crystal structures are difficult for methyl-lysine due to conformational adaptation of pocket residues upon ligand binding.^{29 30}

The first identified chemical probe for methylation reader proteins is the molecule UNC1215 targeting the methyl-lysine reader L3MBTL3, which is an MBT family member.²⁹ This protein contains three MBT domains and among all MBT1 and MBT2 interact with UNC1215 in a unique 2:2 polyvalent mode.

More in detail, the methyl-lysine mimetic group in para interact with the MBT domain of different monomers of L3MBT3 forming a quaternary complex (Figure 1.6).³¹

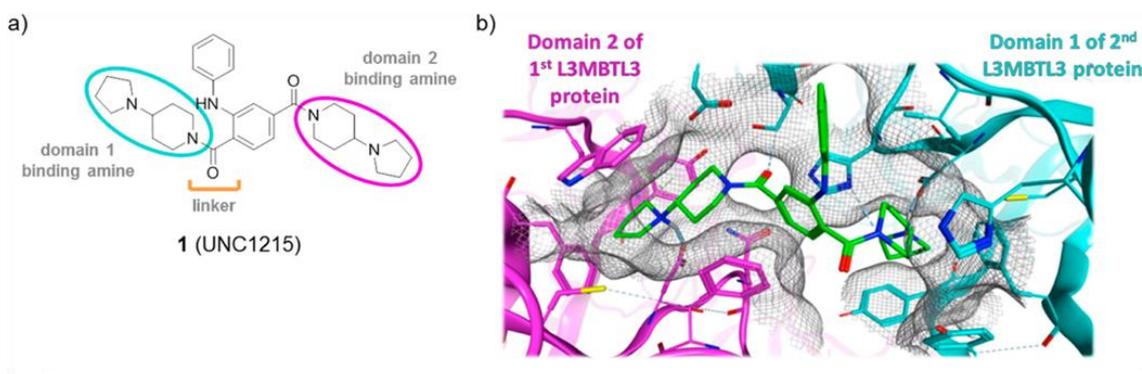


Figure 1.6 a) *UNC1215* chemical structure; b) structure of the co-crystal of *UNC1215* (green) that bind two molecules of *L3MBT3*. The ammine in meta to the aniline portion is positioned in the binding pocket of the MBT2 domain(magenta) while the ammine in ortho to the same aniline portion binds the MBT1 domain of another molecule of *L3MBT3*(cyan).

The binding affinity between the compound and the reader protein L3MBT3 was determined by Isothermal Titration Calorimetry (ITC), exhibiting a K_D of 40 nM. Nevertheless, it was reported that UNC1215 is able to bind other reader proteins, like 53BP1, PHF20 and PHF20L1, showing in this way a low selectivity profile.³² In addition, UNC1215 is the only molecule described as a binder of the PHF20 second Tudor domain with a low micromolar affinity, confirmed also with pulldown experiments using the biotinylated compound.³²

1.4 Epigenetics and drug discovery

In the last few years, epigenetics became a research field particularly pursued thanks to the achievement of several goals in epigenetic drug discovery and together with the continuous new knowledge about the influence of epigenetics in diseases. Furthermore, the speed of developments in drug design indicates the huge demand for this novel avenue to be fully exploited in order to combat diseases.³³

Despite all the progresses and the discoveries made in this field, to date, only seven epidrugs were approved by FDA (histone deacetylase, HDAC and DNA methyl transferase, DNMT) and about twenty compounds are in clinical trials (EZH2, LSD1, DOT1L, PRMT5 and BRD family)³⁴

Since the discovery of Bromodomain and Extra Terminal (BET) inhibitors, a great effort has been spent in investigating the effects of chromatin reader's inhibition, specifically the class of proteins assigned to bind acetylated and methylated residues. Nevertheless, little breakthrough has been made in targeting other chromatin readers.³⁵

Surely, a relevant reason for the lack of epigenetic reader modulators is the druggability of the targets. Unfortunately, few epigenetic targets possess binding sites easily accessible by a putative modulator.

The lack of reader modulators is also caused by the low availability of robust and reliable screening methods to investigate the activity of these molecules. Moreover, in the cellular context, the activity of many epigenetic targets is correlated to the cooperation with other proteins, mainly regarding the “readers”, hence the study of the target alone can be inaccurate.

In addition, the choice of the method is an important aspect to consider, and it depends on the drug discovery stage. It is necessary to find the best compromise between the throughput, the size of the library to be screened and the sample availability. For these reasons, several aspects need to be further investigated with the aim to implement epigenetic drug discovery into clinical management of human disorders.³³

2. AIM OF THE WORK

2.1 Importance to use a screening platform.

As already mentioned, (Chapters 1.4) compared to epigenetic writers and erasers, readers have been less intensively pursued so far as therapeutic targets, especially concerning methylation. Nevertheless, in recent years, epigenetic readers are becoming particularly fascinating as targets in drug discovery.

As to date most of the reader modulators discovered regards acetylation, it is necessary the design, development, and identification of new modulators for epigenetic readers of methylation.

One of the main reason of the lack of inhibitors for this class of reader is the low availability of efficient assay systems for the screening of small-molecule ligands. Some techniques, for example, require large amounts of purified protein or they need specialized equipment and have limited use for high-throughput screening. In addition, all the techniques are not suitable as stand-alone solutions as they are all individually prone to the identification of false positive and false negative.

False positive are the pan assay interference compounds (PAINS), which are a set of 400 different structures that give false readouts in different ways. Some are fluorescent or strongly colored, and in particular assays, they give a positive signal even in the absence of protein.

Others are capable to coat the protein or sequester metal ions that are essential to a protein's function, or they may alter proteins chemically without fitting specifically into a binding site.³⁶

In order to solve these issues, a good approach is the use of orthogonal methodologies with different readouts organized in a modular platform. In this way, crossing the results obtained from each technique it is possible to discover “true hits.”

There are several advantages using this approach. First, crossing the results obtained from technique based on different principles, it is possible to obtain reliable and robust data as each technique suffers of intrinsic limitations that could affect the outputs.

In addition, this approach allow the identification of putative false positive and false negative that interfere with the analysis.

Hence, the optimal workflow for a screening campaign requires, in the first phases of the work, the use of high-throughput techniques suitable for a primary screening, therefore characterized by low sample consuming, easy assay preparation and rapid analysis. Subsequently, the validation of the preliminary results is performed by the use of orthogonal methods, able to deeply characterize the activity of the most promising compounds.

2.2 Aim of the work.

The strategy to use a multiple approach in a drug discovery screening was already considered by different research groups, indeed many works are reported on the use of screening platforms. One of the most suitable example is the work published in 2016 regarding the development of a versatile screening platform for the identification of small molecule modulators for the methyl reader Spindlin1.⁴ This platform could be divided in two levels. A first level is based on biophysical and biochemical assays while the second level consists on cellular assays. Focusing on the first part, the primary screening was carried out using an AlphaLISA assay. Subsequently, a fluorescent polarization (FP) displacement assay was used to confirm the results of the primary screening. In addition, two biophysical techniques, the differential scanning fluorimetry (DSF) and the biolayer interferometry (BLI) were used to measure

parameters as the T_m and kinetic constants respectively. All these techniques were used to confirm the ligands and to identify putative false positive and false negative.

Beside the work already mentioned, also this Ph.D. project is focused on the development of a versatile screening platform composed by biophysical and biochemical techniques to identify new epigenetic modulators for the methylation readers (Figure 2.1). However, the techniques and their position in the workflow are completely different from the work mentioned before. In particular, the primary screening is carried out using the nanoDSF that compared to conventional DSF method is label-free and more sensitive. In order to measure the binding affinity, two gold standard orthogonal techniques were chosen, the Surface Plasmon Resonance (SPR) and the Microscale Thermophoresis (MST). Finally, to evaluate the activity of the library compounds, it was developed an AlphaLISA assay. In the following paragraphs, all these techniques will be described in details showing the benefits and also the disadvantages.

This screening platform was applied to three different reader proteins (PHF20, Spindlin1, and MRG15) to demonstrate the robustness and versatility of the platform. The small libraries used in this work were synthesized in our laboratory, the “Epigenetic Medicinal Chemistry Laboratory” (EMCL).

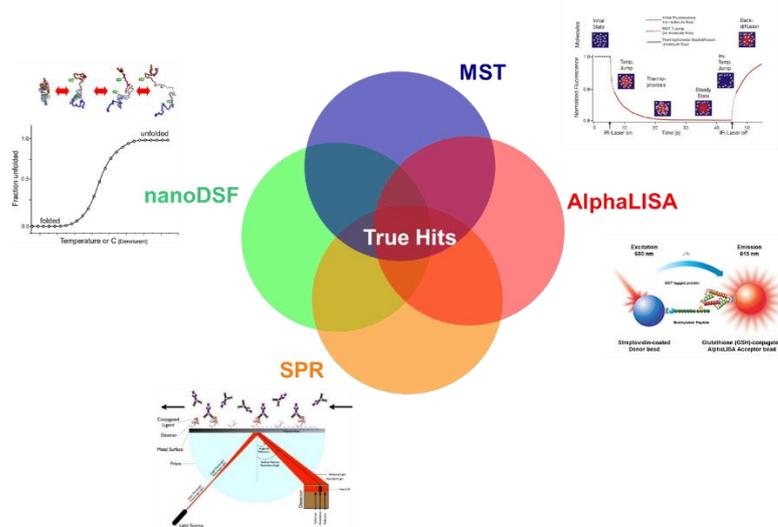


Figure 2.1: Screening platform scheme

2.3 Methods

2.3.1. nanoDSF

The nano Differential Scanning Fluorimetry (nanoDSF) is a biophysical technique based on the study of the thermodynamic stability of proteins that is quantified as melting temperature (T_m). In presence of a putative ligand, this value can change causing a positive or negative shift. Indeed, increasing the temperature, the proteins undergo a sharp, cooperative transition from the native to the denatured state, which is experimentally observable as a characteristic sigmoidal property curve from which the T_m is derived as the inflection point (Figure 2.2).³⁷

The nanoDSF is a label-free technique and the T_m is calculated exploiting the presence of tryptophans in the amino acidic sequence. This fluorescent amino acid, due to its hydrophobic nature, is located in the hydrophobic core but during the denaturation process, it is exposed.

The emission at two wavelengths is detected, at 330 and 350 nm, which corresponds to the covered and exposed tryptophan molecules respectively.

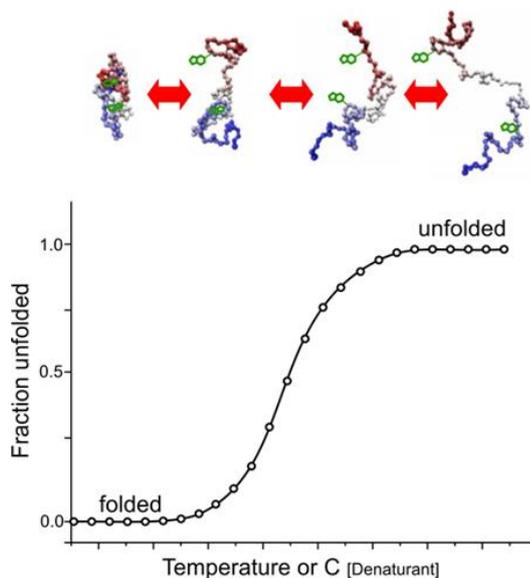


Figure 2.2; *Principle of nanoDSF*

Regarding the benefits, this technique does not need protein labeling and it is carried out in solution. In addition, it requires small amounts of protein, providing the sampling in tiny capillary.

Other benefits such as the fast data analysis and the easy preparation of assay make this technique suitable for a primary screening.

As all the techniques, the nanoDSF has some limitations. Being based on the intrinsic fluorescence determined by the presence of tryptophans in the sequence, not all the proteins can be used in this assay. In addition, this technique could be affected by false negative results, because it is not very sensitive and low affinity interactions cannot be determined. In the end, some detergents can interfere.³⁷

2.3.2 Surface Plasmon Resonance

Another biophysical method to study biomolecular interactions in drug discovery campaigns is the Surface Plasmon Resonance (SPR). It is label-free and capable to measure in real-time binding affinities and kinetic parameters of the interaction between a target biomolecule (ligand) immobilized on a metal surface and the analyte solution flowed over the surface. This surface is typically a thin film of gold on glass support that is called biosensor chip.

The phenomenon at the basis is known as surface plasmon resonance and it occurs when the light hits a metal surface. The light is totally reflexed at a critical incidence angle and its portion is able to excite the electrons of the metal surface layer. The electron movements are now called plasmon, and they propagate parallel to the metal surface. The plasmon oscillation in turn generates an electric field whose range is around 300 nm between the metal surface and sample solution.

Substantially, the resonance phenomenon occurs at a defined angle (SPR angle) that is dependent on the refractive index changes. For this reason, in case of interaction between the

target biomolecules immobilized and the analyte solution, the refractive index at the surface changes and the binding is detected.³⁸

The SPR experiment is usually presented in the form of a sensorgram that reports the resonance or response units (RU) correlated with the time.

At the beginning of the experiment, the immobilized ligand is alone on the sensor chip.

Then the analyte is injected into the flow cell, and if there is binding, it is possible to see an association phase during which binding sites become occupied and the shape of this curve can be used to measure the rate of association (k_{on}). When steady-state is achieved, the RU value corresponds to the changed final critical angle. This maximum RU value relates to the concentrations of immobilized protein and analyte molecules and so it can be used to measure the binding affinity (K_D). When analyte molecules are removed from the continuous flow there is a dissociation phase during which binding sites become unoccupied and the shape of this curve can be used to measure the rate of dissociation (k_{off}). The surface can then be regenerated to start a new experiment (Figure 2.3).³⁹

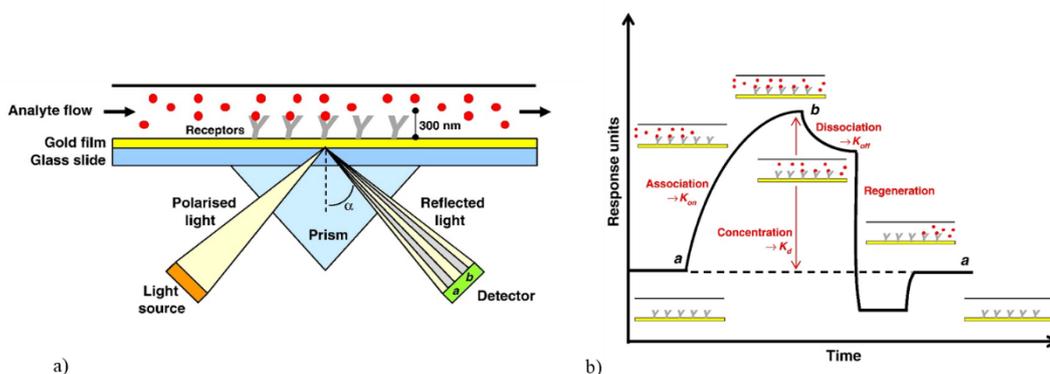


Figure 2.3 a) Principle of SPR and illustration of biosensor components; b) Sensorgram scheme.

Surely, one of the most important benefits is the low consumption of the sample. In addition, it is a technique where the analysis is made in real-time and the target proteins don't need to be labeled. In addition, it is possible to have kinetic information. Regarding the disadvantages, the

SPR requires the immobilization of one of the binding partners. Since immobilization occurs randomly, it is possible that the binding site of the target protein could be hidden.

In addition, being a technique that depends on the refraction index change, any refraction index change, other than from the interaction, can also be detected.

2.3.3 Microscale Thermophoresis

MST is emerging as a sensitive method that can be used to assess biomolecular interactions and is used to study interactions between various binding partners of different molecular sizes: protein-protein, antibody-antigen, protein-DNA, and protein-RNA interactions and the binding of ligand to ternary complexes.

Thermophoresis, also known as the Ludwig-Soret effect, is defined as the direct movement of molecules through a temperature gradient. Thermophoretic properties of a biomolecule are determined by charge, size, and hydration shell. Upon binding to an interacting partner, one or more of these three parameters may be altered, resulting in a change in the initial thermophoretic movement.⁴⁰

To monitor the thermophoresis of biomolecules, the instruments use an infrared (IR) laser, which heats an aqueous sample loaded in a thin glass capillary. The laser focus produces a localized microscopic temperature gradient with a heat range of 2–6 °C and the ligand-protein complexes demonstrate a different motility compared with unbound molecules. This movement through the laser-heated spot is monitored *via* fluorescence of the target protein, either from the intrinsic UV fluorescence of tryptophan, tyrosine, and phenylalanine residues (tryptophan exhibiting the strongest fluorescence) or from labeling the protein with a fluorescent dye through covalent attachment to lysine or cysteine residues, or noncovalent attachment to a polyhistidine tag. Rapid scanning of a series of capillaries containing a constant concentration of fluorescent target molecules and increasing concentrations of ligand enables the

determination of equilibrium binding constants. Alternatively, the ligand can be fluorescently labeled and the unlabeled protein titrated.⁴¹

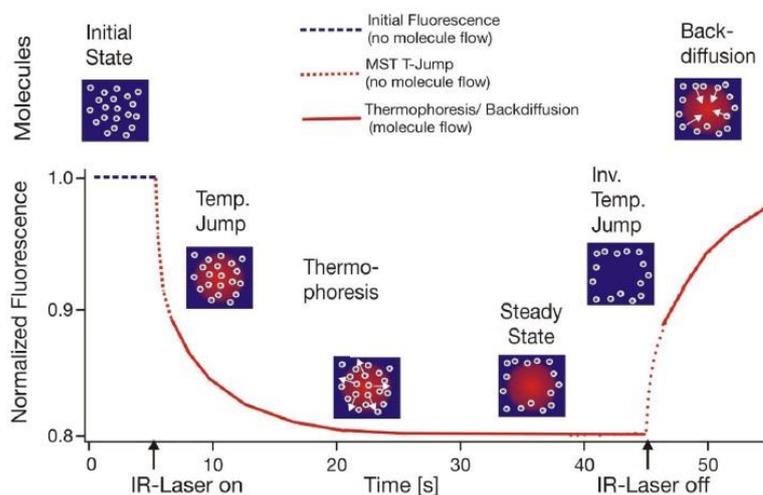


Figure 2.4 Typical signal of an MST experiment

Initially, the molecules are homogeneously distributed and a constant initial fluorescence is detected. Within the first second after activation of the IR laser, the “T-Jump” is observed, which corresponds to a rapid change in fluorophore properties due to the fast temperature change. Subsequently, the thermophoretic movement of the fluorescently labeled molecules out of the heated sample volume can be detected (Figure 2.4).⁴⁰

Among the numerous advantages, the high sensitivity and the rapid analysis are the most relevant. In addition, this technique is immobilization-free and it is possible to work without labeling, using the intrinsic fluorescence. The measurement is made in real-time and it is possible to detect aggregation, sticking and precipitation effects. Nevertheless, also this technique has some limits, for example, some compounds can interfere with protein signal causing adsorption or emission phenomena. In addition, the label-free analysis needs at least one tryptophan in the protein sequence.

2.3.4 AlphaLISA

Alpha (Amplified Luminescent Proximity Homogeneous Assay) is a biochemical technology developed by PerkinElmer.

This technology is based on the measurement of the fluorescence emission following the proximity of two types of beads. These beads, called Donor and Acceptor beads, in presence of a specific analyte, are close enough to generate a chemiluminescent signal.

In particular, Donor beads contain a photosensitizing phthalocyanine that, when irradiated at 680 nm, excites ambient oxygen to a singlet state. Excitation of each Donor beads generates approximately 60,000 oxygen singlets per second, resulting in a highly amplified response upon interaction with Acceptor beads. Nevertheless, the singlet oxygen species are highly unstable (half-life 4 μ s) and able to travel at least 200 nm in solution before decay. If a polystyrene acceptor bead is in proximity, the singlet oxygen will react with the photosensitive dye which covers the bead and the light can be emitted (Figure 2.5).

Depending on the type of assay (AlphaLISA or Alphascreen), the acceptor beads and the emission wavelengths are different. The AlphaLISA assay exploits acceptor beads covered with Europium chelate that emits at 615 nm while the Alphascreen acceptor beads are covered with three different dyes: thioxene, anthracene, and rubrene. The first one reacts initially with singlet oxygen to produce light energy, which is subsequently transferred to anthracene and thence to rubrene. The final compound in the cascade, rubrene, emits light at wavelengths of 520-620 nm.⁴²

The AlphaLISA is an evolution of the AlphaScreen assay designed to meet requirements for high throughput assays for biomarker detection. Indeed, Europium emission is more intense and better defined spectrally (615 nm) than observed in AlphaScreen assays (520-620nm). Consequently, the emission wavelength is less prone to matrix interferences from compounds such as hemoglobin or transferrin.

