

Universiteit
Leiden
The Netherlands

How electrostatic interactions drive nucleosome binding of RNF168 & PSIP : structural studies and their implications for rational drug design

Horn, V.

Citation
Horn, V. (2019, July 3). *How electrostatic interactions drive nucleosome binding of RNF168 & PSIP : structural studies and their implications for rational drug design*. Retrieved from https://hdl.handle.net/1887/74692

Version:	Not Applicable (or Unknown)
License:	Leiden University Non-exclusive license
Downloaded from:	https://hdl.handle.net/1887/74692

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle http://hdl.handle.net/1887/74692 holds various files of this Leiden University dissertation.

Author: Horn, V.
Title: How electrostatic interactions drive nucleosome binding of RNF168 & PSIP : structural studies and their implications for rational drug design
Issue Date: 2019-07-03

How Electrostatic Interactions Drive Nucleosome Binding of RNF168 & PSIP1

Structural Studies and Their Implications for Rational Drug Design

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op woensdag 3 juli 2019

klokke 11:15 uur

door

Velten Horn

geboren te Groß-Umstadt, Duitsland

in 1988

How Electrostatic Interactions Drive Nucleosome Binding of RNF168 & PSIP1
- Structural Studies and Their Implications for Rational Drug Design

Velten Horn

Doctoral thesis, Leiden University, 2019
Cover design: Klaus Schüller
ISBN number: 978-94-91546-16-7
© 2019, Velten Horn
Printed by boekdrukken.com

Promotor:	prof. dr. M. Ubbink
Co-Promotor:	dr. H. van Ingen
Promotiecommissie:	prof. dr. H.S. Overkleeft (voorzitter)
	prof. dr. N.I. Martin (secretaries)
	prof. dr. T. Sixma (Nederlands Kanker Instituut)
	dr. K.M. Bonger (Radboud Universiteit Nijmegen)
	prof. dr. R. Boelens (Universiteit van Utrecht)

The research described in this thesis was performed in the Macromolecular Biochemistry Department of the Leiden Institute of Chemistry in Leiden, the Netherlands and the NMR Spectroscopy Research Group of the Bijvoet Center for Biomolecular Research in Utrecht, the Netherlands.

Table of Contents

LIST OF ABBREVIATIONS **2**

Chapter 1

INTRODUCTION **5**

Chapter 2

STRUCTURAL BASIS OF SPECIFIC H2A K13/K15 UBIQUITINATION BY RNF168 ..... **41**

Chapter 3

MIMICKING THE NUCLEOSOMAL CONTEXT IN A PEPTIDE: INCREASING THE BINDING AFFINITY TOWARDS THE EPIGENETIC READER PSIP1 ..... **79**

Chapter 4

A NOVEL SCREENING APPROACH FOR PEPTIDES TARGETING THE TRIMETHYLLYSINE RECOGNITION DOMAIN OF PSIP1 ..... **111**

Chapter 5

GENERAL CONCLUSIONS & PERSPECTIVES **137**

ENGLISH SUMMARY **147**

NEDERLANDSE SAMENVATTING **150**

CURRICULUM VITAE **153**

LIST OF PUBLICATIONS **154**

List of Abbreviations

6-FAM	5,6-carboxyfluoresceine
AIR	ambiguous interaction restraints
ATP	adenosine triphosphate
Bp	basepairs
BS	binding sequence
CPMG	Carr-Purcell-Meiboom-Gill sequence
Cryo-EM	cryogenic electron microscopy
CSP	chemical shift perturbation
DCM	dichloromethane
DIC	N,N'-Diisopropylcarbodiimide
DSB	double-strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EPR	electron paramagnetic resonance
FMOC	fluorenylmethyloxycarbonyl
FRET	Förster resonance energy transfer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOAt	1-Hydroxy-7-azabenzotriazole
HSQC	heteronuclear single quantum coherence
IEX	ion exchange
IM-MS	ion-mobility spectrometry–mass spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
KAc	acetylated lysine
Kme	methylated lysine
LC-MS	liquid chromatography–mass spectrometry
MD	Molecular dynamics
Mmt	monomethoxytrityl