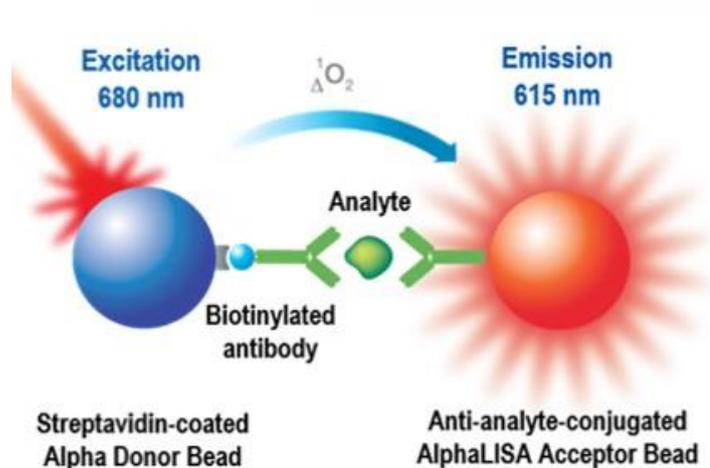


Figure 2.5 Principle of AlphaLISA technology.

In addition, the bead surface is coated with latex-based hydrogels containing reactive aldehydes. This coating both reduces non-specific binding but additionally provides a functionalized surface to which a variety of different ligand or receptor binding partners may be affixed. Collectively, these surface properties allow quantitation of a wide range of enzymes, protein-protein interactions, as well as DNA and RNA hybridization.⁴³

As with any assay technology, there are advantages and limitations also for Alpha assays. The main advantage of the technology is that it is applicable to a very broad range of analytes. All the assays are homogeneous, rapid and robust, being more sensitive than previously reported immunoassay methods. Furthermore, AlphaScreen assays do not require insertion of large fluorescent epitope tags that may sterically hinder bio molecular interactions. AlphaScreen and notably AlphaLISA can be employed to measure enzyme activity in crude biological fluid extracts such as cell lysates, serum and plasma and a variety of cellular/body fluid matrices that do not easily affect the assay readout.

The main disadvantage of Alpha technology is that it is sensitive to intense light or long exposure to ambient light, nevertheless this problem can be easily overcome by simple assay

adjustments. In addition, singlet oxygen can be sequestered by compounds in screening libraries that can scavenge radical oxygen.⁴³

3. RESULTS AND DISCUSSION

3.1. Tudor domains of PHF20

3.1.1. *Background*

Among several reader proteins, Tudor domains of PHF20 are members of the Royal superfamily (Chapter 1.2.2). The interest for the reader domains of PHF20 derived from a partnership with Dr. M. Bedford (Department of Epigenetics & Molecular Carcinogenesis, Smithville, Texas, USA). With the aim to identify new modulators for methylation reader proteins, it was published a work based on a “library on library” screening approach, that consists of a screening of a library of small molecule compounds against a pool of methyl lysine reader domains spotted on a microarray plate.⁴⁴

Starting from this work, the compounds EML638 and EML643 (analogs not biotinylated of EML405 and EML417 respectively) were selected among all the tested compounds because of their interesting binding to the Tudor domains of PHF20. In particular, EML638 showed a low selectivity interacting with the Tudor domains of different proteins (PHF20, PHF20L1, 53BP1, SPIN1) and also with some chromodomains, while EML417 selectively binds to the Tudor domains of PHF20. The structures of the compounds are reported in Figure 3.1.1.

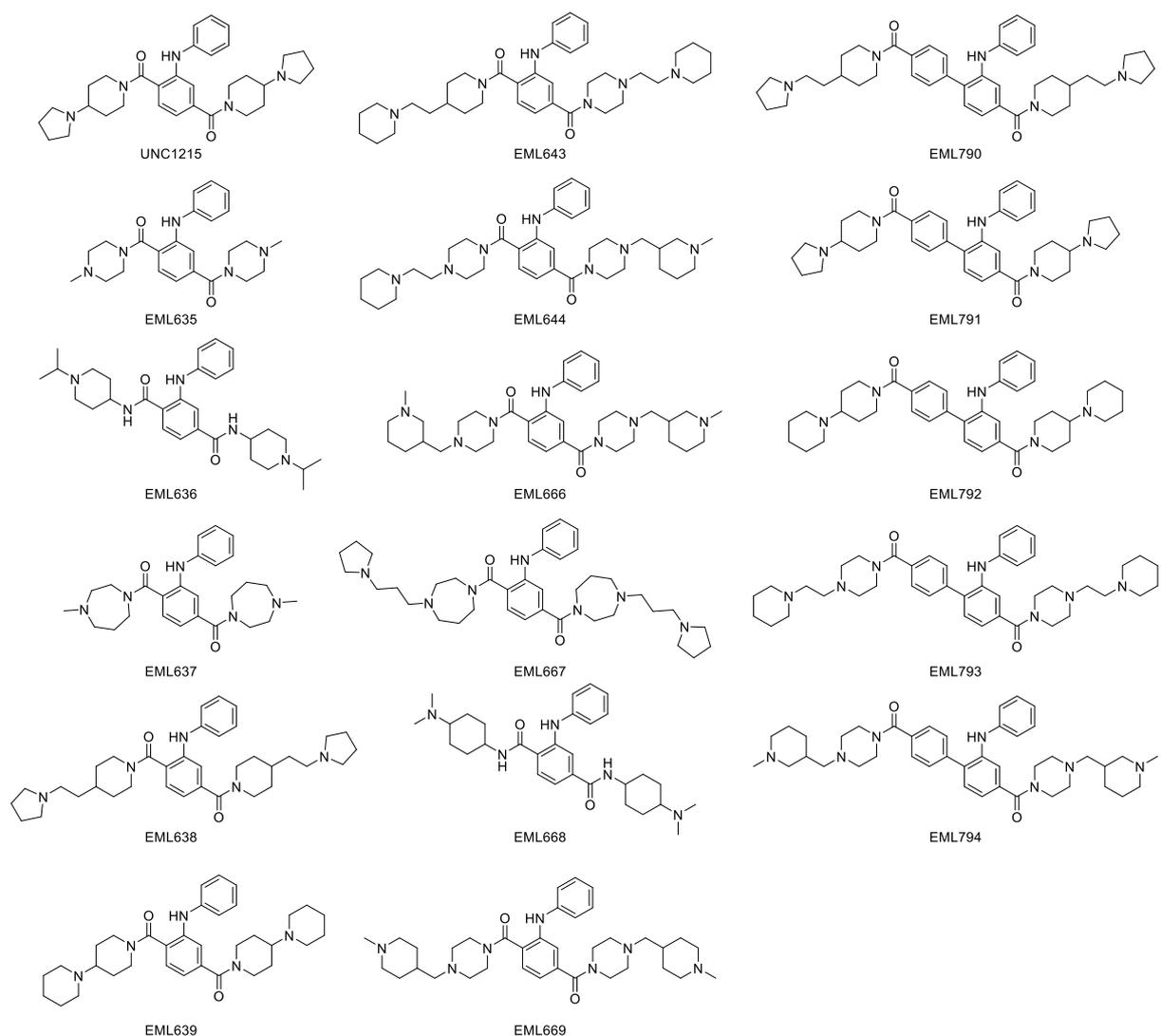


Figure 3.1.1 Library of small-molecule compounds screened for PHF20.

Despite the ability to screen compounds against thousands of immobilized targets, unfortunately, microarray technology gives only qualitative information regarding the binding of compounds. Considering that all the knowledge of PHF20 was obtained studying different sequences of this protein, we decided to screen the same library of compounds using these sequences of PHF20 (GST-Tudor1-2, Tudor1-2, Tudor2 His-tagged and His6-Tudor 2 mutate from cysteine to serine).

The use of different sequences allows to elucidate several aspects. First, the influence of the GST tag. This tag (molecular weight above 26KDa) is able to stabilize and increase the

solubility of the fused protein of interest. Nevertheless, considering that is heavy twice the sequence of the protein (around 35 KDa against 17 KDa of Tudor1-2), the GST tag could influence the conformation of the protein, preventing its accessibility by small-molecule ligands.

Another aspect to consider is the presence of the Tudor1 domain. As reported in different papers and already discussed in Cap.1, for this domain is not reported a reader activity. Hence it is not known if the presence of Tudor1 domain could disturb the binding of the Tudor2 domain with its substrates.

In addition, as already mentioned, PHF20 is able to homodimerize exploiting the presence of two Cysteines in the Tudor2 domain sequence. The homodimer formation increases the affinity against the dimethylated substrate, for this reason, we wanted to understand if the mutation of the two Cysteines in Serines could affect the binding.

In the end, considering that these sequences are not commercially available, a relevance part of this work was focused on the expression and purification of all four sequences of PHF20 (Figure 3.1.2).

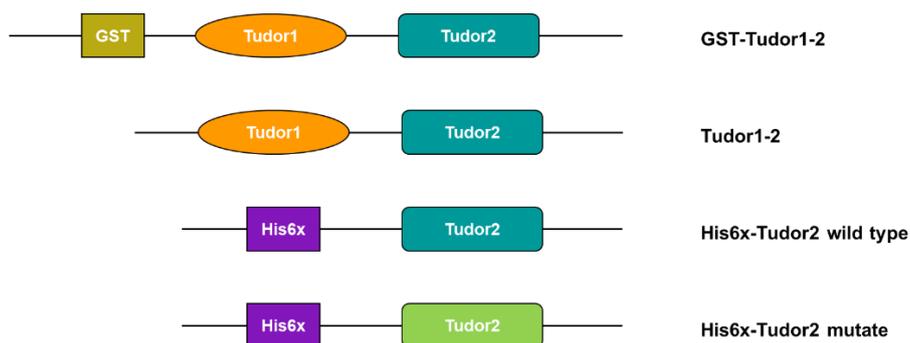


Figure 3.1.2 Sequences of PHF20 scheme

3.1.1.1. Expression and purification of GST-Tudor1-2

The expression of this protein was made using the expression vector pGEX-4T-1 containing the DNA sequence of Tudor1-2 domains of PHF20, generously provided by Dr. M. Bedford (Figure 3.1.3).

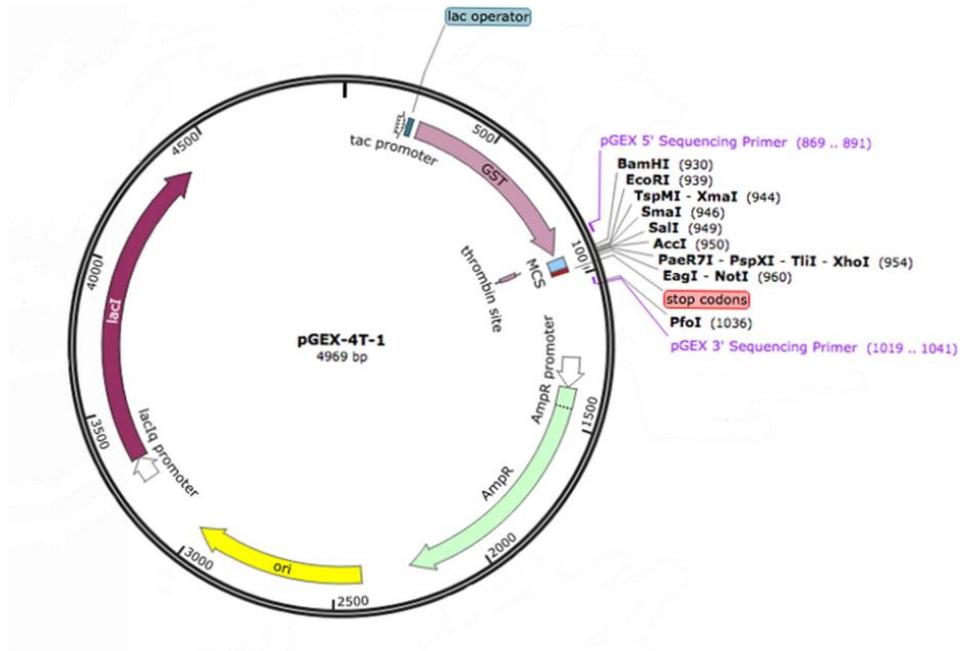


Figure 3.1.3 Plasmid map for the expression of PHF20 Tudor domains (aa 1-150) as GST tagged protein.

This vector allowed the protein expression with the GST-tag at the N-terminal region, improving the purification process. In addition, a specific cleavage site is located between the GST-tag and protein sequence, allowing to obtain also the Tudor1-2 sequence without the tag.

The vector was transformed into BLR(DE3)pLysS competent cell and the expression was conducted in LB medium. Expression tests were performed at different temperatures, times of growth and inducer concentration with the aim to define the optimal condition that was reached using 1mM IPTG for 4h at 37°C. Subsequently, the protein was purified by affinity chromatography using a ÄKTA purifier system and a 1mL GSTrap HP column (Figure 3.1.4)

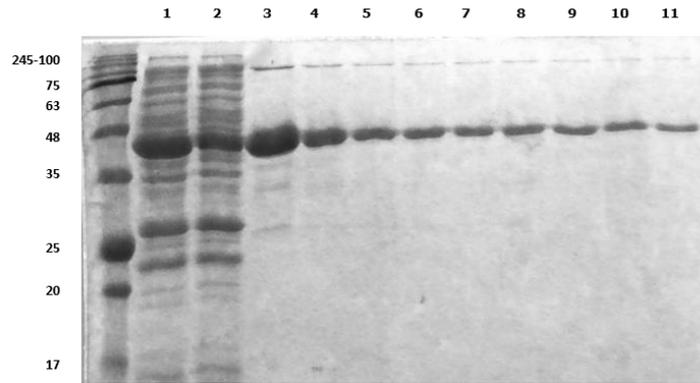


Figure 3.1.4 SDS-PAGE of the purification of the protein by affinity chromatography using GStrap HP column and an AKTA FPLC at 1 mL/min. Lane 1: lysate; Lane 2: unbound fraction; Lanes 3-11: elution fractions

3.1.1.2. GST cleavage and purification of Tudor1-2 domains

The sequence Tudor1-2 was obtained exploiting the presence of a cleavage site recognized by the Thrombin. The Thrombin CleanCleavage kit, that consists in agarose beads covered with Thrombin, was used to remove the GST-tag. The cleavage was confirmed via SDS-PAGE (Figure 3.1.5) and the purification of the protein free from its tag was performed exploiting their different isoelectric point through ionic exchange chromatography using a Mono Q column (Figure 3.1.6).

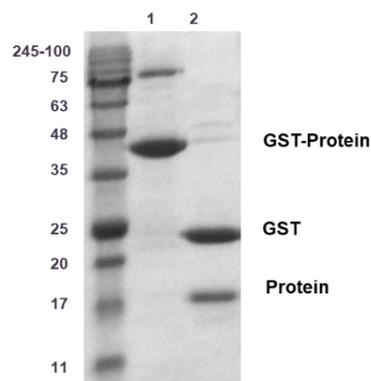


Figura 3.1.5. SDS-PAGE after the GST cleavage, which was performed incubating the protein with Thrombin beads for 6 h at room temperature. Lane 1: protein before the cleavage; Lane 2: protein after the cleavage.

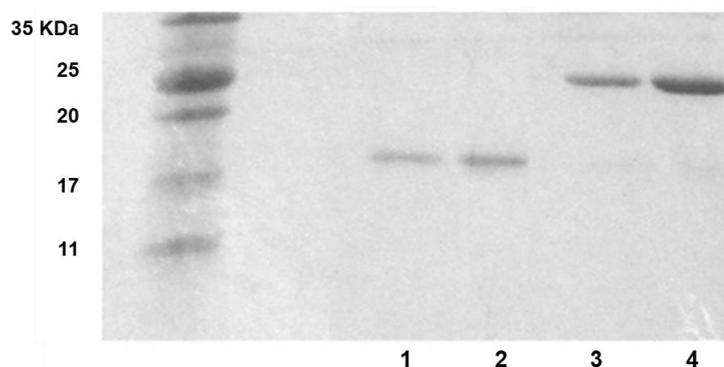


Figure 3.1.6 SDS-PAGE of the separation of the protein from GST by anionic exchange chromatography using a Mono Q column. Lanes 1-2: Tudor1-2; Lanes 3-4: GST.

3.1.1.3. Expression and purification of His –Tudor2 sequence.

Regarding His-Tudor2 wild type, the expression was made using the expression vector pTEV-(pET15b) generously provided by Dr.G. Mer. The vector directs the expression of a recombinant protein with an N-terminal hexahistidine (His tag) that is removable by the tobacco etch virus (TEV) protease.

The vector was transformed in BLR(DE3)pLysS, and the expression of the protein was induced in E.coli. The expression tests were performed at different temperatures, times of growth and different concentrations of the inducer but, unfortunately, in all the conditions used, there was no expression of protein.

Subsequently, it was used another E.Coli strain, the BL21(DE3)pLysS. The expression tests were performed again, but this time conducting the expression at O.D. 0.8, 1 mM IPTG overnight at 15 °C the protein was expressed in low amount. With the aim to improve this condition, it was performed expression tests at different concentrations of IPTG and since there was no improvement, we decided to proceed with the purification of the protein from the lysate

using affinity chromatography and a HisTrap HP column. Unfortunately, the protein showed instability during the purification steps.

Once again, it was changed cells strain, in particular, the vector was transformed in Rosetta(DE3)pLysS designed to enhance the expression of eukaryotic proteins that contain codons rarely used in E.coli. The expression tests showed a huge improvement, indeed almost all the conditions showed the expression of the protein. Hence, it was used the best condition for the purification but also in this case the protein degraded.

At this point, it was performed tests at different concentrations of IPTG, times of growth, optical density and different purification temperatures. Performing the induction of the expression at O.D. 0.6, 10 μ M IPTG for 6 hours at 15 °C and purifying it at 4°C, the protein was not degraded (Figure 3.1.7).

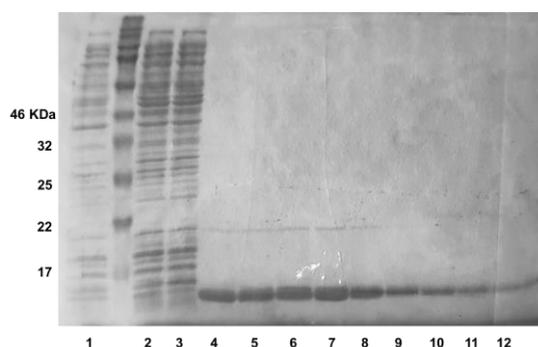


Figure 3.1.7 *SDS-PAGE of the purification of the protein by affinity chromatography using HisTrap HP column and an AKTA FPLC at 1 mL/min. Lane 1: before Induction; Lane 2: lysate; Lane 3: unbound fraction; Lanes 4-12: elution fractions*

3.1.1.4. *Expression and purification of His –Tudor2 mutate sequence.*

The last protein was the His-Tudor mutate from cysteines to serines. Considering the good results obtained with the Tudor2 wild type, also for the mutated sequence it was used the Rosetta(DE3)pLysS as E.coli strain.

With the aim to define the optimal conditions for protein expression, it was performed expression tests at different temperatures, times of growth and concentrations of the inducer. After the determination of the best expression conditions, the protein purification was performed from the lysate using affinity chromatography, taking advantage of the ÄKTA purifier system (Figure 3.1.8). This sequence, differently from the wild type, was obtained performing the purification at rt, hence it is clear that the mutate is more stable than its analog wild type.

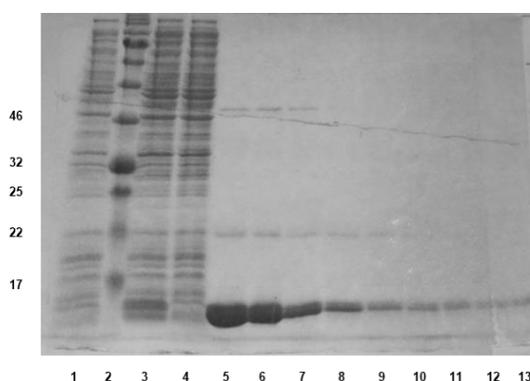


Figure 3.1.8 *SDS-PAGE of the purification of the protein by affinity chromatography using HisTrap HP column and an AKTA FPLC at 1 mL/min. Lane 1: lysate; Lane 3: unbound fraction; Lanes 4-13: elution fractions*

3.1.2. Protein stability

Once obtained all the proteins, the protein stability was verified doing a melting profile with nanoDSF. These experiments were performed using standard treated capillary, with a temperature range from 20 °C to 95 °C at a heating rate of 1 °C per minute.

The sequence His-Tudor2 wild type shows two T_m values, which the lower is around 30 °C (Figure 3.1.9). This result suggests that this sequence is unstable, explaining also the reason why after the purification steps the protein was degraded. Since all the techniques of the

screening platform require incubation time at room temperature, this sequence was excluded from the platform avoiding misleading results.

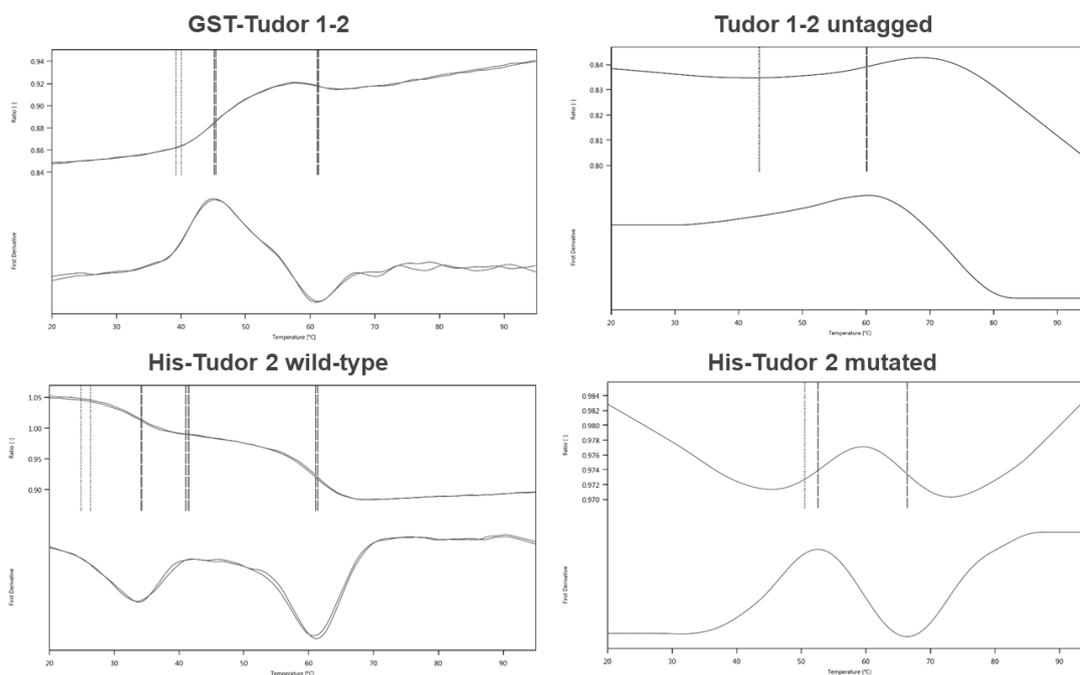


Figure 3.1.9 Comparison of melting profiles for the four different sequences of PHF20

3.1.3. nanoDSF as primary screening

The first method used in this screening platform was the nanoDSF. Considering the easy assay preparation and the fast data analysis, this technique is suitable for a primary screening. All three sequences (GST-Tudor1-2, Tudor1-2, and His-Tudor2 mutate) were used with the aim to evaluate the effect of small-library of UNC1215 analogs designed and synthesized in EMCL. Before proceeding with the library screening it was necessary to validate the method by checking the protein interaction with the UNC1215, which is the only molecule described to interact with the Tudor domain of PHF20.

Seeing a clear interaction between the proteins and the well-known inhibitor UNC1215, the library of compounds was screened at 100 μ M concentration with a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C at a heating rate of 1 $^{\circ}$ C per minute. It was chosen as cut-off a $\Delta T_m \geq \pm 0.5$ $^{\circ}$ C that is sufficient to establish a binding event.

COMPOUNDS	GST-TUDOR1-2	TUDOR1-2	TUDOR2 MUT
UNC1215	0.5	1.8	- 5.2
EML635	0.3	0.2	- 0.2
EML636	0.2	0.7	- 2.3
EML637	0.7	3.8	- 1.5
EML638	0.3	0.5	- 3.9
EML639	0.5	0.3	- 4.7
EML643	0.5	0.2	- 3.0
EML644	0.6	0.6	- 4.7
EML666	0.5	1.7	- 5.3
EML667	1.1	4.9	- 5.4
EML668	0.3	0.9	- 2.3
EML669	0.9	4.1	- 3.9
EML790	0.5	0.5	- 7.1
EML791	0.6	0.4	- 4.0
EML792	0.1	0.3	- 3.4
EML793	0.3	1.2	- 6.8
EML794	0.5	0.6	- 5.6

 Binding ($\Delta T_m \geq \pm 0.5$ °C)
 No Binding

Table 3.1 *nanoDSF* assays were performed using the Prometheus. The small molecule screening was performed at the fixed dose of 100 μ M with a protein concentration of 1.65 mg/mL for GST-Tudor1-2, 0.20 mg/mL for Tudor1-2, and 1.2 mg/mL for His-Tudor2 mutate. The measurements were performed using standard capillaries.

As it is shown in Table 3.1 and as we expected, the compounds bind differently to each protein sequence. Most of the molecules bind the sequence Tudor2 mutated. In addition, the EML792 is selective for Tudor2 mutate, probably the monomer form is preferred for the binding of this compound. There are some compounds that bind two sequences of PHF20 as EML636, EML638, EML639, EML643, EML668, and EML791. While there are other compounds that are able to bind all the sequences of PHF20 such as EML637, EML644, EML666, EML667, EML669, EML790, EML794 and of course UNC1215. Furthermore, EML635 is the only molecule that does not bind any sequence of PHF20, showing clearly that it is unable to bind this protein.

It is necessary to consider that this technique, as for the microarray, is able to detect the binding but it is unable to quantify the binding affinity.

Hence for the determination of the binding affinity, it was developed an MST assay for each sequence of PHF20.

3.1.4. Microscale Thermophoresis assay (MST)

The Microscale Thermophoresis is the second biophysical technique of the platform, and it is useful to study bio molecular interactions determining the binding affinity. The proteins were labeled exploiting their structure features, therefore the GST-Tudor 1-2 and Tudor 1-2 were labeled on the lysine using the Monolith Protein Labeling Kit RED-NHS (Amine Reactive) kit while the Tudor2 mutate was labeled on the His-tag using Monolith His-Tag Labeling Kit RED-tris- NTA kit. The protein labeling was carried out following the procedure suggested on the protocol with a protein:dye ratio of 1:3.

To define the best conditions for the assays, different tests have been performed with the aim to investigate the type of capillaries, the buffer and the protein concentration to use. Hence regarding GST-Tudor1-2, the best condition is reached using premium coated, PBS 1X, Tween 0.05% as buffer and the protein concentration of 1 nM.

For the sequence Tudor1-2 it was preferred the use of premium coated capillaries, Hepes pH 7.5 50 mM, NaCl 150 mM, Tween 0.05% as buffer and the protein concentration of 5 nM. Regarding Tudor 2 mutate, the best condition is reached using standard capillaries, PBS 1X, Tween 0.05% as buffer and the protein concentration of 5 nM.

Before evaluating the interaction with the molecules, it was important to validate the screening method by checking the interaction between each protein and the histone H4K20me2, the natural substrate of PHF20. This experiment was carried out maintaining constant the concentration of the proteins while the peptide (at maximum concentration of 200 μ M) was

serially diluted. After confirming the interaction with peptide, the interaction of each protein with the library of compounds was evaluated. Hence, for each sequence, the protein concentration was kept constant while the compounds (at maximum concentration of 200 μ M) were serially diluted.

Compounds	GST-Tudor1-2 K_D (μ M \pm SD)	Tudor1-2 K_D (μ M \pm SD)	His-Tudor2 mut K_D (μ M \pm SD)
Peptide	11.71 \pm 3.46	1.35 \pm 0.57	2.40 \pm 0.13
UNC1215	29.32 \pm 1.27	30.05 \pm 6.70	66.30 \pm 7.50
EML635	14.61 \pm 3.61	-	32.00 \pm 8.30
EML636	30.53 \pm 2.83	15.92 \pm 3.45	-
EML637	-	-	137.75 \pm 17.32
EML638	67.95 \pm 14.92	17.04 \pm 2.94	-
EML639	-	10.36 \pm 1.98	97.50 \pm 7.93
EML643	29.15 \pm 9.25	-	-
EML644	20.15 \pm 6.23	15.42 \pm 0.95	-
EML666	32.25 \pm 2.76	27.20 \pm 3.86	21.60 \pm 5.09
EML667	27.22 \pm 2.83	56.75 \pm 18.60	12.71 \pm 2.56
EML668	-	-	1.95 \pm 2.05
EML669	21.35 \pm 3.65	30.51 \pm 7.92	8.14 \pm 3.68
EML790	13.25 \pm 0.64	15.05 \pm 3.56	154.00 \pm 34.82
EML791	12.83 \pm 1.13	6.31 \pm 0.98	51.95 \pm 5.66
EML792	9.32 \pm 3.95	2.43 \pm 0.75	-
EML793	9.95 \pm 4.57	-	104.32 \pm 32.93
EML794	18.95 \pm 0.64	6.93 \pm 1.21	21.53 \pm 3.04

Table 3.2: MST assays were performed using the Monolith NT. The small molecule screening was performed at the fixed dose of 100 μ M.

Looking at the results obtained for each sequence of PHF20 (Table 3.2), some considerations can be done. Approximately most of the compounds are able to interact with three different sequence of PHF20, in particular, there are some compounds which are selective for a particular sequence, such as EML637 and EML668 against the Tudor2 mutate while EML643 against GST-Tudor1-2. Moreover, other compounds are able to bind two of the three sequences (EML635, EML636, EML637, EML638, EML639, EML644, EML792, and EML793). Instead, there are other compounds that bind all three sequences (EML666, EML667, EML669, EML790, EML791, and EML794).

In addition, in most cases for each compound, the K_D values relative to the different sequences are substantially different. It is not unexpected because, as already mentioned, the presence of the GST tag, the Tudor1 domain and the capability of the protein to homodimerize can affect the binding.

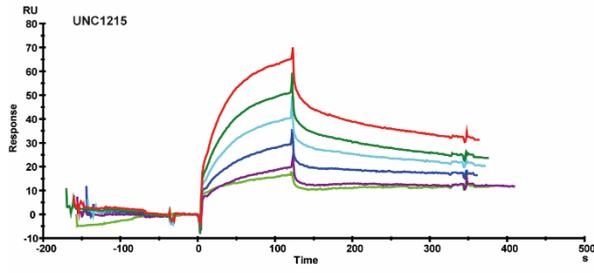
3.1.5. Surface plasmon resonance

With the aim to validate the MST results, it was used an orthogonal biophysical technique, the Surface Plasmon Resonance (SPR). This method allows the measurement of interactions in real-time with high sensitivity and without labeling procedures but immobilizing the protein on the sensor surface while a solution of ligand flows over the surface.

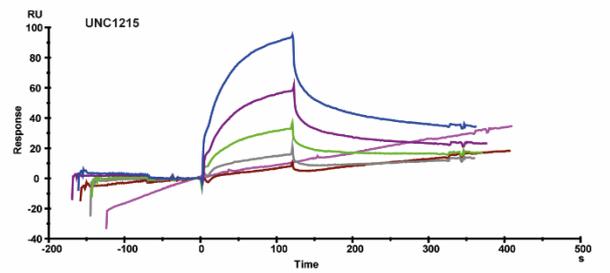
In this experiment two sequences of PHF20 were used, Tudor1-2 and Tudor2 mutate, that were immobilized using amine coupling approach. As this immobilization procedure affects randomly the amine groups of the protein, the immobilization of the GST-Tudor1-2 could give problems. Indeed, the GST-tag could hide the binding site of the protein, so this sequence was excluded from the analysis.

The proteins were immobilized on the flow cells of the biosensor chip (CM5), and the library compounds were injected at various concentrations (from 3.12 μ M to 100 μ M) over the protein surface. The binding of each compound was read in real-time as the change in mass at the sensor surface. After the injections, running buffer was allowed to flow over the surface, and the dissociation of compounds from the surface was observed. The SPR experiments were performed in PBS pH 7.4, 0.005% Tween. The sensorgrams related to the binder compounds are reported in Figure 3.1.10.

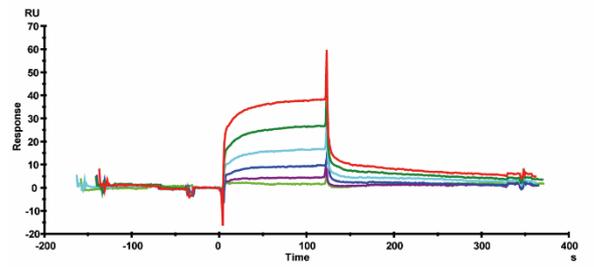
TUDOR1-2



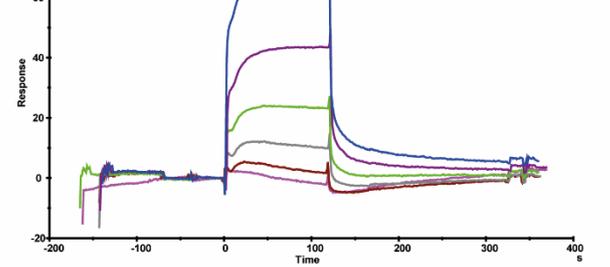
TUDOR2 MUTATE



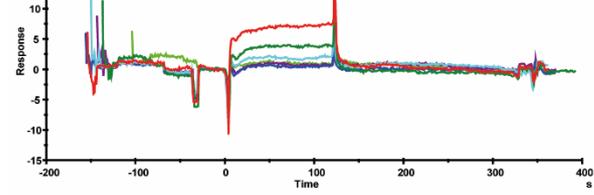
EML636



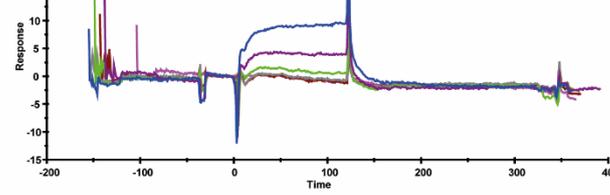
EML636



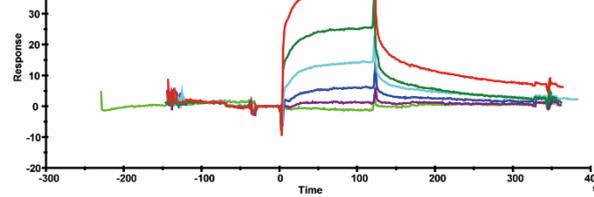
EML637



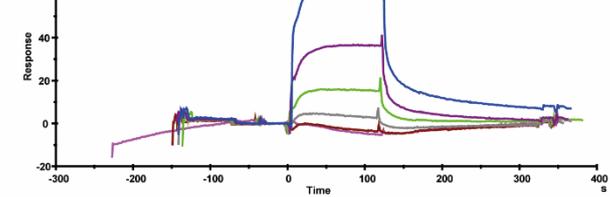
EML637



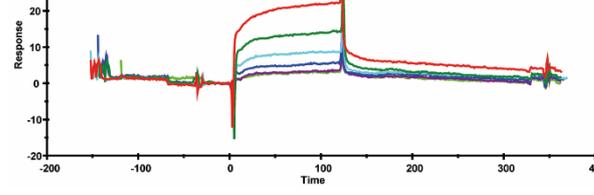
EML638



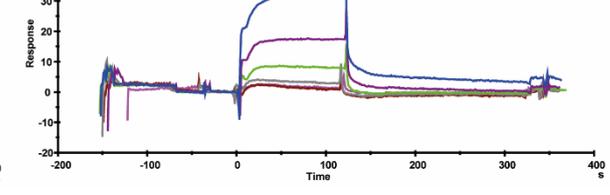
EML638



EML639

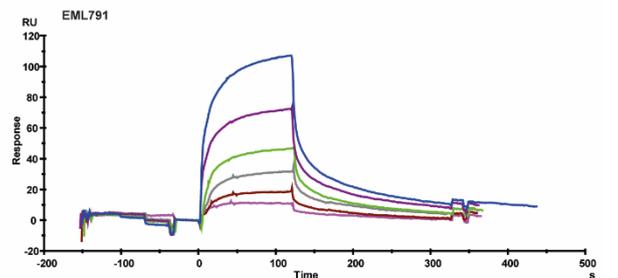
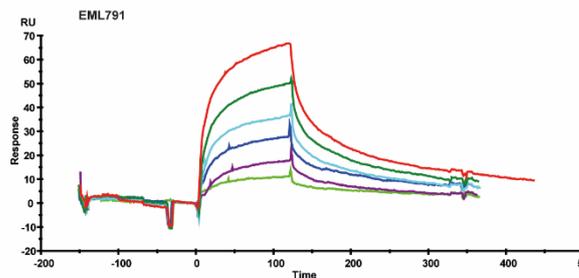
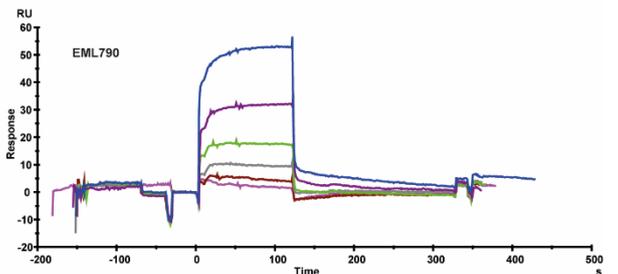
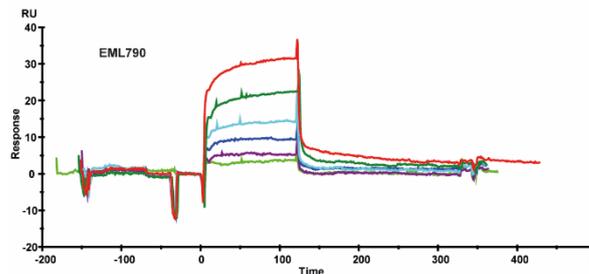
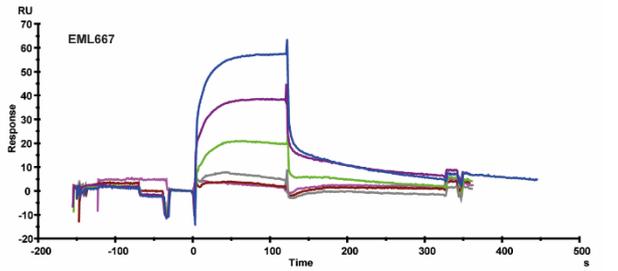
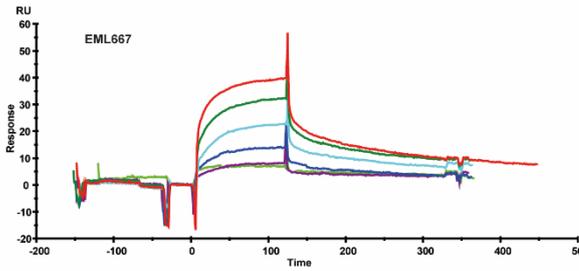
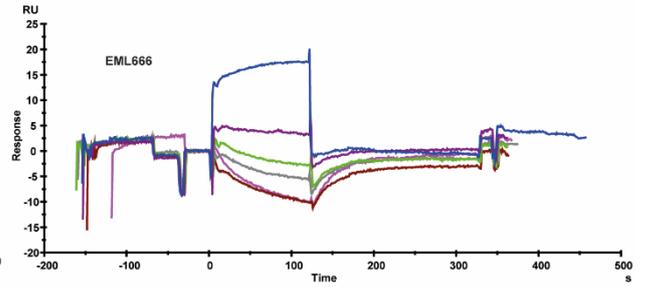
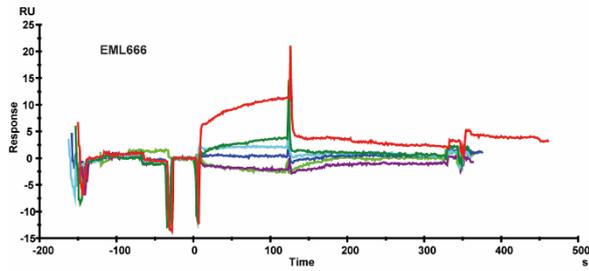
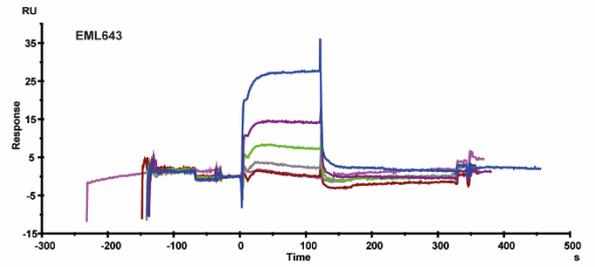
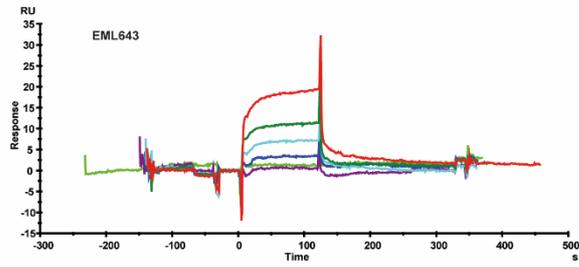


EML639



TUDOR1-2

TUDOR2 MUTATE



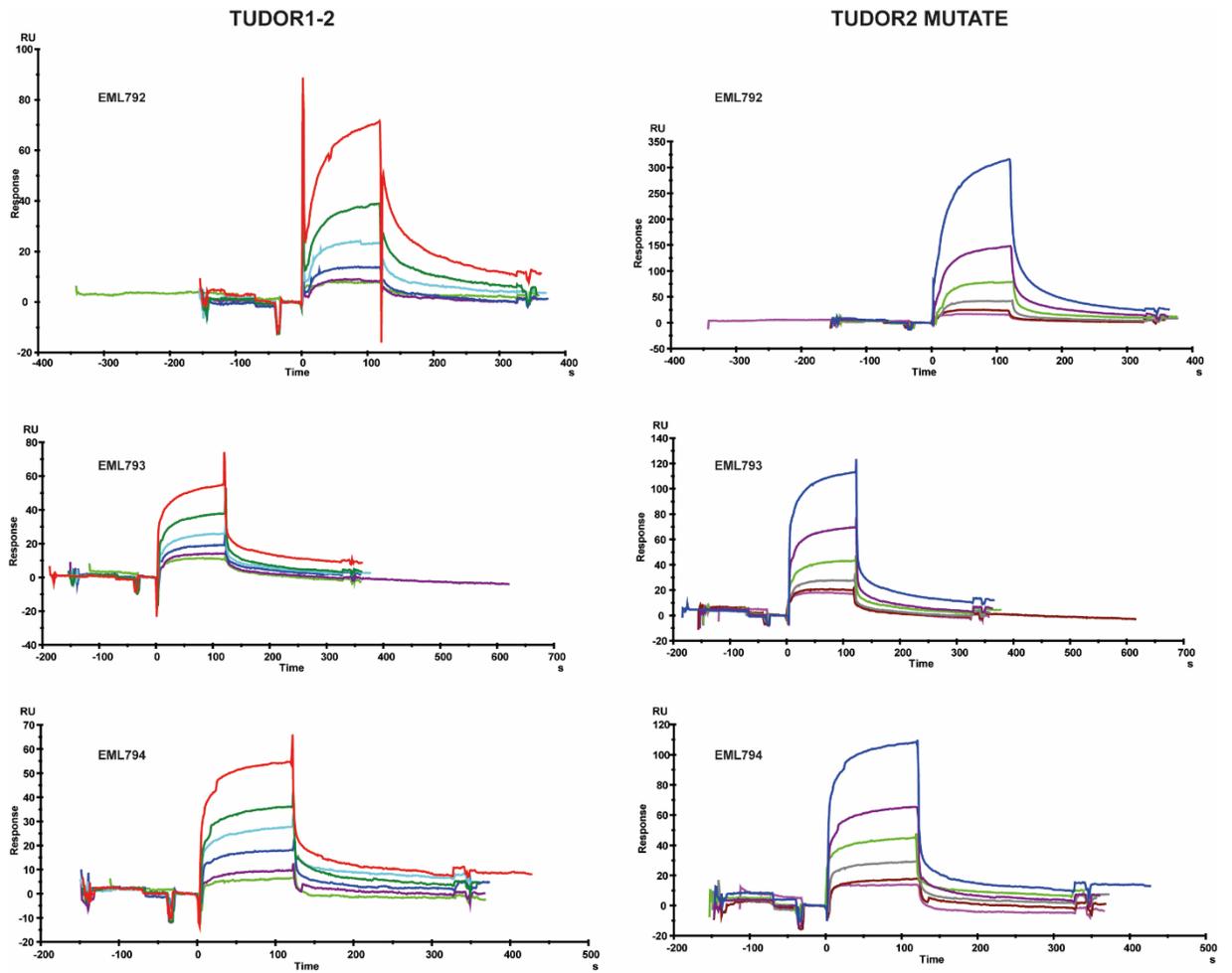


Figure 3.1.10 Sensorgrams of all the compounds able to bind the two sequence of PHF20.