MST	microscale thermophoresis
NCP	nucleosome core particle
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PCS	pseudocontact chemical shift
PPI	protein-protein interactions
PRE	paramagnetic relaxation enhancement
PMSF	phenylmethylsulfonyl fluoride
PTM	post translational modification
SANS	small-angle neutron scattering
SASD	surface accessible distance
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulfate
Sph	phosphorylated serine
ssNMR	solid-state NMR
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TIPS	triisopropylsilane
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimized spectroscopy
TSP	trimethylsilylpropanoic acid
Ub	ubiquitin
UPLC	ultra performance liquid chromatography
XL-MS	cross-link-based mass spectrometry

1

Introduction

Based on the review article: **V. Horn** & H. van Ingen (2019) *Recognition of nucleosomes by chromatin factors: Lessons from data-driven docking-based structures of nucleosome-protein complexes*

Accepted for publication (DOI: 10.5772/intechopen.81016)

ABSTRACT

The function of chromatin ultimately depends on the many chromatin-associated proteins and protein complexes that regulate all DNA-templated processes such as transcription, repair and replication. As the molecular docking platform for these proteins, the nucleosome is the essential gatekeeper to the genome. As such, the nucleosome-binding activity of a myriad of proteins is essential for a healthy cell.

Here, we review the molecular basis of nucleosome-protein interactions and classify the different binding modes available. The structural data needed for such studies not only comes from traditional sources such as X-ray crystallography, but also increasingly from other sources. In particular, we highlight how sparse interaction data, derived from for example NMR or mutagenesis, is used in data-driven docking to drive the modeling of the complex into an atomistic structure. This approach has opened up detailed insights for several nucleosome-protein complexes that were intractable or recalcitrant to traditional methods. Throughout the thesis presented here, we follow this integrative approach to face yet unmet challenges in the structure elucidation of the nucleosome-binding ubiquitin ligase RNF168. Furthermore, we use data-driven docking to rational design modulators of the epigenetic reader protein PSIP1. These structures guide the formation of new hypotheses and advance our understanding of chromatin function at the molecular level.

Introduction

The packaging of DNA into chromatin represents one of the most fundamental layers of the biology of the cell. It provides the required structural compaction of DNA to fit in the nucleus and plays crucial roles in controlling cell fate and protecting genome integrity. The fundamental unit of chromatin is the nucleosome in which 147 base pairs (bp) of DNA are wrapped around an octameric protein complex composed of two copies of histone proteins H2A, H2B, H3 and H4 [1-3]. Herein, H2A-H2B and H3-H4 form heterodimeric complexes due to hydrophobic interactions. H3-H4 dimers then further associate to assemble a tetramer that, together with two H2A-H2B dimers completes the octameric histone complex. Each histone protein possesses three contact points with the nucleosomal DNA, where arginine side chains are inserted into a minor groove roughly each helical turn, demanding the involved minor grooves to adapt a bent conformation. Nucleosomes are arranged as beads-on-a-string forming 10 nanometer (nm) wide fiber that subsequently condenses into higher order structures [4] (Figure 1-1). Chromatin state and DNA accessibility are determined at the nucleosome level. Changes therein are mediated through interactions of histone proteins and nucleosomal DNA with a wide range of protein complexes that control the structure of chromatin. They interpret, write and erase post translational modifications, act as ATP-dependent nucleosome remodelers or recruit subsequent chromatin effector proteins. This allows changes in the functional state of chromatin and regulation of DNA templated processes. While promoting a large variety of effects on chromatin structure, nucleosome interacting proteins share the molecular basis of recognizing and binding the nucleosome. Understanding the basis of chromatin dynamics therefore demands understanding the molecular basis of nucleosome-protein interactions.