COMPOUNDS	TUDOR 2 MUT			TUDOR 1-2		
	K _D (μM ± SD)	K _{on} [M ⁻¹ s ⁻¹]	K _{off} [s ⁻¹]	K _D (μM ± SD)	K _{on} [M ⁻¹ s ⁻¹]	K _{off} [s ⁻¹]
UNC1215	66.5 ± 4.9 μM	1.23E+02	8.17E-03	8.6 ± 2.1 μM	4.53E+02	3.88E-03
EML635	-			-		
EML636	2.3 ± 1.6 μM	5.09E+00	1.13E-05	41.5 ± 5.8 μM	2.30E+02	9.56E-03
EML637	-			-		
EML638	24.6 ± 3.7 μM	2.18E+00	5.36E-05	81.4 ± 8.1 μM	1.17E+02	9.49E-03
EML639	14.5 ± 1.9 μM	1.45E+00	2.10E-05	20.7 ± 3.2 μM	4.51E+02	9.35E-03
EML643	14.7 ± 0.9 μM	1.03E+00	2.51E-04	251 ± 10.2 μM	8.84E+01	2.21E-02
EML644	-			-		
EML666	-			-		
EML667	68.3 ± 6.6 μM	1.58E+02	1.08E-02	16.2 ± 4.5 μM	4.83E+02	7.83E-03
EML668	-			-		
EML669	-			-		
EML790	3.6 ± 1.1 μM	2.12E-01	1.67E-05	98.2 ± 6.3 μM	1.54E+02	1.51E-02
EML791	26.2 ± 3.9 μM	5.45E+02	1.41E-02	15.3 ± 1.5 μM	7.08E+02	1.08E-02
EML792	268 ± 9.1 μM	9.57E+01	2.56E-02	175 ± 13.2 μM	8.71E+01	1.53E-02
EML793	77.3 ± 7.3 μM	1.97E+02	1.52E-02	22.5 ± 2.3 μM	5.33E+02	1.20E-02
EML794	52.3 ± 4.1 μM	3.15E+02	1.65E-02	27.9 ± 2.6 μM	4.15E+02	1.16E-02

Table 3.3 SPR results of the experiments performed screening the library of compounds on the proteins Tudor2 mutate and Tudor1-2

Table 3.3 shows the results of all SPR experiments performed. It is clear that both proteins bind the same compounds in the library. In particular, for some compounds, the K_D values are comparable (EML639, EML791, EML792, EML794) while for others they are significantly different (UNC1215, EML636, EML638, EML643, EML667, EML790, EML793). Probably the presence of the Tudor1 domain and the presence of the mutation influence in different ways the binding of compounds.

Nevertheless, all the techniques used at this point of the screening are able to measure the affinity of these molecules for the proteins. To deeply characterize these small-molecule compounds, we decided to evaluate the ability of the compounds to prevent the interaction of the protein with the natural ligand (histone H4K20me2). For this purpose, we designed and developed an AlphaLISA assay, as it is not commercially available an optimized kit for methylation readers.

3.1.6. AlphaLISA

AlphaLISA is based on the measurement of fluorescence emission following the proximity of two types of beads. To date, it is not reported an optimized protocol for the study of the binding of reader domains of PHF20 with this technology, hence, it was necessary the design, development, and optimize the Alpha assay. Among the four different sequences of PHF20, the assay was developed for the protein GST-Tudor1-2. Regarding the beads, the assay was designed using streptavidin-coated donor beads able to interact with the biotinylated peptide substrate of PHF20 (H4K20me2) and acceptor beads coated with glutathione that interacts with the GST tag of the protein.

It is fundamental that the protein is stable and folded during the experiments, for this reason, it is necessary to determine the correct buffer and pH to use. For this purpose, first of all, the stability of the protein was tested in different buffers using the nanoDSF assay. In particular, the melting temperature (T_m) of the protein was measured in three different buffers (Tris, Hepes, Phosphate buffer) at different pH thus obtaining ten different conditions.

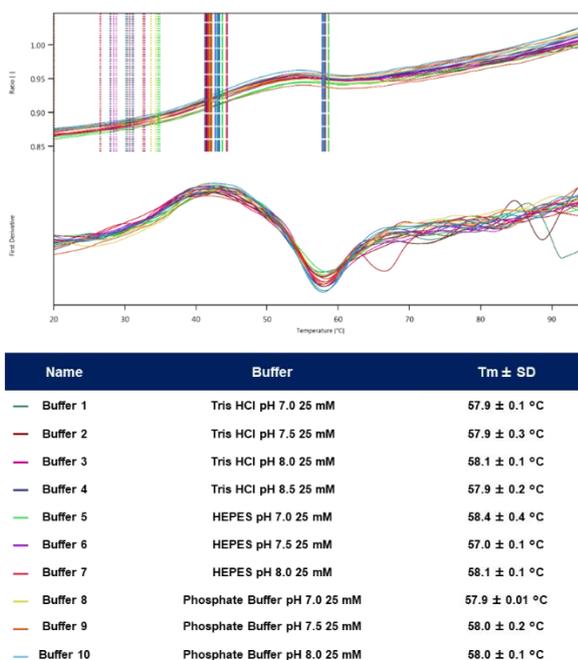


Figure 3.1.11 Determination of protein stability in different buffer and pH using nanoDSF method.

Figure 3.1.11 shows that there were no significant differences between the different buffers used, as T_m values are comparable. However, we decided to use buffer number 6 as Perkin Elmer recommends this buffer for interaction studies (25 mM Hepes pH 7.5, 100 mM NaCl, 0.1% Tween20).

After optimizing the assay buffer it was necessary to understand the best conditions in terms of protein and peptide concentration. First, it was performed a cross-titration with the aim to determine the optimal concentrations of peptide for the assay. In this experiment, while the concentrations of beads were kept constant, the concentration of the binding partners was varied. It is possible to see from the graph of Figure 3.1.12, that the best condition was reached using the peptide at 5 nM.

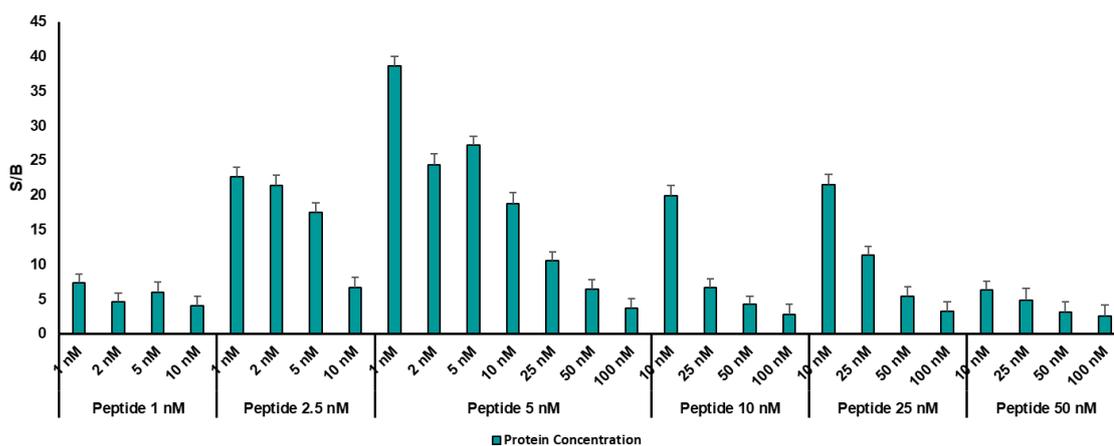


Figure 3.1.12 Cross-titration experiments for the binding of the protein with the biotinylated peptide H4K20me2, using Streptavidin Donor beads and Glutathione Acceptor beads.

At this point, to determine the optimal concentration of protein, it was performed a new experiment keeping constant the peptide concentration at 5 nM and using the protein concentration from 200 nM to 0.03 nM. The hook point was reached at the concentration of 1

nM of protein (Figure 3.1.13). In an AlphaLISA assay, to avoid beads saturation, it would be appropriate to use a concentration of protein lower than the one that results in the hook point.

For the above mentioned reasons, the optimized concentrations of protein and peptide to be used in the following assays were 0.5 and 5 nM respectively.

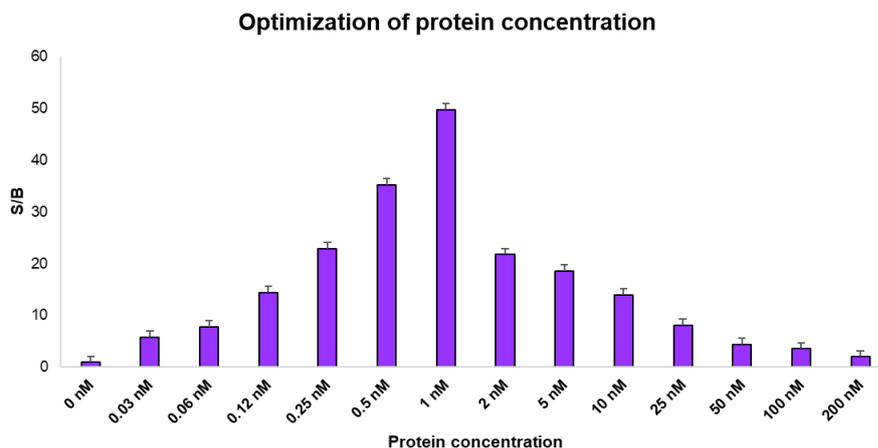


Figure 3.1.13 Optimization of protein concentration keeping constant the peptide concentration.

The incubation time is another parameters that should be investigated to have a good interaction, hence different incubation times were evaluated (30, 60, 90 and 120 min) and the results showed that the best signal-to-noise ratio was obtained with an incubation time of 30 min (Figure 3.1.14).

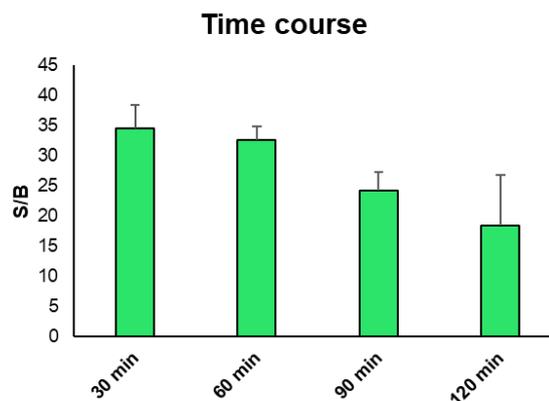


Figure 3.1.14 Variation of Alpha signal in dependence of the incubation time.

It was also necessary to establish the DMSO tolerance as the platform should screen small molecule compounds, so different concentrations of DMSO were tested starting from a concentration of 0.1% up to 5%. In all cases, it is possible to see a good response from the assay. (Figure 3.1.15)

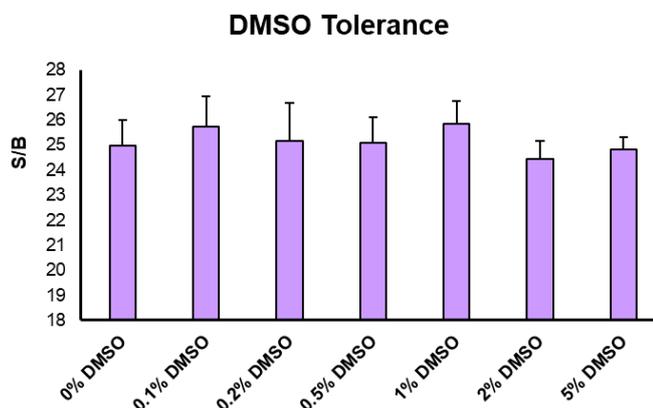


Figure 3.1.15 Determination of DMSO tolerance

Once determined the best conditions (assay buffer to use, protein and substrate concentrations and incubation time) the following step was the validation of the assay. For this reason, it was performed a displacement assay. This experiment was made using a fixed concentration of protein and biotinylated peptide and varying the concentration of a non-biotinylated peptide. It was found that the IC₅₀ is 27 nM. (Figure 3.1.16)

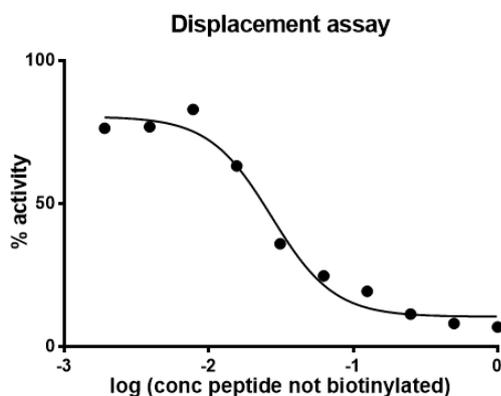
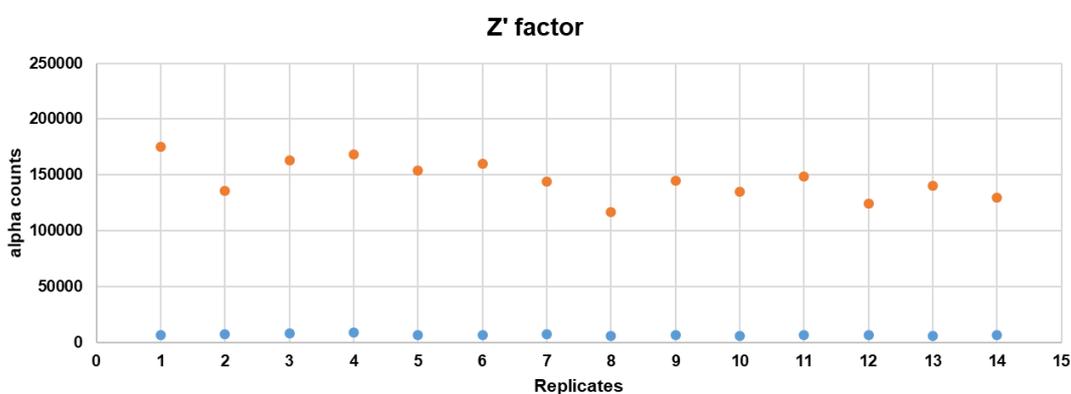


Figure 3.1.16 Displacement assay with not-biotinylated peptide

Finally, with the aim to verify the quality of the assay, statistical parameters were measured: signal-to-noise ratio (S/N), signal-to-background ratio (S/B) and the Z'-factor. To calculate these parameters the assay was performed in two conditions: blank and control (without and with protein respectively). The test was performed in duplicate and it was repeated independently for 7 days (Figure 3.1.17). The assay developed for PHF20 has a Z' factor of 0.61 (an assay is considered to be excellent if it has a Z'-factor between 0.5 and 1) and this means that it has good reproducibility, reliability and it can be used in drug discovery processes.



$$S/N = \frac{\text{mean signal} - \text{mean background}}{SD \text{ of background}}$$

$$S/B = \frac{\text{mean signal}}{\text{mean background}}$$

$$Z' = 1 - \frac{(3SD \text{ of positive control}) + (3SD \text{ of negative control})}{\text{mean of positive control} - \text{mean of negative control}}$$

- S/N= 171.4
- S/B= 21.1
- Z' factor= 0.61

Figure 3.1.17 Measurement of statistical parameters like Z'Factor, S/N and S/B.

Once determined the best conditions for the assay and established the reproducibility and quality of the method, the next step was the evaluation of the effect of UNC1215 analogs against the reader activity of the Tudor domain of PHF20. The activity of the compounds was preliminarily determined by conducting fixed-dose assays at 100 μ M. As shown in Figure 3.1.18, most molecules significantly inhibit protein-substrate interaction, therefore, we decided to measure the IC₅₀ for each molecule. (Table 3.4)

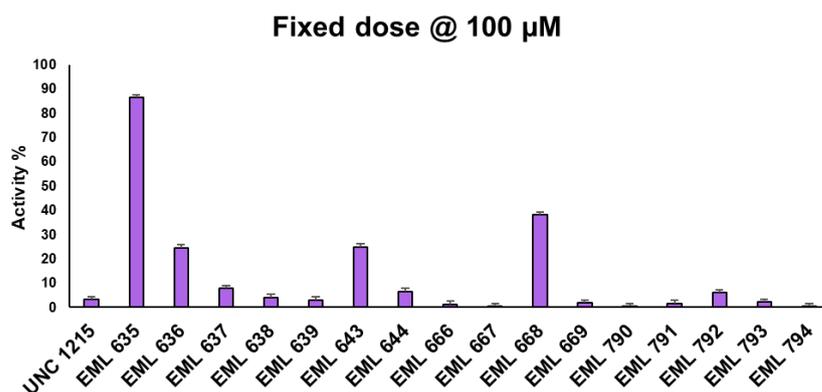


Figure 3.1.18 Fixed-dose assay performed on the library of compounds.

COMPOUNDS	IC ₅₀ \pm SD
UNC1215	2.5 \pm 0.4 μ M
EML635	40.6 \pm 4.2 μ M
EML636	13.1 \pm 1.9 μ M
EML637	51.3 \pm 6.2 μ M
EML638	2.1 \pm 0.7 μ M
EML639	5.3 \pm 0.9 μ M
EML643	26.8 \pm 3.3 μ M
EML644	3.0 \pm 1.1 μ M
EML666	4.1 \pm 0.8 μ M
EML667	9.6 \pm 2.8 μ M
EML668	35.8 \pm 5.1 μ M
EML669	8.2 \pm 2.4 μ M
EML790	1.3 \pm 0.6 μ M
EML791	0.3 \pm 0.1 μ M
EML792	7.4 \pm 2.7 μ M
EML793	4.9 \pm 1.1 μ M
EML794	0.6 \pm 0.2 μ M

Table 3.4 IC₅₀ values determined for the compounds of the library.

Considering the results obtained from this technique, we can say that all the molecules possess an IC₅₀ around low millimolar range, so all the molecules seem to be able to inhibit the reader activity of the target protein.

Here it is possible to understand the importance to use a screening platform. For example, if we had used only this technique to select the best molecules of the library surely we would have identified false positive. To clarify this consideration in the next paragraph I will describe the crosschecking of the results obtained from all the techniques of the screening platform.

3.1.7. Crosschecking

Finally, we compared all data obtained for the different sequences of PHF20. Considering the GST-Tudor1-2 protein, analyzing the results obtained using nanoDSF, AlphaLISA and MST we find that 8 compounds are positive in all the assays. Regarding the Tudor2 mutate protein and Tudor1-2, by cross-checking the nanoDSF, MST and SPR techniques, 7 and 5 compounds, respectively have been identified as real positives.

Combining all the results obtained for each protein in a Venn diagram (Figure 3.1.19), we can conclude that there are 3 compounds (EML667, EML790, and EML794) that can be considered hits, as these compounds bind all the three protein sequences. Unfortunately, it is difficult to make structure-activity relationships, because our set of molecules is too small. Nevertheless, expanding the library, in the future it will be possible to understand which chemical groups determine the affinity and the selectivity.

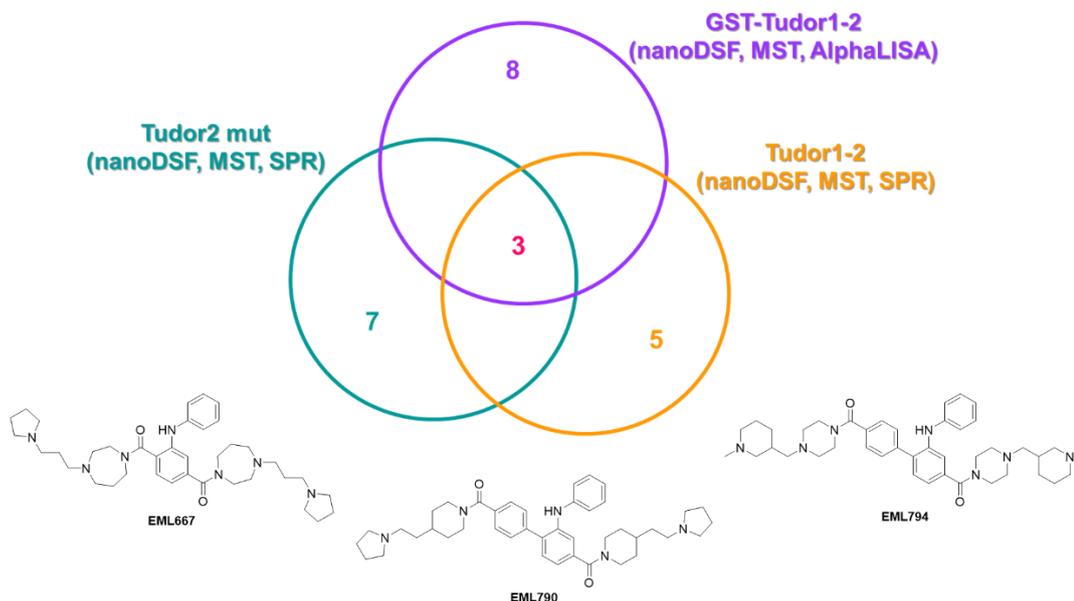


Figure 3.1.19: Venn diagram for the identification of the compounds that bind all three sequences of PHF20.

3.2 Spindlin1

3.2.1 Background

Spindlin1 is a chromatin reader protein that possesses three Tudor domains. In our previous work performed in collaboration with M. Bedford, we identified the compound EML405 as a new ligand of Spindlin1. On the basis of this result, we decided to test a small library of compounds (Figure 3.2.1) structurally related to EML638 (EML405 without the biotin) as new ligands of Spindlin1.

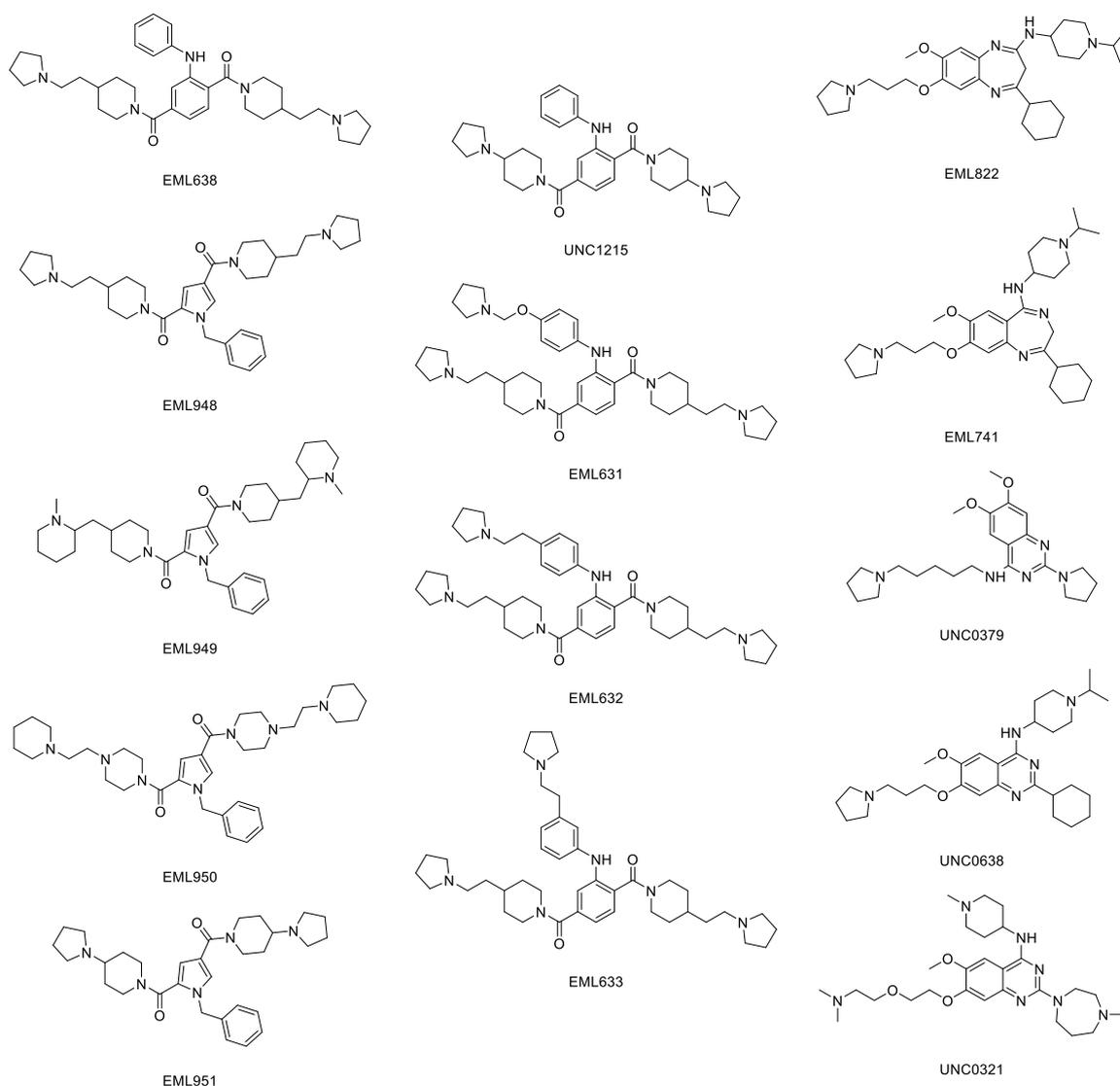


Figure 3.2.1 Small molecule library screened for Spindlin1 and MRG15

As performed for PHF20, Spindlin1 was expressed and purified in house. In addition, once the protein was obtained in a large amount and its quality was confirmed, an MST assay was optimized and validated for this target. These experiments were performed during my internship spent in NanoTemper, Munich.

3.2.2 Expression and purification of Spindlin1

Before proceeding with the evaluation of the binding of compounds, it has been necessary to produce proteins in large amounts. Also for this protein, it was used the carrier fusion protein GST as tag. The plasmid vector was transformed in BLR(DE3)pLysS and the expression of the protein was induced in *E.coli*. With the aim to define the optimal conditions for protein expression, several expression tests at different temperatures and times of growth were performed. The best condition was given using 100 μ M IPTG at 16°C overnight.

Subsequently, this condition was used for the scale-up, thus we proceeded with the purification of the proteins from the lysate using affinity chromatography, taking advantage of the ÄKTA purifier system. As it is possible to see in Figure 3.2.2 the protein is highly expressed using these conditions.

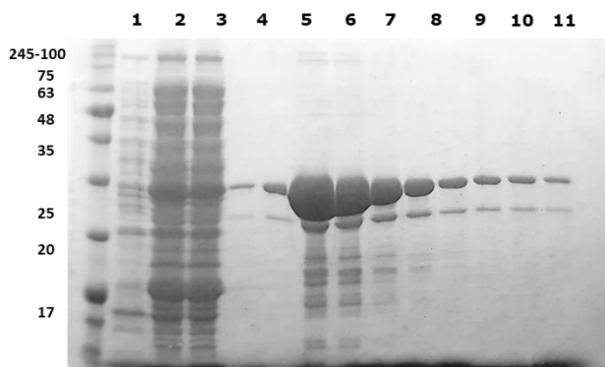


Figure 3.2.2 SDS-PAGE of the purification of Spindlin1 by affinity chromatography using GStrap HP column and an AKTA FPLC at 1 mL/min. Lane 1: before induction; Lane 2: lysate Lane 3 unbound fraction; Lanes 4-11: elution fractions

As it is possible to see from Figure 3.2.2 the protein was highly expressed, therefore, the following step was to check the protein stability. Also in this case it was exploited the nanoDSF performing a melting profile (Figure 3.2.3). It is clear that the protein is correctly folded, with a T_m value of 50°C demonstrating that the protein is stable.

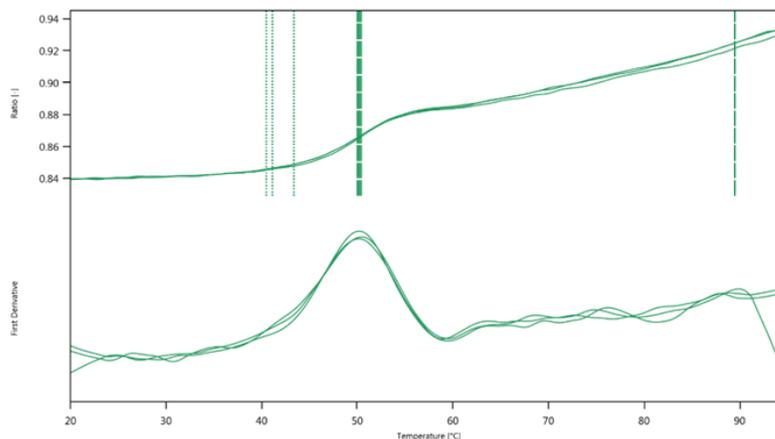


Figure 3.2.3 *Melting profile of GST-Spindlin1*

3.2.3 Design, optimization, and validation of MST assay for Spindlin1

The protein was labeled using the Monolith protein labeling kit RED-NHS 2nd generation and the buffer used for the elution was 50 mM Tris-Cl pH 7.8, 150 mM NaCl, 10 mM MgCl₂, 0.05% Tween (MST-T buffer). It was adopted the standard procedure reported in the labeling protocol which suggests a protein:dye ratio of 1:3.

Once labeled, the protein folding was checked comparing the melting profile of the labeled and unlabeled protein. Figure 3.2.4 shows that the two proteins have similar melting temperatures, confirming the correct folding of the protein after the labeling.

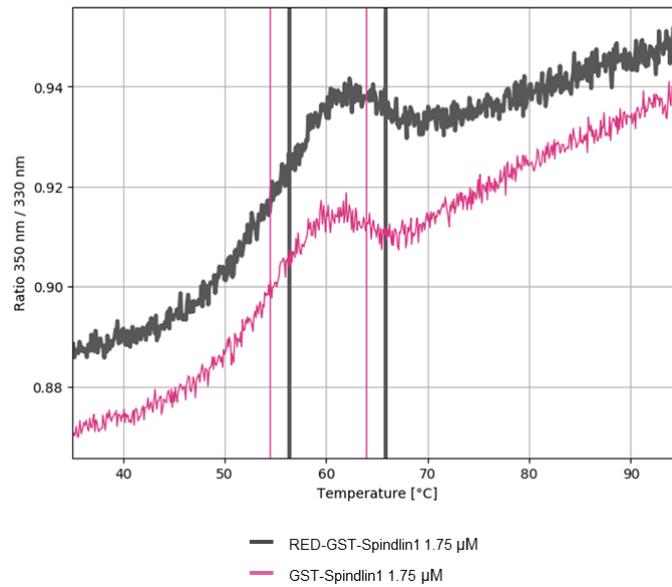


Figure 3.2.4 Comparison of the labeled and unlabelled protein. This experiment was performed using Tycho NT.6.

After the labeling procedure, it was measured the degree of labeling (DOL) which was 0.4, outside the optimal range (0.5-1). Nevertheless, a pretest was performed with the aim to check if the fluorescence intensity of the protein was high enough, if the protein aggregate or if there is sticking to the walls of the capillaries. This experiment was performed using the Monolith NT.115pico and final concentration of protein at 5 nM. The fluorescence was high enough (8000 counts) and it means that despite a low DOL, this labeled protein can be used for the following experiments. In addition, phenomena of sticking and aggregation were not present (Figure 3.2.5).

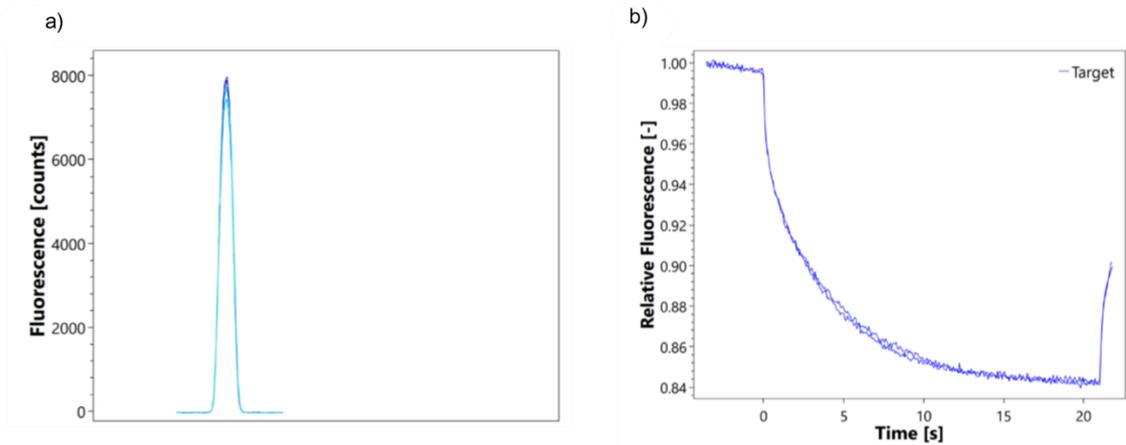


Figure 3.2.5 Pretests for *Spindlin1* after labeling. **a)** Capillary scan and **b)** MST traces.

To optimize the experimental conditions, different tests were performed changing the type of capillaries and the buffer. In Figure 3.2.6 it is reported the MST profile of the protein in MST-T buffer, in premium and standard capillaries. As it is clear in Figure 3.2.6, the use of standard capillaries caused adsorption and aggregation, therefore, the following experiment were performed in premium coated capillaries.

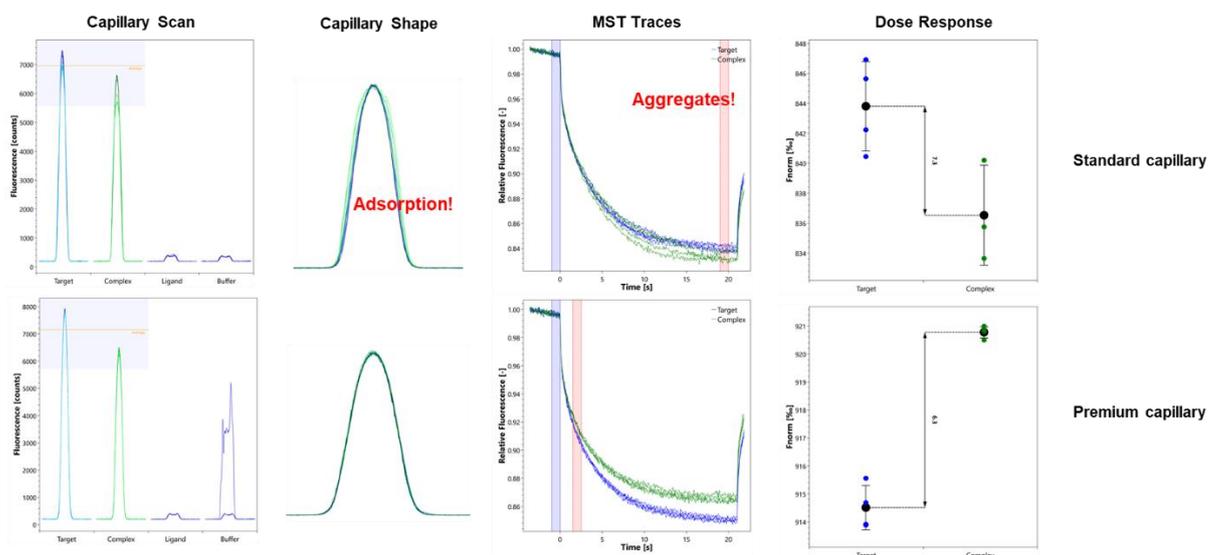


Figure 3.2.6 Assay optimization for *SPIN1*: binding check of *UNC1215* using standard and premium capillaries

3.2.4 Assay validation

Before evaluating the interaction with the molecules, it was important to validate the screening method by checking the interaction of the protein with the reference compound UNC1215. The affinity of the protein for UNC1215 was determined maintaining constant the concentration of the protein while the inhibitor (at maximum concentration of 400 μM) was serially diluted. The affinity of UNC1215 for SPIN1 was $21.3 \mu\text{M} \pm 4.6$. (Figure 3.2.7).

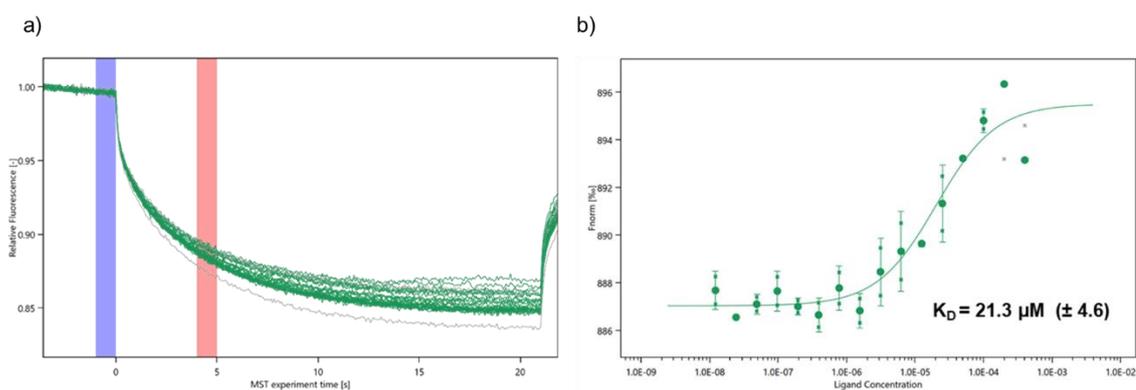


Figure 3.2.7 Binding affinity experiment between *Spindlin1* and UNC1215 **a)** MST traces and **b)** Dose-response.

3.2.5 Small library screening

Once optimized and validated the assay, the putative modulators of Spindlin1 were first tested in a single point screening. The molecules were tested in duplicate at a fixed dose of 100 μM together with DMSO as negative control (2%) and UNC1215 as positive control. Then, applying a statistical differentiator of three standard deviations, the potential binders were selected. Among 17 compounds, five molecules behave as binders; hence these molecules were titrated starting from 400 μM against the target protein and the K_D values were measured (Figure 3.2.8).

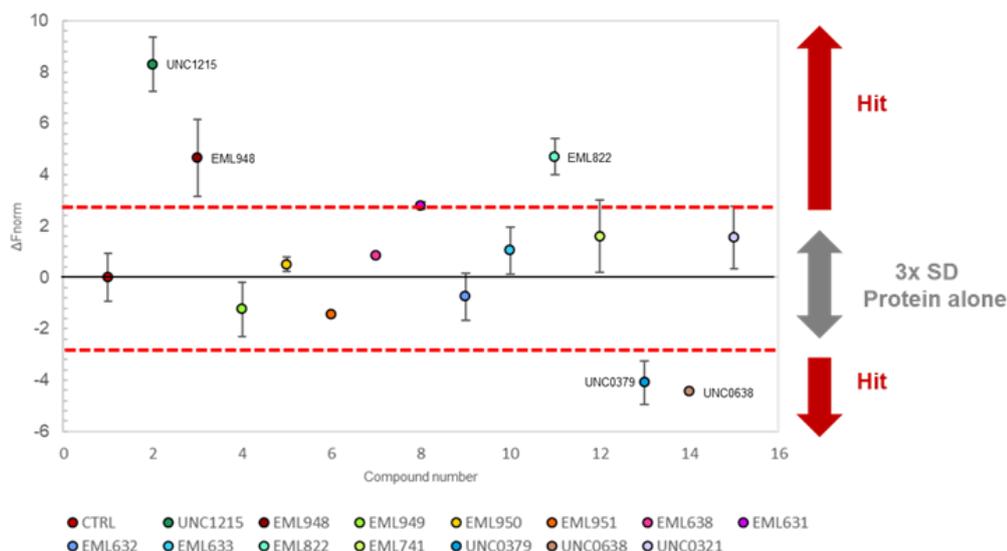


Figure 3.2.8 Single point screening, *Spindlin1* vs. inhibitors;

Compound	K_D Medium MST power 5 sec MST on time
EML948	2.4 $\mu\text{M} \pm 0.4$
UNC1215	21.3 $\mu\text{M} \pm 4.6$
UNC0379	33.8 $\mu\text{M} \pm 10$
UNC0638	40 $\mu\text{M} \pm 11$

Table 3.5 K_D values of compounds

Looking at the K_D values reported in Table 3.5 it is clear that all the selected compounds have a binding affinity in the micromolar range. Clearly, the most interesting compound of the set is the EML948 which possesses a higher binding affinity compared to the other compounds. These results are still preliminary as for this protein we used only one method for the screening of the compounds. The application of the other techniques is still ongoing in the laboratory.