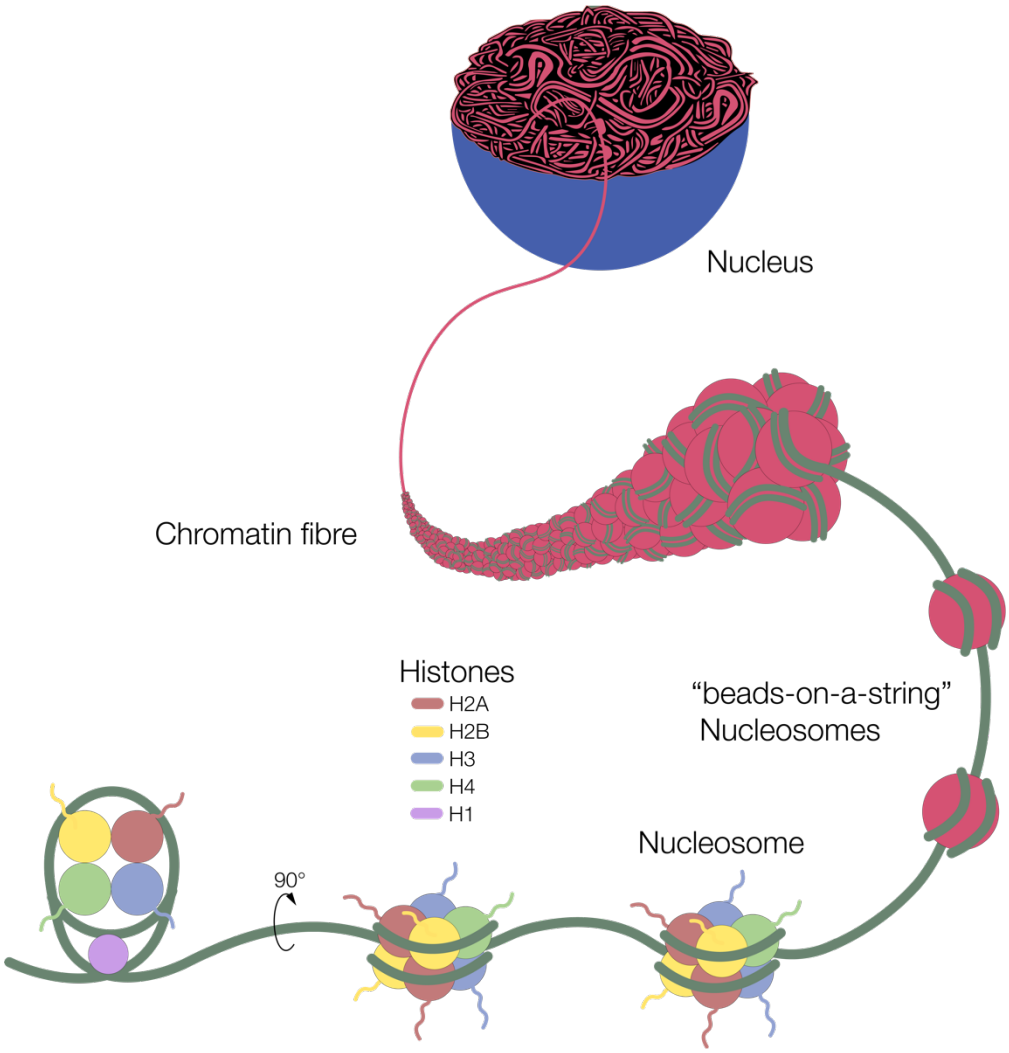


Figure 1-1. Schematic depiction of chromatin organization in eukaryotic nuclei.

147 bp of DNA are wrapped around an octameric complex of histone proteins with two copies of a H2A/H2B dimer and one H3/H4 tetramer, forming a nucleosome. Additionally, histone H1 is positioned on the nucleosome dyad. Nucleosomes are displayed in a "beads-on-a-string" conformation, that further condensates into higher order chromatin fibers in the nucleus.

Insights into the molecular mechanistic basis of how histone-modifying enzymes install or remove post-translational modifications (writers and erasers, respectively) and how these modifications are recognized by effector proteins (readers) are of immense

interest, especially in drug development. Deregulation of these proteins is connected to pathological outcome, including cardiovascular diseases, neurological disorders, metabolic disorders and cancer [5]. So-called epigenetic drugs that target the nucleosome interaction of these chromatin factors offer new therapeutic potential [6-9]. A selection of epigenetic drugs including those currently undergoing clinical trial are described in detail elsewhere [10]. Advancement in their development requires insights into the underlying molecular mechanism of nucleosome recognition, enabling control over subsequent modification of the chromatin state.

In this chapter, we will review the molecular basis of nucleosome-protein interactions, focusing on the different binding epitopes presented by the nucleosome. After an overview of the nucleosome-protein structures determined by crystallography or cryo-electron microscopy (cryo-EM), we highlight several studies in which experimental data from nuclear magnetic resonance spectroscopy (NMR), cross-link-based mass spectrometry (XL-MS) or mutational analysis were used to build atomistic structural models of nucleosome complexes. Throughout, we emphasize the role of these data-driven models in deepening our understanding of nucleosome recognition.

NUCLEOSOME BINDING EPITOPES

Consisting of DNA and histone proteins, the nucleosome offers a selection of distinct interaction surfaces. These serve as platforms for the binding of effector proteins with high levels of specificity (Figure 1-2). Most interaction sites are situated on the histone proteins, that possess a globular tertiary structure with exposed, disordered N-terminal tails. Histone tails are known to carry a wide range of covalent, post-translational side chain modifications (PTMs), serving as specific recognition sites for effector proteins. PTMs include lysine and arginine mono-, di- and trimethylation, lysine acetylation, serine and threonine phosphorylation as well as lysine ubiquitination [11, 12]. This cosmos of modifications maintains a dynamic nature through the reversibility of the covalent modifications. Modified histones are recognized by so-called reader protein domains specific for the respective modification (Figure 1-2A). Nucleosome interacting proteins can possess more than one reader domain which allows cross-talk between different post-translational modifications. Examples of PTM reader domains are Chromo, Tudor, PHD and MBT domains for methylated lysine residues (Kme1,Kme2, Kme3), bromodomains for acetylated lysine (KAc) residues and 14-3-3 proteins for phosphorylated serine (Sph) [11, 13] (Table 1-1). The most recent addition to the list is YEATS domains that recognize crotonylated lysine [14-16]. Reader domains are

defined conserved motifs that are able to complex a specific modification. The "Royal Family" of reader domains are in this respect a particularly instructive example. This superfamily includes the Chromo, MBT, PWWP and plant agenet domains that bind methylated lysine (Tudor, Chromo, MBT, PWWP, plant agenet) or arginine (Tudor) residues. Most domains of this family contain a barrel-shaped structure formed by 3 to 5 antiparallel β-strands that holds a cluster of aromatic residues that form the so-called aromatic cage [17]. The aromatic cage presents an electron-rich yet hydrophobic surface that is well suited to bind methylated lysines through cation-π interactions [18]. We will address the molecular basis of the aromatic cage containing PWWP domain of PSIP1 in this thesis as well (see Chapter 3 and Chapter 4).

The detailed description of structural features and similarities of epigenetic reader domains, as well as their substrate specificity have been subject to literature reviews [19-21].

Table 1-1. Overview of selected reader domains for post translational modifications[a].