3.3 MRG15

3.3.1 Background

MRG15 is a transcription factor involved in embryonic development, cell proliferation, and cellular senescence. The same approach used for the protein Spindlin1 was applied also to this protein, with the aim to identify new ligands. Hence, the protein was first of all expressed and purified, then the screening of the small molecule compounds was performed using MST during my internship at NanoTemper technologies.

3.3.2 Expression and purification of MRG15

The plasmid vector was transformed in BLR(DE3)pLysS and the expression of the protein was induced in *E.coli* as a GST-tagged protein. The optimal conditions for protein expression were defined, performing expression tests at different temperatures (25°C and 37°) and times of growth (2h, 4h, 6h and overnight). The best results were achieved using for the induction 100 µM IPTG at 37°C overnight.

After the determination of the best expression conditions, the purification of the protein was performed from the lysate using affinity chromatography, taking advantage of the ÄKTA purifier system. The purification was confirmed *via* SDS-PAGE (Figure 3.3.1)

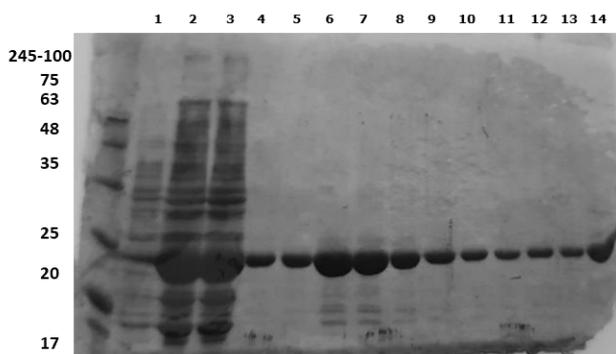


Figure 3.3.1 SDS-PAGE of the purification of MRG15 by affinity chromatography using GStrap HP column and an AKTA FPLC at 1 mL/min. Lane 1: before induction; Lane 2:lysate Lane 3 unbound fraction; Lanes 4-11: elution fractions

The stability of the protein was verified doing a melting profile with nanoDSF. These experiments were performed using standard treated capillary, with a temperature range from 20 °C to 95 °C at a heating rate of 1 °C per minute.

Looking the curves in Figure 3.3. it is clear that the protein is correctly folded, the only T_m value detected is around 55°C hence the protein is stable.

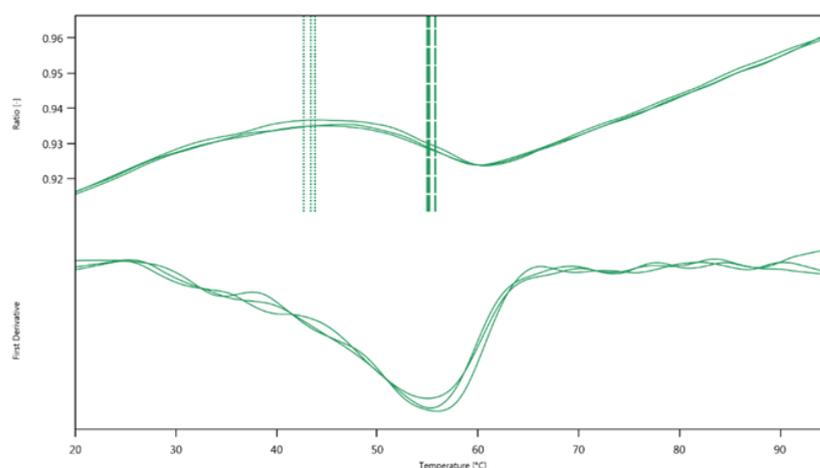


Figure 3.3.2 *Melting profile of GST-MRG15*

3.3.3 *Design, optimization, and validation of MST assay for MRG15*

3.3.3.1 *Protein labeling*

Before to start with MST experiments, for this protein it was necessary to perform several labeling reactions to achieve a good result. Also for this protein, for the labeling it was exploited the presence of the lysines, hence the Monolith protein labeling kit RED-NHS 2nd generation was used.

First, we used the standard labeling procedure (1:3 protein:dye ratio) with a protein concentration of 10 μM, the pretest was performed but the initial fluorescence was too low (3.3.3a)

For this reason, we decided to increase the protein:dye ratio to 1:5 and the protein concentration to 20 μM . Looking at the thermal stability (3.3.3b), it was clear that the labeling affected negatively the protein folding. In addition, photo-enhancement phenomena were detected during the experiments, which are common when protein:dye ratio is too high.

Finally, the protein was labeled maintaining the concentration at 20 μM but the protein-dye ratio was reduced to 1:3. The protein stability was checked and a pretest was performed. As it is showed in Figure 3.3.3 c, the labeling procedure did not change the protein integrity, and this data was confirmed by a degree of labeling of 0.6 therefore, this preparation of labeled protein was used for the following experiments.

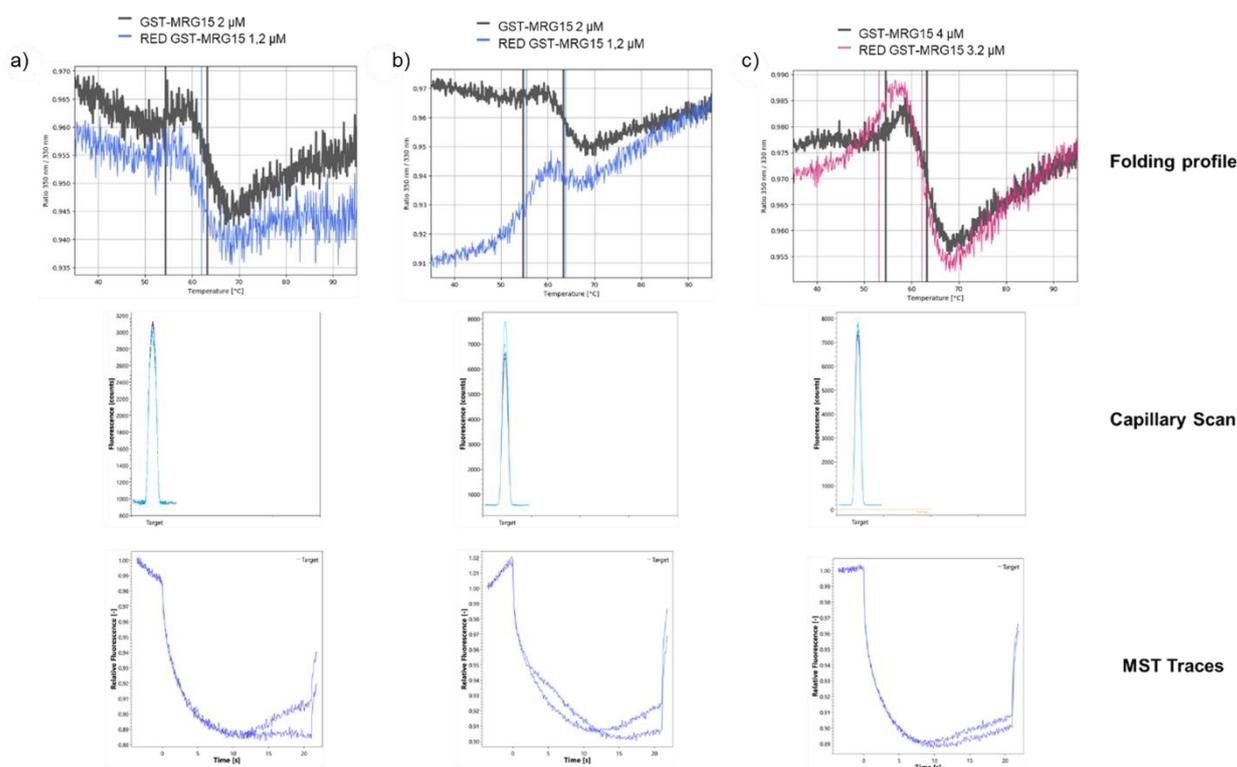


Figure 3.3.3 Folding profile, capillary scan and MST traces of MRG15 labeled using **a)** protein-dye ratio 1:3 and 10 μM as starting concentration; **b)** protein-dye ratio 1:5 and 20 μM as starting concentration; **c)** protein-dye ratio 1:3 and 20 μM as starting concentration;

3.3.3.2 Buffer optimization

Determined the optimal protein-dye ratio to use, we wanted to understand the best buffer to use in this assay. The buffer optimization was conducted performing a single point screening with and without the reference UNC1215 in different buffers (MST-T; PBS including 0.05% of Tween; 50 mM Tris-Cl pH 8.0, 50 mM NaCl, 0.05% Tween). The Amplitude and the standard deviation of the normalized fluorescence were analyzed. Checking the amplitude in Figure 3.3.4 it is clear that the best condition is reached using PBS-T as assay buffer.

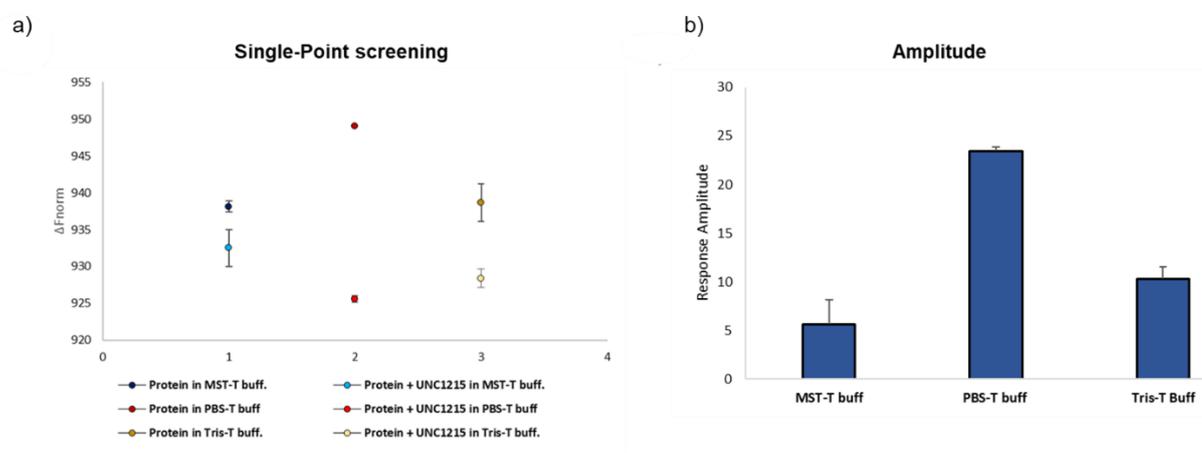


Figure 3.3.4; a) Single point screening and b) amplitude comparison used for the buffer optimization;

Nevertheless, with the aim to improve the reproducibility of the MST signal, we decided to evaluate the eventual effect of the BSA. The addition of the BSA (0.05 mg/mL) to the assay buffer markedly improved MST signals (Figure 3.3.5) so it was used as a stabilizing agent in the following experiments.

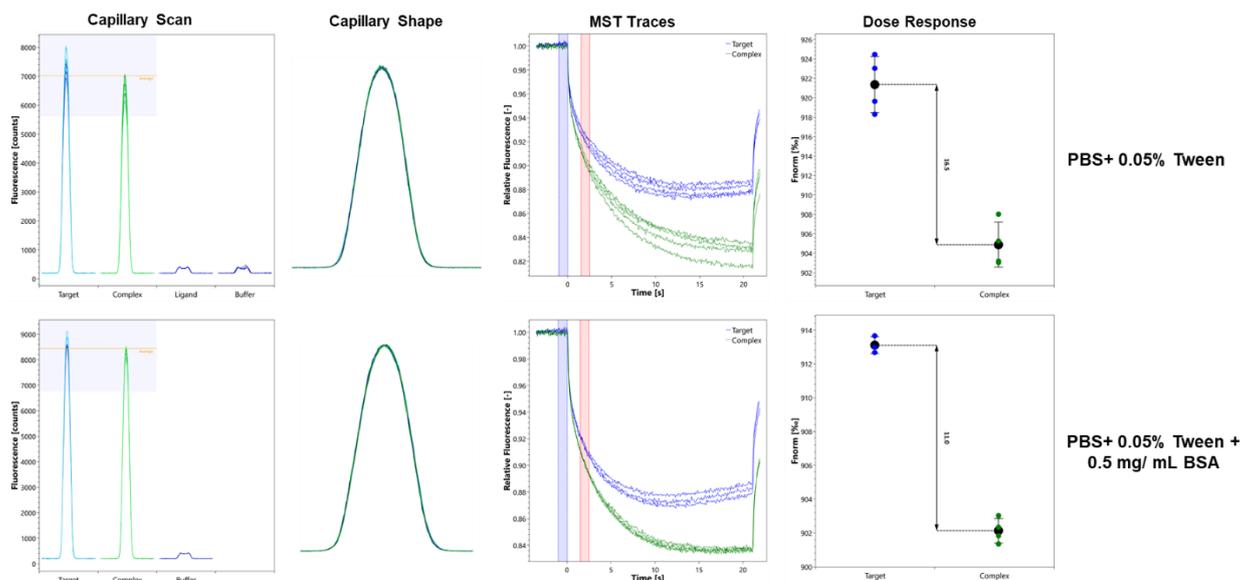


Figure 3.3.5 Improvement caused by adding BSA in the assay buffer

3.3.3.3 Assay validation

After the optimization of all the experimental conditions, we decided to validate the assay measuring the binding affinity between MRG15 and the reference compound UNC1215.

This experiment was performed diluting the compound starting from 400 μM and the labeled protein was kept constant at 5 nM as final concentration. Before the MST measurement, samples were incubated for 10 minutes at RT. The affinity of UN1215 for Spindlin1 was estimated at 43.1 nM \pm 5.3. (Figure 3.3.6)

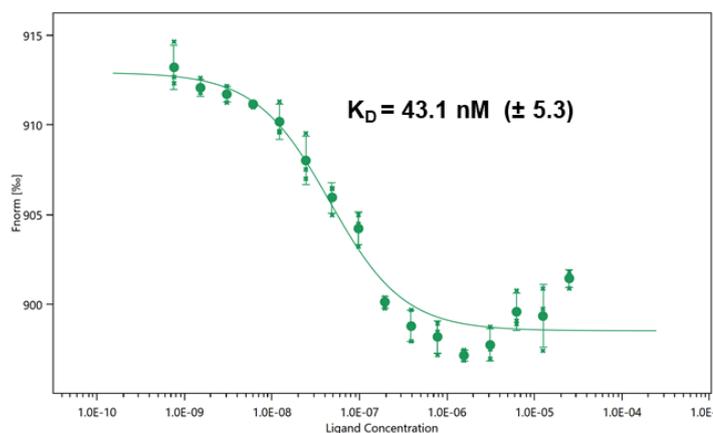


Figure 3.3.6 Binding affinity of MRG15 for UNC1215

3.3.4 Small library screening

Once optimized and validated the assay, finally it was possible to proceed to the screening of the library.

Initially, it was performed a single-dose screening with the aim to select the putative binders of MRG15. All the compounds were tested in duplicate at a fixed dose of 100 μ M and it was applied a statistical cut-off of three standard deviations. From this experiment eleven molecules behave as binders (Figure 3.3.7).

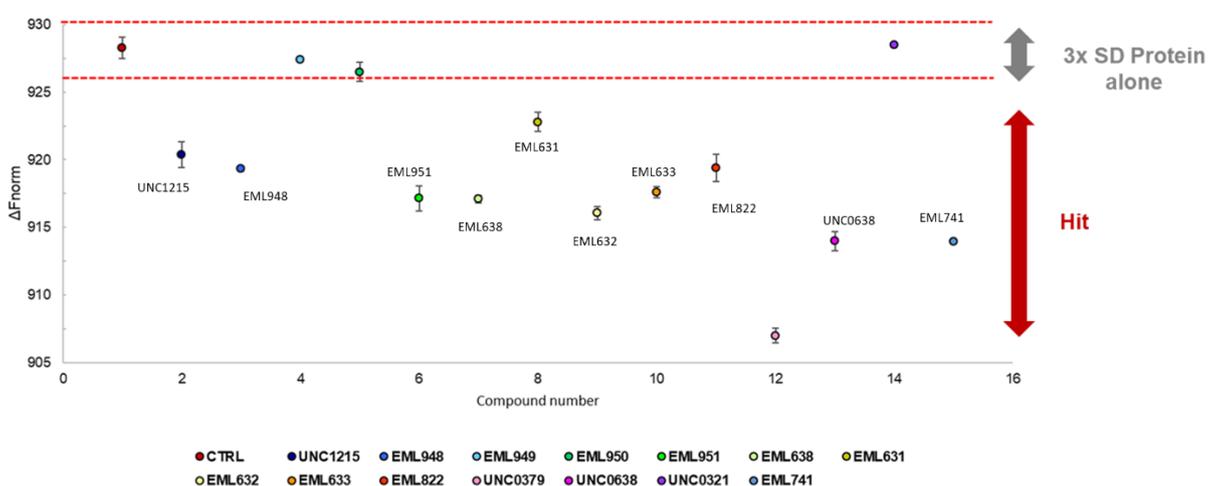


Figure 3.3.7 Single-point screening of the compounds against MRG15

To determine the K_D values, the compounds were titrated starting from 400 μ M and the labeled protein was kept constant at 5 nM as final concentration. Before the MST measurement, samples were incubated for 10 minutes at RT. The binding curves and the K_D values are reported in Figure 3.3.8.

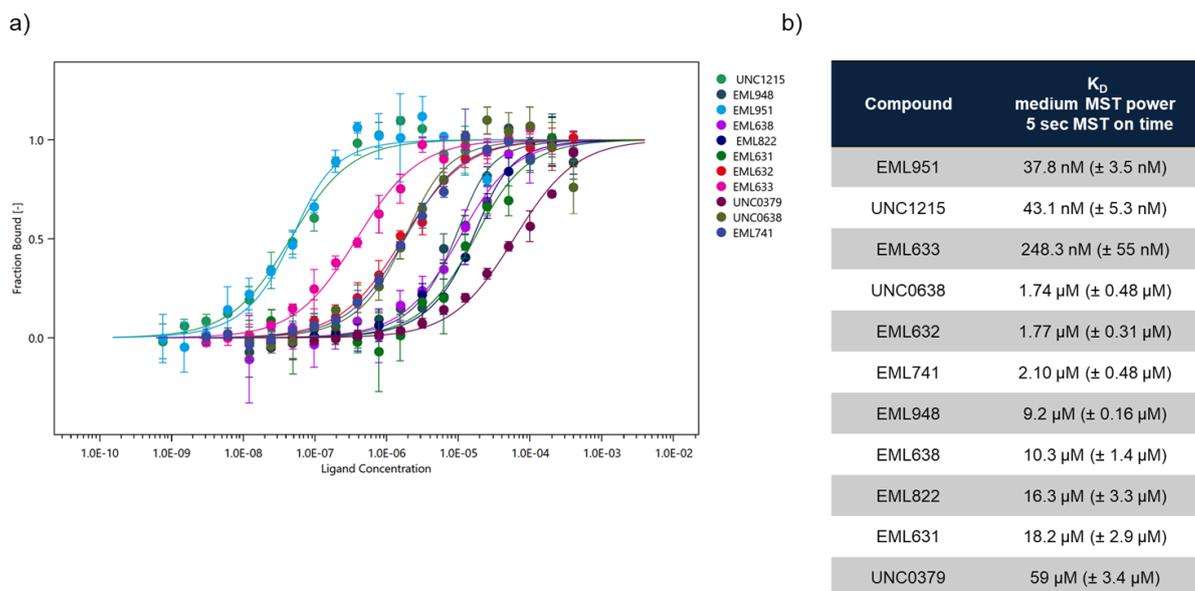


Figure 3.3.8 a) Binding curves of the compounds for MRG15; b) K_D values calculated for each interaction.

Looking at the K_D values it is possible to say that there are several compounds like EML948, EML663 and the reference UNC1215 that possess a nanomolar affinity, while there are other compounds (EML632, EML741, EML948, EML638, EML822, EML631, UNC0638, and UNC0379) that bind the target with a micromolar affinity. In addition, the most interesting molecule of this set is the EML948 that is able to bind MRG15 with a higher affinity than UNC1215. Also regarding MRG15, these results are still preliminar, we used only one method for the screening of the compounds. The application of the other techniques is still ongoing in the laboratory.