(sub)Domain	Modification	Protein	Function
Royal Family			
Tudor	Kme1, Kme2, Kme3, Rme2	53BP1	DNA damage response [22]
		TDRD3	Transcription activation [23]
MBT	Kme1, Kme2	L3MBTL1	Transcriptional repression [24 25]
PWWP	Kme3	PSIP1	Transcriptional co-activation, DNA repair [26 27]
Chromo	Kme, Kme2, Kme3	CHD1	Chromatin remodeling [28 29]
		HP1	Heterochromatin [30]
		MRG15	Splicing [31]
Plant agenet	Kme, Kme2, Kme3	FMRP	DNA damage response [32]
Bromodomain			
	KAc	BRD2/3	Transcriptional regulation [33]
14-3-3			
	Sph	14-3-3γ	Transcriptional activation [34]

[a] See ref. [21, 35, 36] for more in-depth discussion.

Reader domains can, in addition to the post translational modification, show specificity for a defined amino acid sequence motif around the epigenetic mark that supports complex formation. For example, the WD40 domain of the EED (embryonic ectoderm development) protein selectively reads out trimethylated lysine in a A-R-K-S sequence motif (as for H3K27me3) but not in a R-T-K-Q motif (as for H3K4me3) [37].

Next to histone tails, the nucleosome also possesses intrinsic docking platforms on its histone surface. The most prominent of these is composed of histones H2A and H2B. While the histone octamer is overall highly positively charged, there is a patch on the H2A-H2B dimer surface formed by acidic residues with negative surface charge. This structural feature is named the acidic patch and engages in a manifold of interactions with specific binding domains (Figure 1-2B), including the tail of histone H4 of adjacent nucleosomes that promotes chromatin compaction. A common feature observed for acidic patch interacting proteins is a positively charged arginine residue that interacts with a triad of acidic residues on H2A (Glu60, Asp89, Glu91). This is referred to as the arginine anchor [38]. It is often supported by surrounding positively charged residues interacting with acidic H2A/H2B interface residues.

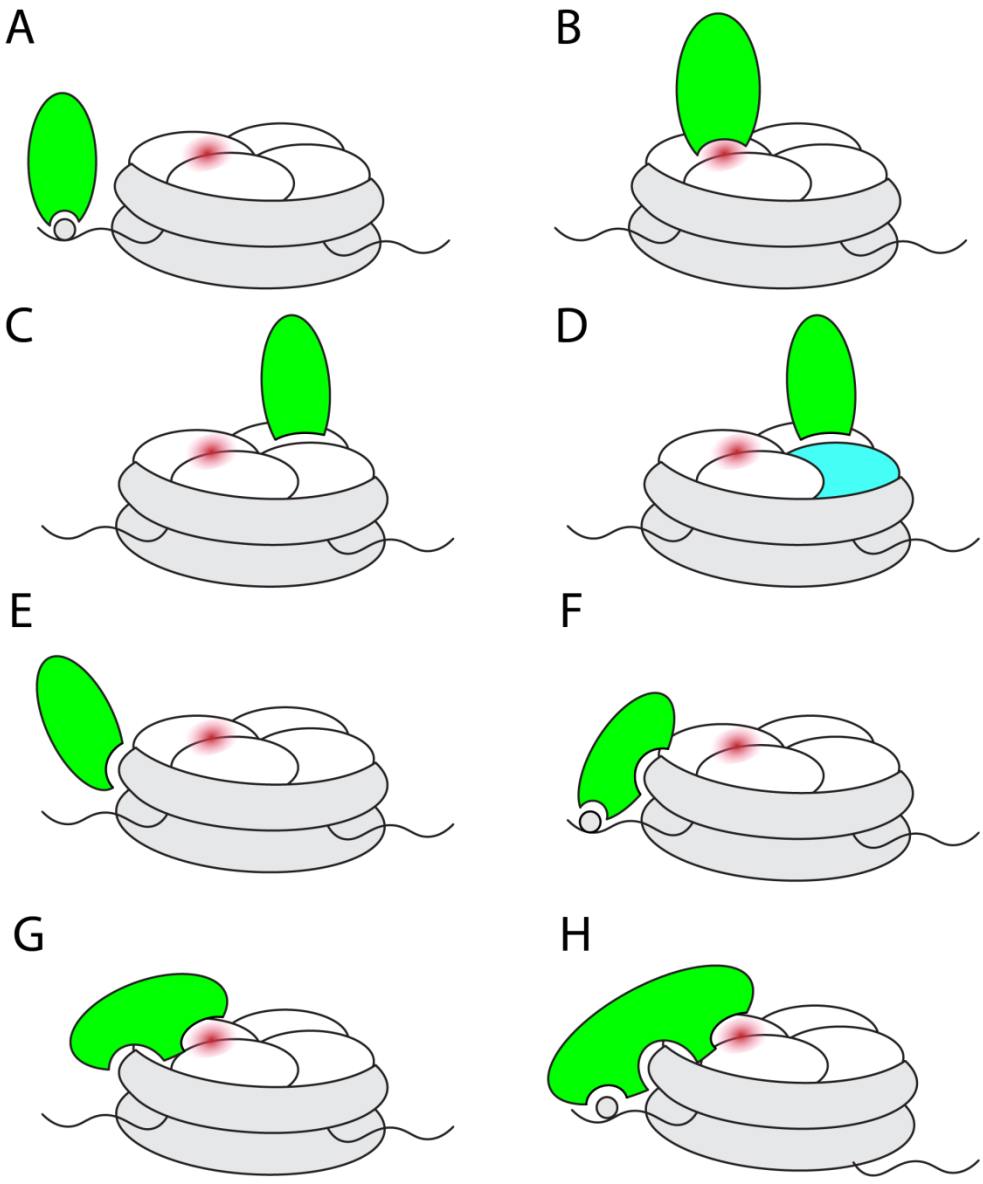


Figure 1-2. A schematic depiction of different modes of nucleosome recognition.

Reported types of epitopes are histone tails including PTMs (A), the H2A-H2B acidic patch (B), the canonical histone surface (C), specific surface motifs formed by histone variants (D) or nucleosomal DNA (E). A manifold of synergetic combinations of binding epitopes are known, such as histone mark and DNA (F), acidic patch and DNA (G) or all three epitopes (H).

Other parts of the histone core surface may also mediate protein-nucleosome interactions (Figure 1-2C). First, a solvent exposed cleft between H4 and H2B was shown to be involved in binding interactions with Sir3 or 53BP1 [39, 40]. These proteins bind simultaneously to both the H4-H2B cleft and the acidic patch using one nucleosome binding domain for each epitope. Second, incorporation of non-canonical histones in nucleosomes introduces specific interaction surfaces that allow histone variant-specific nucleosome binding (Figure 1-2D). An example hereof are CENP-N and CENP-C that recognize the incorporated histone H3 variant CENP-A [41, 42].

Finally, the nucleosomal DNA is a major protein interaction site. First, it forms the binding site of linker histone H1 [43-45]. Second, it is often involved in additional synergistic interactions to nucleosome binding domains (Figure 1-2E). Recent studies have furthermore identified transcription factor proteins that primarily bind to nucleosome DNA. These so-called pioneer factors bind their DNA target sites while embedded in the nucleosome [46-48]. The structural details of these are however still lacking.

Throughout the advances in studies on nucleosome binding, it has become clear that binding of effector proteins in many cases involves interactions of nucleosome-binding domains to multiple nucleosome epitopes (Figure 1-2G-H). However, due to their size and complexity as well as the stability and dynamics of complex formation, the nucleosome is a challenging system for structural biology.