3.4 GST tag interference check

Considering that both proteins (SPIN1 and MRG15) are GST tagged, it was possible to think about an aspecific interaction of the compounds with the GST. To confirm the reliability of our data, the MST results obtained for both proteins were cross-checked, and only the compound EML948 was able to interact with both proteins (Figure 3.4.1). For this compound it was evaluated the interaction with the GST tag.

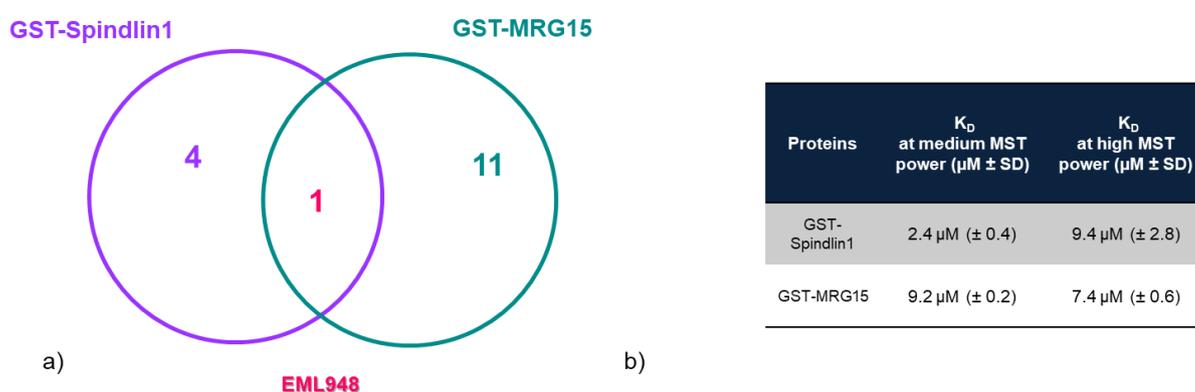


Figure 3.4.1 a) Venn diagram of Spindlin1 and MRG15 binders; b) K_D values of EML948 against Spindlin1 and MRG15.

In this case, considering that the concentration of the GST was too low to perform a labeling procedure, we decided to perform the binding experiments without the labeling of the protein, using the Monolith label-free.

Initially, it was checked the thermal stability of the GST tag via nanoDSF (Figure 3.4.2).

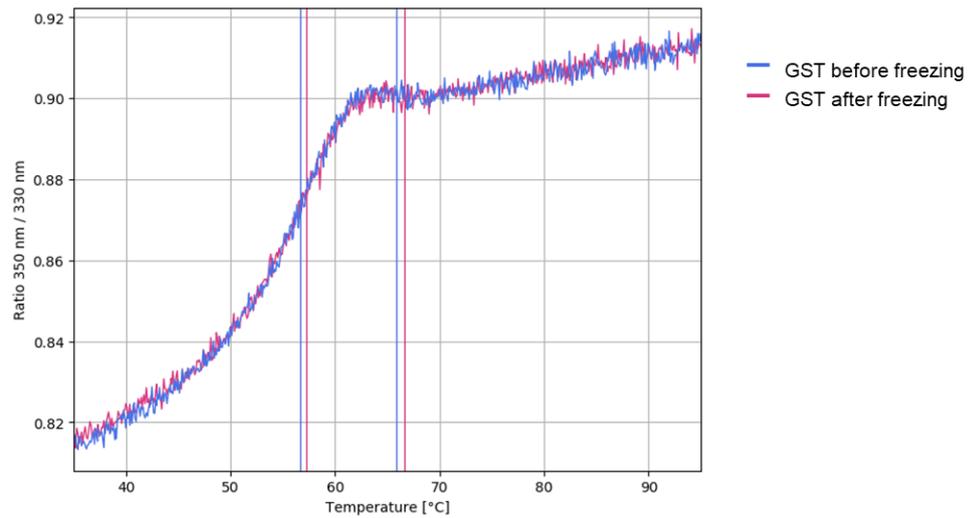


Figure 3.4.2 Comparison of the frozen and thawed GST tag. This experiment was performed through Tycho NT.6. with a temperature range from 35 °C to 95 °C at a heating rate of 1 °C per minute.

Subsequently, it was performed a pretest with the aim to understand if the protein was fluorescent enough to perform the measurements and if the compound ELM948 and the assay buffer interfere with the analysis (Figure 3.4.3)

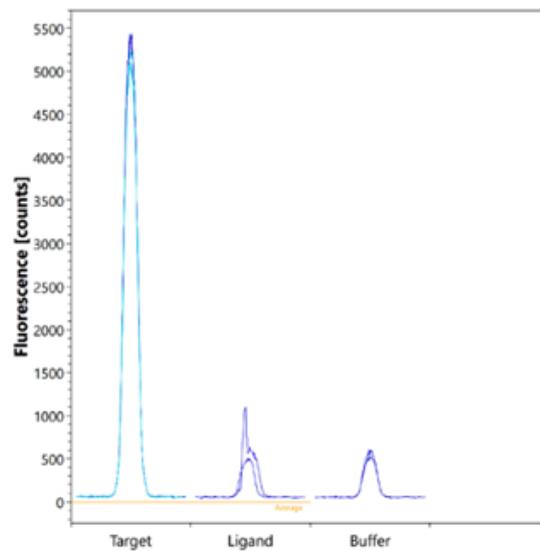


Figure 3.4.3 Pretests of GST tag for label-free measurements.

Once confirmed that the fluorescence of the protein was good and that the molecule and the assay buffer do not interfere with the analysis, it was measured the binding affinity between GST-tag and EML948.

This latter experiment was performed maintaining constant the concentration of the protein while the compound (at maximum concentration of 100 μM) was serially diluted. Figure 3.4.4 show that no binding was detected. This result demonstrates that the interaction of the compound is not related to the tag hence it is possible to conclude that EML948 is a real binder for both proteins.

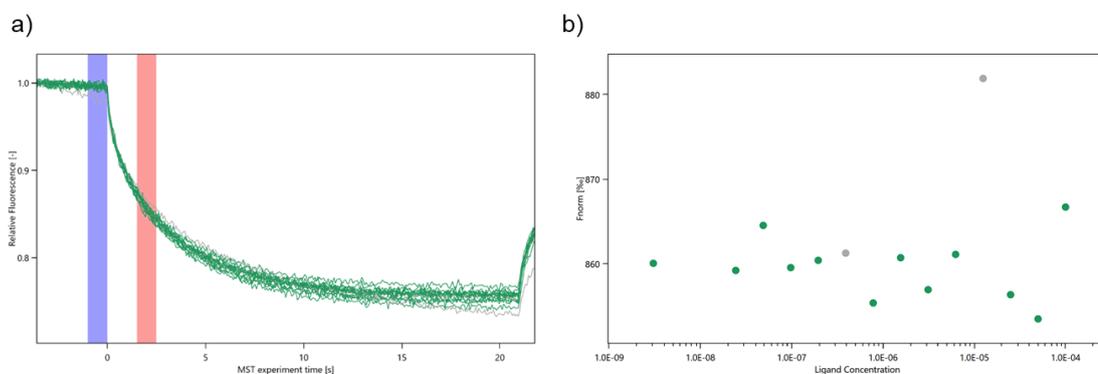


Figure 3.4.4 a) *MST traces* and b) *Dose-response experiments between GST and EML948*

4. CONCLUSIONS

Epigenetic proteins, playing a key role in all the biological processes, if aberrantly expressed or altered, are usually linked to the outbreak of several diseases.

Hence, considering the influence of these proteins in physiological and pathological processes, great efforts have been done in the drug discovery with the aim to develop small molecule modulators of chromatin-modifying proteins. Nevertheless, the number of reported modulators is limited and refers only to specific classes of epigenetic enzymes, like epigenetic “writers” and “erasers”.

So far, several studies are in progress to identify small molecule chemical probes for epigenetic readers but this research area is affected by several problems, one of the most important is the low availability of robust and reliable screening methods to investigate the activity of these molecules.

For this reason, this work was focused on the development of a versatile screening platform composed by biochemical and biophysical methods. Using orthogonal techniques, organized in a modular fashion, it is possible to overcome the limitations of each technique, allowing the identification of small molecule modulators for methylation reader proteins.

In particular, the platform is composed by the nanoDSF, the Microscale thermophoresis, the Surface plasmon resonance, and the Alpha technology. This multiple approach was applied to three different epigenetic readers, PHF20 and Spindlin1 that contain Tudor domains and MRG15 that contain one chromodomain.

Regarding PHF20, a small library of compounds synthesized in EMCL was screened using different sequences of this protein with the aim to understand the effect caused by different

factors as the presence of the GST tag, the presence of a not functional domain and the eventual differences between the monomeric and dimeric forms of the protein. First, several experiments were performed to optimize the expression and the purification of the proteins. In order to check the stability of each protein, nanoDSF experiments were performed and the sequence Tudor2 wild type, as unstable, was discarded from the study. Thus, the screening platform was applied on three sequences of PHF20 (GST-Tudor1-2, Tudor1-2, His-Tudor2 mutate).

As a primary screening technique, it was used the nanoDSF. Nevertheless, giving only qualitative results, the nanoDSF was coupled with MST, a technique that allows the determination of the K_D . To confirm the data obtained, it was also used the SPR, an orthogonal method. In this case we decided to exclude the GST tagged sequence from the analysis because the tag could hide the binding site.

Finally, to evaluate the ability of the compounds to inhibit the reader activity of PHF20, it was designed, developed, and validate an AlphaLISA assay.

At this point, crossing the result obtained from each technique and each sequence of PHF20 it was possible to identify three true hits, **EML667**, **EML790**, and **EML794**.

Regarding Spindlin1 and MRG15, also in this case, the work started with the expression and purification of the proteins. Once obtained and verified the quality of the proteins, they were used for MST experiments.

Several experiments were performed in order to find the optimal experimental conditions to be used in the screening of compounds (protein-dye ratio, assay buffer, and type of capillary to use). Optimized these conditions, a new library of compounds was screened for these two proteins. First, it was performed a single point screening to select the putative binders and subsequently the binding affinity was measured. Several compounds were identified as modulators of these two proteins, with an affinity between low nanomolar and micromolar.

So far, the binding experiments with the other techniques are still in progress, hence only after obtaining all the results it will be possible to confirm the identification of real hits for these two proteins.

In conclusion, this multiple approach proved to be widely usable because of its robustness, flexibility, and reliability and for sure the application of this platform would provide new opportunities to increase drug discovery success and productivity in the epigenetic field.

5. MATERIALS AND METHODS

5.1 PHF20

5.1.1 *Expression and purification of GST Tudor1-2*

The vector, generously provided by M. Bedford, was transformed into BLR(DE3)pLysS competent cells and the protein was overexpressed in lysogeny broth (LB) in the presence of 100 µg/mL ampicillin. Cells were grown at 37 °C to OD600 of 0.8 and induced by isopropyl-1-thio-D-galactopyranoside (IPTG, 100 µM). After 4 h at 37 °C, cells harbouring the expressed protein were pelleted at 5000g for 15 min at 4 °C and resuspended in 40 mL of Lysis Buffer (PBS buffer pH 7.4), adding protease inhibitors cocktail (Sigma). Cells were lysed using a sonicator (Vibra-Cells, Sonics) and cell debris were pelleted at 10000g for 30 min at 4 °C. The clarified lysate was filtered using a 0,45 µm syringe filter and loaded onto a 1-mL GSTrap HP column (GE Healthcare, # 17-5281-01) using an AKTA FPLC (GE Healthcare) at the flow rate of 1 mL/min. GST-protein was eluted using the elution buffer (Tris HCl 50 mM, glutathione reduced 10 mM, pH 8.0) over 10 column volumes. Fractions containing the protein were confirmed by sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE) and then pooled.

5.1.2 *GST cleavage and purification of Tudor1-2*

The cleavage of GST was performed using ThrombinCleanCleavage Kit (Sigma, # RECOMT), following the procedure of the technical datasheet. Briefly, 100 µL aliquot of a 50% (v/v) suspension of thrombin agarose resin was centrifuged at 500g and washed three times with 500 µL of Cleavage Buffer 1X, afterwards the resin was resuspended in 100 µL of

Cleavage Buffer 10X, 1 mg of the protein was added and the final volume of 1 mL was reached with water. The cleavage reaction was incubated at room temperature, with gentle agitation to keep beads suspended, for 6h. Analysis of cleavage reaction was performed by SDS-PAGE. The cleaved protein was then separated from the GST by anionic exchange chromatography using a Mono Q™ column (GE Healthcare, # 17-5166-01): a 1-mL bed volume of Mono Q™ 5/50 GL was equilibrated in buffer A (Tris-HCl 20 mM pH 8.0); the cleaved protein was loaded onto the column at 1 mL/min. The column was washed with 5 column volumes of buffer A and eluted with buffer B (Tris-HCl 20 mM, NaCl 1M, pH 8.0), allowing the separation of the native protein from the cleaved GST.

5.1.3 Expression and purification of His-Tudor2 wild type

The vector was transformed into Rosetta(DE3)pLysS competent cells and the protein was overexpressed in lysogeny broth (LB) in the presence of 100 µg/mL ampicillin. Cells were grown at 37 °C to OD600 of 0.6 and induced by isopropyl-1-thio-D-galactopyranoside (IPTG, 10 µM) and incubated for 6h at 15 °C. Cells harbouring the expressed protein were pelleted at 5000g for 15 min at 4 °C and resuspended in 40 mL of Lysis Buffer (20 mM Phosphate buffer, 200 mM NaCl, 30 mM Imidazole, pH 7.4), adding protease inhibitors cocktail (Sigma). Cells were lysed using a sonicator (Vibra-Cells, Sonics) and cell debris were pelleted at 10000g for 30 min at 4 °C. The clarified lysate was filtered using a 0,45 µm syringe filter and loaded onto a 1-mL HisTrap HP column (GE Healthcare, # 17-5247-01) using an AKTA FPLC (GE Healthcare) at 4 °C and with the flow rate of 1 mL/min. His Tudor 2 wild type was eluted using the elution buffer (Phosphate buffer 20 mM, NaCl 200 mM, Imidazole 500 mM, pH 7.4) over 10 column volumes. Fractions containing the protein were confirmed by sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE) and then pooled.

5.1.4 Expression and purification of His-Tudor2 mutate

The vector was transformed into Rosetta(DE3)pLysS competent cells and the protein was overexpressed in 1 L of LB at 37 °C by the addition of 0.1 mM IPTG when the OD₆₀₀= 0.8. Leaving the cell culture overnight at 15 °C, cells harbouring the expressed protein were pelleted at 5000g for 15 min at 4 °C and resuspended in 40 mL of Lysis Buffer (Phosphate buffer 20 mM, NaCl 200 mM, Imidazole 30 mM, pH 7.4), adding protease inhibitors cocktail (Sigma). Cells were lysed using a sonicator (Vibra-Cells, Sonics) and cell debris were pelleted at 10000g for 30 min at 4 °C. The clarified lysate was filtered using a 0.45 µm syringe filter and loaded onto a 1-mL HisTrap HP column (GE Healthcare, # 17-5247-01) using an AKTA FPLC (GE Healthcare) at flow rate of 1 mL/min. His Tudor 2 wild type was eluted using the elution buffer (Phosphate buffer 20 mM, NaCl 200 mM, Imidazole 500 mM, pH 7.4) over 10 column volumes. Fractions containing the protein were confirmed by sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE) and then pooled.

5.1.5 nanoDSF experiments

Protein thermal stability was measured in a label-free fluorimetric analysis using the Prometheus NT.48 (NanoTemper Technologies). Briefly, the shift of intrinsic tryptophan fluorescence of proteins upon temperature-induced unfolding was monitored by detecting the emission fluorescence at 330 and 350 nm. Thermal unfolding was performed in nanoDSF grade standard glass capillaries (NanoTemper Technologies) at a heating rate of 0.5 °C per minute in a range from 20 °C to 95 °C, using 20 µM of protein and 100 µM of compounds. Protein melting points (T_m) were calculated from the first derivative of the ratio of tryptophan emission intensities at 330 and 350 nm.

5.1.6 *Microscale thermophoresis*

5.1.6.1 *GST-Tudor1-2 labeling*

Fluorescence labeling of GST-Tudor1-2 was performed following the protocol for *N*-hydroxysuccinimide (NHS) coupling of the RED-NHS 1st generation dye (MO-L001, NanoTemper Technologies) to lysine residues. Briefly, 100 μ L of a 10 μ M solution of protein in labeling buffer (NaHCO₃ 130 mM, NaCl 50 mM, pH 8.2) was mixed with 100 μ L of 30 μ M 1st generation dye in labeling buffer and incubated for 30 min at room temperature (r.t.). Unbound fluorophores were removed by size-exclusion chromatography with MST buffer as running buffer.

Prior to MST experiment, aliquots of NT647-His Tudor2 mutata were thawed on ice and centrifuged for 15 min at 4 °C and 12000g to remove protein aggregates.

5.1.6.2 *Tudor1-2 labeling*

Tuor1-2 was labeled following the protocol for *N*-hydroxysuccinimide (NHS) coupling of the RED-NHS 1st generation dye (MO-L001, NanoTemper Technologies) to lysine residues. Briefly, 100 μ L of a 10 μ M solution of Tudor1-2 protein in labelling buffer was mixed with 100 μ L of 30 μ M NT647-NHS fluorophore (NanoTemper Technologies) in labelling buffer and incubated for 30 min at room temperature (RT). Unbound fluorophores were removed by size-exclusion chromatography with MST buffer as running buffer (Tris 50 mM, NaCl 150 mM, MgCl₂ 10 mM, pH 7.8) as running buffer.

Prior to MST experiment, aliquots of NT647-His Tudor2 mutata were thawed on ice and centrifuged for 15 min at 4 °C and 12000g to remove protein aggregates.

5.1.6.3 Tudor2 mutate labeling

His-Tudor2 wild type was labeled following the protocol for tris-Nitrilotriacetic acid (tris-NTA) coupling the RED-NT-647dye (MO-L008, NanoTemper Technologies) to His tagged proteins. The procedure is the following: 100 μ L of a 200 nM solution of protein in PBS buffer, 0.05 % Tween 20, pH 7.4 was mixed with 100 μ L of 100 nM 1st generation dye in the same buffer used for the protein and incubated for 30 min at room temperature (r.t.). For storage, aliquots of labeled His-Tudor2 mutate were frozen in 10 μ L aliquots at -80 °C. Prior to MST experiments, aliquots of proteins were thawed on ice and centrifuged for 15 min at 4 °C at 12000 g to remove protein aggregates.

5.1.6.4 Assay development for MST screening

The MST experiments were performed on the Monolith NT.115Pico instrument using 5 nM protein concentration, premium coated capillaries, 20% LED power. All the data were analyzed at medium MST power.

For the binding curves with the all the small molecules, compounds stock (20 mM) in DMSO were serially diluted in assay buffer starting from 200 μ M, keeping constant the concentration of DMSO (1%). Each dilution was mixed 1:1 with a solution of several concentration of labeled proteins (1nM for PHF20, 5nM for Tudor1-2 and His-Tudor2 mutate) to yield a final volume of 20 μ L. The reaction mixtures were incubated 10 minutes at room temperature and loaded into premium-coated capillaries.

K_D values were calculated from compound concentration-dependent changes in normalized fluorescence (F_{norm}) of proteins after 2.5 s of thermophoresis. Each binding experiment was performed in triplicate and the K_D values obtained are the mean of these three different

experiments. Data were analyzed using MO Affinity Analysis software (NanoTemper Technologies).

5.1.7 *Surface plasmon resonance*

SPR analyses were performed on a Biacore 3000 optical biosensor equipped with research-grade CM5 sensor chips (Biacore AB). The sequences Tudor1-2- and His-Tudor2 were used in this analysis. Proteins (30 $\mu\text{g}/\text{mL}$ in potassium acetate 10 mM, pH 4.5) were immobilized on individual flow cells of the sensor chip at a flow rate of 10 $\mu\text{L}/\text{min}$ using standard amine coupling protocols to obtain densities of 7–8 kRU. Myoglobin was used as a negative control, and one flow cell was left empty for background subtraction. All compounds were dissolved in DMSO (100%) to obtain 20 mM solutions, and they were diluted in PBS with 0.005% Tween20 while maintaining a final 0.1% DMSO concentration. Binding experiments were performed at 25 °C using a flow rate of 30 $\mu\text{L}/\text{min}$ with 120 s association monitoring and 200 s dissociation monitoring.

Surface regeneration was performed when necessary by a 10 s injection of 5 mM NaOH. The simple 1:1 Langmuir binding fit model of BIAevaluation software (version 4.1) was used for determining equilibrium dissociation constants (K_D) and kinetic dissociation (k_d) and association (k_a) constants using Equations (1) and (2), where R represents the response unit, and C is the concentration of the analyte.

$$dR/dt = K_a (R_{max} - R) - k_d R \quad (1)$$

$$K_D = k_d / K_a \quad (2)$$

5.1.8 *AlphaLISA assay optimization*

All AlphaLISA experiments described were performed at room temperature in white 384-well Opti-Plates (PerkinElmer, # 6007299) in a final volume of 25 μ L. For the detection of GST fusion protein, the anti-GST coated Acceptor beads (PerkinElmer, # AL110C) were used, while the detection of biotinylated peptide [Lys(Me₂)₂₀] - Histone H4 (8 - 30) (Anaspec, # AS-65419-1) was performed using the Streptavidin Donor Beads (PerkinElmer, # 6760002S). Equal concentrations of Acceptor beads and Donor beads were used in every experiments (5 μ g/mL final concentration of each bead). All incubation steps with AlphaScreen beads were performed at room temperature for 1 h under subdued lighting conditions, and finally, the assay plates were read in an EnSpire Multilabel plate reader (PerkinElmer).

The effect of UNC1215 analogues on Tudor activity was determined at a fixed 100 μ M concentration, incubating the compounds with 25 nM of the protein and 100 nM of the substrate in reaction buffer (Tris-HCl 20 mM, NaCl 25 mM, Tween20 0.05%, pH 8.0) for 2 h.

5.2 Spindlin1

5.2.1 *Protein expression and purification*

The vector, generously provided by M. Bedford, was transformed into BLR(DE3)pLysS competent cells and the protein was overexpressed in lysogeny broth (LB) in the presence of 100 μ g/mL ampicillin. Cells were grown at 37 °C to OD₆₀₀ of 0.8 and induced by isopropyl-1-thio-D-galactopyranoside (IPTG, 100 μ M) and incubated overnight at 16 °C. Cells

harbouring the expressed protein were pelleted at 5000g for 15 min at 4 °C and resuspended in 40 mL of Lysis Buffer (PBS buffer pH 7.4), adding protease inhibitors cocktail (Sigma). Cells were lysed using a sonicator (Vibra-Cells, Sonics) and cell debris were pelleted at 10000g for 30 min at 4 °C. The clarified lysate was filtered using a 0,45 µm syringe filter and loaded onto a 1-mL GStrap HP column (GE Healthcare, # 17-5281-01) using an AKTA FPLC (GE Healthcare) at the flow rate of 1 mL/min. GST-protein was eluted using the elution buffer (Tris HCl 50 mM, reduced glutathione 10 mM, pH 8.0) over 10 column volumes. Fractions containing the protein were confirmed by sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE) and then pooled.

5.2.2 *Spindlin1 labeling for MST measurements*

Fluorescence labeling of Spindlin1 was performed following the protocol for N-hydroxysuccinimide (NHS) coupling of the RED-NHS 2nd generation dye (MO-L011, NanoTemper Technologies) to lysine residues. Briefly, 100 µL of a 20 µM solution of MRG15 protein in labeling buffer (NaHCO₃ 130 mM, NaCl 50 mM, pH 8.2) was mixed with 100 µL of 60 µM 2nd generation dye in labeling buffer and incubated for 30 min at room temperature (r.t.). Unbound fluorophores were removed by size-exclusion chromatography with MST buffer as running buffer. The degree of labeling was determined using extinction coefficients $\epsilon = 82927.5 \text{ M}^{-1} \text{ cm}^{-1}$ for the protein and $\epsilon = 195000 \text{ M}^{-1} \text{ cm}^{-1}$ for the dye, with a correction factor C_{corr} of 0.04 at 280 nm, and a path length d of 1 cm,

$$C_{280} = [A_{280} - (A_{650} \times 0.04)] / \epsilon_{\text{protein}} \times d$$

$$\text{DOL} = A_{650} / 195000 \text{ M}^{-1} \text{ cm}^{-1} \times C_{280}$$

For storage, Spindlin1 was frozen in 10 µL aliquots at -80 °C. Prior to MST experiments, aliquots of Spindlin1 were thawed on ice and centrifuged for 15 min at 4 °C at 12000 g to remove protein aggregates.

5.2.3 Protein thermal stability measurements

To verify if the labeling of Spindlin1 with the dye would have altered the stability of the protein; thermal unfolding profiles of both proteins were recorded using Thyco NT.6 (NanoTemper technologies). This test was performed using the proteins at the concentration of 1.75 μM in MST buffer and the thermal unfolding was analyzed in a heating range from 35 to 95 $^{\circ}\text{C}$ with a thermal ramp of 30 $^{\circ}\text{C}/\text{min}$. The inflection temperature was calculated, together with a profile similarity percentage.

5.2.4 Assay development for MST screening

The MST experiments for Spindlin1 were performed on the Monolith NT.115Pico instrument using 5 nM protein concentration, premium coated capillaries, 20% LED power. All the data were analyzed at medium MST power.

For the binding curves with all the small molecules, compounds stock (20 mM) in DMSO were serially diluted in assay buffer starting from 400 μM , keeping constant the concentration of DMSO (2%). Each dilution was mixed 1:1 with a solution of 5 nM labeled proteins to yield a final volume of 20 μL . The reaction mixtures were incubated 10 minutes at room temperature and loaded into premium-coated capillaries.

K_{D} values were calculated from compound concentration-dependent changes in normalized fluorescence (F_{norm}) of proteins after 5 s of thermophoresis. Each binding experiment was performed in triplicate and the K_{D} values obtained are the mean of these three different experiments. Data were analyzed using MO Affinity Analysis software (NanoTemper Technologies).

5.3 MRG15

5.3.1 Protein expression and purification

MRG15 protein was expressed in BLR(DE3)pLysS cells in lysogeny broth (LB) in the presence of 100 µg/mL ampicillin. Cells were grown at 37 °C to OD600 of 0.8 and induced by isopropyl-1-thio-D-galactopyranoside (IPTG, 100 µM) and incubated overnight at 37 °C. Cells harbouring the expressed protein were pelleted at 5000 g for 15 min at 4 °C, washed twice with PBS, and frozen at -80°C. For protein purification, the pellet was thawed on ice, suspended in 40 mL of lysis buffer (PBS buffer pH 7.4) supplemented with protease inhibitors cocktail. Cells were lysed using a sonicator (Vibra-Cells, Sonics) (amplitude 30%) for 20 min on ice, and cell debris were pelleted at 10000 g for 30 min at 4 °C. The clarified lysate was filtered using a 0.45 µm syringe filter and loaded onto a 1 mL GStrap HP column (GE Healthcare, # 17528101) at the flow rate of 1 mL/min, previously conditioned with the binding buffer (PBS buffer pH 7.4). The protein was eluted in isocratic mode using the elution buffer (Tris-HCl 50 mM, reduced glutathione 10 mM, pH 8.0). Fractions containing the protein were confirmed by sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE), pooled, and dialyzed overnight at 4 °C to remove glutathione.

5.3.2 Protein labeling for MST measurements

Fluorescence labeling of MRG15 was performed following the protocol for N-hydroxysuccinimide (NHS) coupling of the RED-NHS 2nd generation dye (MO-L01, NanoTemper Technologies) to lysine residues. Briefly, 100 µL of a 20 µM solution of MRG15 protein in labeling buffer (NaHCO₃ 130 mM, NaCl 50mM, pH 8.2) was mixed with 100 µL of 60 µM 2nd generation dye in labeling buffer and incubated for 30 min at room temperature (r.t.). Unbound fluorophores were removed by size-exclusion chromatography with MST buffer as running buffer. The degree of labeling was determined using extinction coefficients $\epsilon = 65507$

M-1 cm-1 for the protein and $\epsilon = 195000 \text{ M}^{-1}\text{cm}^{-1}$ for the dye, with a correction factor C_{corr} of 0.04 at 280 nm, and a path length d of 1 cm, using the following equations:

$$C_{280} = [A_{280} - (A_{650} \times 0.04)] / \epsilon_{\text{protein}} \times d$$

$$\text{DOL} = A_{650} / 195000 \text{ M}^{-1}\text{cm}^{-1} \times C_{280}$$

For storage, MRG15 was frozen in 10 μL aliquots at $-80 \text{ }^\circ\text{C}$. Prior to MST experiments, aliquots of MRG15 were thawed on ice and centrifuged for 15 min at $4 \text{ }^\circ\text{C}$ at 12000 g to remove protein aggregates.

5.3.3 Protein thermal stability measurements

To verify if the labeling of MRG15 with the dye would have altered the stability of the protein, thermal unfolding profiles of both proteins were recorded using Thyco NT.6 (NanoTemper technologies). This test was performed using the proteins at the concentration of 3 μM in MST buffer and the thermal unfolding was analyzed in a heating range from 35 to 95 $^\circ\text{C}$ with a thermal ramp of 30 $^\circ\text{C}/\text{min}$. The inflection temperature was calculated, together with a profile similarity percentage.

5.3.4 Assay development for MST screening

The MST experiments for MRG15 were performed on the Monolith NT.115Pico instrument using 5 nM protein concentration, premium coated capillaries, 20% LED power. All the data were analyzed at medium MST power.

For the binding curves with all the small molecules, compounds stock (20 mM) in DMSO were serially diluted in assay buffer starting from 100 μM , keeping constant the concentration of DMSO (2%). Each dilution was mixed 1:1 with a solution of 5 nM labeled proteins to yield a final volume of 20 μL . The reaction mixtures were incubated 10 minutes at room temperature and loaded into premium-coated capillaries.

K_D values were calculated from compound concentration-dependent changes in normalized fluorescence (F_{norm}) of proteins after 5 s of thermophoresis. Each binding experiment was performed in triplicate and the K_D values obtained are the mean of these three different experiments. Data were analyzed using MO Affinity Analysis software (NanoTemper Technologies).

BIBLIOGRAPHY

1. Bird, A., Perceptions of epigenetics. *Nature* **2007**, *447* (7143), 396-398.
2. Biswas, S.; Rao, C. M., Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology* **2018**, *837*, 8-24.
3. Mahmood, N.; Rabbani, S. A., DNA Methylation Readers and Cancer: Mechanistic and Therapeutic Applications. *Frontiers in Oncology* **2019**, *9* (489).
4. Wagner, T.; Greschik, H.; Burgahn, T.; Schmidtkunz, K.; Schott, A.-K.; McMillan, J.; Baranauskienė, L.; Xiong, Y.; Fedorov, O.; Jin, J.; Oppermann, U.; Matulis, D.; Schüle, R.; Jung, M., Identification of a small-molecule ligand of the epigenetic reader protein Spindlin1 via a versatile screening platform. *Nucleic Acids Research* **2016**, *44* (9), e88-e88.
5. Pérez-Salvia, M.; Esteller, M., Bromodomain inhibitors and cancer therapy: From structures to applications. *Epigenetics* **2017**, *12* (5), 323-339.
6. Fujisawa, T.; Filippakopoulos, P., Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology* **2017**, *18* (4), 246-262.
7. Wagner, T.; Robaa, D.; Sippl, W.; Jung, M., Mind the Methyl: Methyllysine Binding Proteins in Epigenetic Regulation. *ChemMedChem* **2014**, *9* (3), 466-483.
8. Cochran, A. G.; Conery, A. R.; Sims, R. J., Bromodomains: a new target class for drug development. *Nature Reviews Drug Discovery* **2019**, *18* (8), 609-628.
9. Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J., How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology* **2007**, *14* (11), 1025-1040.
10. Musselman, C. A.; Lalonde, M.-E.; Côté, J.; Kutateladze, T. G., Perceiving the epigenetic landscape through histone readers. *Nature Structural & Molecular Biology* **2012**, *19* (12), 1218-1227.
11. Lu, R.; Wang, G. G., Tudor: a versatile family of histone methylation 'readers'. *Trends in Biochemical Sciences* **2013**, *38* (11), 546-555.
12. Chen, C.; Nott, T. J.; Jin, J.; Pawson, T., Deciphering arginine methylation: Tudor tells the tale. *Nature Reviews Molecular Cell Biology* **2011**, *12*, 629.
13. Cui, G.; Park, S.; Badeaux, A. I.; Kim, D.; Lee, J.; Thompson, J. R.; Yan, F.; Kaneko, S.; Yuan, Z.; Botuyan, M. V.; Bedford, M. T.; Cheng, J. Q.; Mer, G., PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nature Structural & Molecular Biology* **2012**, *19*, 916.
14. Taniwaki, M.; Daigo, Y.; Ishikawa, N.; Takano, A.; Tsunoda, T.; Yasui, W.; Inai, K.; Kohno, N.; Nakamura, Y., Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer. *International journal of oncology* **2006**, *29* (3), 567-75.
15. Pallasch, C. P.; Struss, A. K.; Munnia, A.; König, J.; Steudel, W. I.; Fischer, U.; Meese, E., Autoantibodies against GLEA2 and PHF3 in glioblastoma: tumor-associated autoantibodies correlated with prolonged survival. *International journal of cancer* **2005**, *117* (3), 456-9.
16. Adams-Cioaba, M. A.; Li, Z.; Tempel, W.; Guo, Y.; Bian, C.; Li, Y.; Lam, R.; Min, J., Crystal structures of the Tudor domains of human PHF20 reveal novel structural variations on the Royal Family of proteins. *FEBS Letters* **2012**, *586* (6), 859-865.
17. Zhang, T.; Ah Park, K.; Li, Y.; Sun Byun, H.; Jeon, J.; Lee, Y.; Hee Hong, J.; Man Kim, J.; Huang, S.-M.; Choi, S.-W.; Kim, S.-H.; Sohn, K.-C.; Ro, H.; Hoon Lee, J.; Lu, T.; Stark, G. R.; Shen, H.-M.; Liu, Z.-g.; Park, J.; Min Hur, G., PHF20 regulates NF- κ B signalling by disrupting recruitment of PP2A to p65. *Nature Communications* **2013**, *4* (1), 2062.
18. Tang, N.; Ma, L.; Lin, X.-Y.; Zhang, Y.; Yang, D.-L.; Wang, E.-H.; Qiu, X.-S., Expression of PHF20 protein contributes to good prognosis of NSCLC and is associated with Bax expression. *Int J Clin Exp Pathol* **2015**, *8* (10), 12198-12206.
19. Zhao, Q.; Qin, L.; Jiang, F.; Wu, B.; Yue, W.; Xu, F.; Rong, Z.; Yuan, H.; Xie, X.; Gao, Y.; Bai, C.; Bartlam, M.; Pei, X.; Rao, Z., Structure of human spindlin1. Tandem tudor-like domains for cell cycle regulation. *The Journal of biological chemistry* **2007**, *282* (1), 647-56.

20. Wang, J.-X.; Zeng, Q.; Chen, L.; Du, J.-C.; Yan, X.-L.; Yuan, H.-F.; Zhai, C.; Zhou, J.-N.; Jia, Y.-L.; Yue, W.; Pei, X.-T., SPINDLIN1 Promotes Cancer Cell Proliferation through Activation of WNT/TCF-4 Signaling. *Molecular Cancer Research* **2012**, *10* (3), 326-335.
21. Su, X.; Zhu, G.; Ding, X.; Lee, S. Y.; Dou, Y.; Zhu, B.; Wu, W.; Li, H., Molecular basis underlying histone H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. *Genes Dev* **2014**, *28* (6), 622-636.
22. Franz, H.; Greschik, H.; Willmann, D.; Ozretić, L.; Annette Jilg, C.; Wardelmann, E.; Jung, M.; Buettner, R.; Schüle, R., The histone code reader SPIN1 controls RET signaling in liposarcoma. *Oncotarget; Vol 6, No 7* **2015**.
23. Shanle, E. K.; Shinsky, S. A.; Bridgers, J. B.; Bae, N.; Sagum, C.; Krajewski, K.; Rothbart, S. B.; Bedford, M. T.; Strahl, B. D., Histone peptide microarray screen of chromo and Tudor domains defines new histone lysine methylation interactions. *Epigenetics & Chromatin* **2017**, *10* (1), 12.
24. Eissenberg, J. C., Structural biology of the chromodomain: Form and function. *Gene* **2012**, *496* (2), 69-78.
25. Xie, T.; Zmyslowski, Adam M.; Zhang, Y.; Radhakrishnan, I., Structural Basis for Multi-specificity of MRG Domains. *Structure* **2015**, *23* (6), 1049-1057.
26. PENA, A. N.; PEREIRA-SMITH, O. M., The Role of the MORF/MRG Family of Genes in Cell Growth, Differentiation, DNA Repair, and Thereby Aging. *Annals of the New York Academy of Sciences* **2007**, *1100* (1), 299-305.
27. Leung, J. K.; Berube, N.; Venable, S.; Ahmed, S.; Timchenko, N.; Pereira-Smith, O. M., MRG15 activates the B-myb promoter through formation of a nuclear complex with the retinoblastoma protein and the novel protein PAM14. *The Journal of biological chemistry* **2001**, *276* (42), 39171-8.
28. Zhang, P.; Du, J.; Sun, B.; Dong, X.; Xu, G.; Zhou, J.; Huang, Q.; Liu, Q.; Hao, Q.; Ding, J., Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Research* **2006**, *34* (22), 6621-6628.
29. Greschik, H.; Schüle, R.; Günther, T., Selective targeting of epigenetic reader domains. *Expert Opinion on Drug Discovery* **2017**, *12* (5), 449-463.
30. Zhu, H.; Wei, T.; Cai, Y.; Jin, J., Small Molecules Targeting the Specific Domains of Histone-Mark Readers in Cancer Therapy. *Molecules* **2020**, *25* (3), 578.
31. James, L. I.; Korboukh, V. K.; Krichevsky, L.; Baughman, B. M.; Herold, J. M.; Norris, J. L.; Jin, J.; Kireev, D. B.; Janzen, W. P.; Arrowsmith, C. H.; Frye, S. V., Small-Molecule Ligands of Methyl-Lysine Binding Proteins: Optimization of Selectivity for L3MBTL3. *Journal of Medicinal Chemistry* **2013**, *56* (18), 7358-7371.
32. James, L. I.; Baryte-Lovejoy, D.; Zhong, N.; Krichevsky, L.; Korboukh, V. K.; Herold, J. M.; MacNevin, C. J.; Norris, J. L.; Sagum, C. A.; Tempel, W.; Marcon, E.; Guo, H.; Gao, C.; Huang, X.-P.; Duan, S.; Emili, A.; Greenblatt, J. F.; Kireev, D. B.; Jin, J.; Janzen, W. P.; Brown, P. J.; Bedford, M. T.; Arrowsmith, C. H.; Frye, S. V., Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nature Chemical Biology* **2013**, *9* (3), 184-191.
33. Ganesan, A.; Arimondo, P. B.; Rots, M. G.; Jeronimo, C.; Berdasco, M., The timeline of epigenetic drug discovery: from reality to dreams. *Clinical Epigenetics* **2019**, *11* (1), 174.
34. Tomaselli, D.; Lucidi, A.; Rotili, D.; Mai, A., Epigenetic polypharmacology: A new frontier for epi-drug discovery. *Medicinal Research Reviews* **2020**, *40* (1), 190-244.
35. Mio, C.; Bulotta, S.; Russo, D.; Damante, G., Reading Cancer: Chromatin Readers as Druggable Targets for Cancer Treatment. *Cancers* **2019**, *11* (1).
36. Baell, J.; Walters, M. A., Chemistry: Chemical con artists foil drug discovery. *Nature* **2014**, *513* (7519), 481-483.
37. Magnusson, A. O.; Szekrenyi, A.; Joosten, H.-J.; Finnigan, J.; Charnock, S.; Fessner, W.-D., nanoDSF as screening tool for enzyme libraries and biotechnology development. *The FEBS Journal* **2019**, *286* (1), 184-204.
38. Nguyen, H. H.; Park, J.; Kang, S.; Kim, M., Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel, Switzerland)* **2015**, *15* (5), 10481-510.

39. Patching, S. G., Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2014**, *1838* (1, Part A), 43-55.
40. Rainard, J. M.; Pandarakalam, G. C.; McElroy, S. P., Using Microscale Thermophoresis to Characterize Hits from High-Throughput Screening: A European Lead Factory Perspective. *SLAS discovery : advancing life sciences R & D* **2018**, *23* (3), 225-241.
41. Asmari, M.; Ratih, R.; Alhazmi, H. A.; El Deeb, S., Thermophoresis for characterizing biomolecular interaction. *Methods* **2018**, *146*, 107-119.
42. Wilson, J.; Rossi, C. P.; Carboni, S.; Fremaux, C.; Perrin, D.; Soto, C.; Kosco-Vilbois, M.; Scheer, A., A Homogeneous 384-Well High-Throughput Binding Assay for a TNF Receptor Using Alphascreen Technology. *Journal of Biomolecular Screening* **2003**, *8* (5), 522-532.
43. Eglén, R. M.; Reisine, T.; Roby, P.; Rouleau, N.; Illy, C.; Bosse, R.; Bielefeld, M., The use of AlphaScreen technology in HTS: current status. *Curr Chem Genomics* **2008**, *1*, 2-10.
44. Bae, N.; Viviano, M.; Su, X.; Lv, J.; Cheng, D.; Sagum, C.; Castellano, S.; Bai, X.; Johnson, C.; Khalil, M. I.; Shen, J.; Chen, K.; Li, H.; Sbardella, G.; Bedford, M. T., Developing Spindlin1 small-molecule inhibitors by using protein microarrays. *Nature Chemical Biology* **2017**, *13* (7), 750-756.