CRYSTAL CLEAR – LESSONS FROM CRYSTALLOGRAPHY AND SINGLE PARTICLES

A key role in the research of protein interactions are high-resolution three-dimensional structures of the complexes, typically obtained by crystallography and, increasingly, cryo-electron microscopy. These structures enable the identification of binding sites and intermolecular interactions, offering a guided approach to design binding deficient mutants or competitive binders. The history of nucleosome structural biology peaked with the publication of the high-resolution crystal structure of the nucleosome in 1997 [3]. Luger *et al.* achieved crystallization of the nucleosome together with a palindromic version of human α-satellite DNA. This milestone study provided the foundation to also study the structures of nucleosomes together with chromatin factors in complexes. Table 2 lists the structures of nucleosome-protein complexes solved to date by crystallography and cryo-electron microscopy [39, 49-57]. The most recent

addition to this ever-growing list are the spectacular structures of the complex between the INO80 chromatin remodeler and the nucleosome [58, 59]. Below, we discuss a few cardinal studies to highlight the different nucleosomal binding modes of effector proteins.

THE FIRST CRYSTAL STRUCTURE OF A NUCLEOSOME COMPLEX

The first high-resolution structure of a nucleosome-protein complex was the crystal structure of a peptide model of Kaposi's Sarcoma-Associated Herpesvirus LANA N-terminal region bound to the nucleosome [60]. The binding site identified in this study was the acidic patch. The atomistic resolution allowed to identify intermolecular side chain interactions including the arginine anchor, bound to the acidic triad, an arginine binding motif composed of Glu60, Asp89 and Glu91 on the acidic patch. Ever since, the LANA-nucleosome has become a golden standard for comparisons with other acidic patch interactions [49, 54]. Importantly, LANA is used to investigate the acidic patch binding ability of other proteins by competitive binding [61-63]. This exact epitope happened to be the binding interface also for the first full protein domain that was crystalized in its nucleosome-bound state.

THE FIRST CRYSTAL STRUCTURE OF A NUCLEOSOME-BOUND PROTEIN DOMAIN

The first structure of a protein bound to the nucleosome was the RCC1-nucleosome complex published by the Tan lab in 2010. RCC1 (regulator of chromosome condensation) is essential during mitosis by recruiting Ran GTPase, which plays a role in nucleus reorganization, to the nucleosome [64, 65]. A comparison with LANA highlighted the crucial and conserved interaction of arginine residues with the acidic patch triad [49]. RCC1 binds to the acidic patch using the canonical arginine anchor, here contained in a loop, and also binds the nucleosomal DNA through its N-terminal tail. Such synergetic interactions have been observed later in many other nucleosome binding proteins [49, 54, 66-69]. This study was the first to show such complexity of nucleosomes as interaction platforms due to RCC1 interacting with both, acidic patch and nucleosomal DNA. It also highlights the importance of properly defining the boundaries of binding domains to capture all binding epitopes in order to reveal possible synergetic interactions and fully understand complex formation and subsequent effects on chromatin structure.

Specificity of Effector Protein Orientation in Nucleosome Complex Formation

Besides determining the binding mode, synergetic interactions can also provide the structural basis for specificity of effector protein activity. This was shown in the crystal structure, also from the Tan lab, of the polycomb repressive complex 1 (PRC1) that ubiquitinates H2A K119 in a highly specific manner [54]. On its surface the nucleosome displays various lysine residues that can be ubiquitinated by the respective writer proteins. However, the downstream response wildly differs depending on the position of the ubiquitinated lysine. Thus, target specificity is of high importance for ubiquitin writer proteins. In case of PRC1, this is based on two distinct binding processes. For one there is the interaction between acidic patch and the arginine anchor of the Ring1B/Bmi1 subunit. In addition, the E2 subunit UbcH5c engages the nucleosomal DNA. Both contributions combined are responsible for exact positioning of the catalytic center of the ubiquitin carrying E2 to the target H2A K119 (Figure 1-3E).

Besides LANA, RCC1 and PRC1, other crystal structures of nucleosome complexes offered further insights into nucleosome recognition. In particular, the structure of the nucleosome complex of the SAGA DUB deubiquitination module showed a non-canonical acidic patch binding. Morgan *et al.* found that the SAGA nucleosome-binding DUB module possesses three equally crucial arginine residues distributed over an α-helix [55] (Figure 1-3D). This perhaps points towards yet other acidic patch interaction modes.

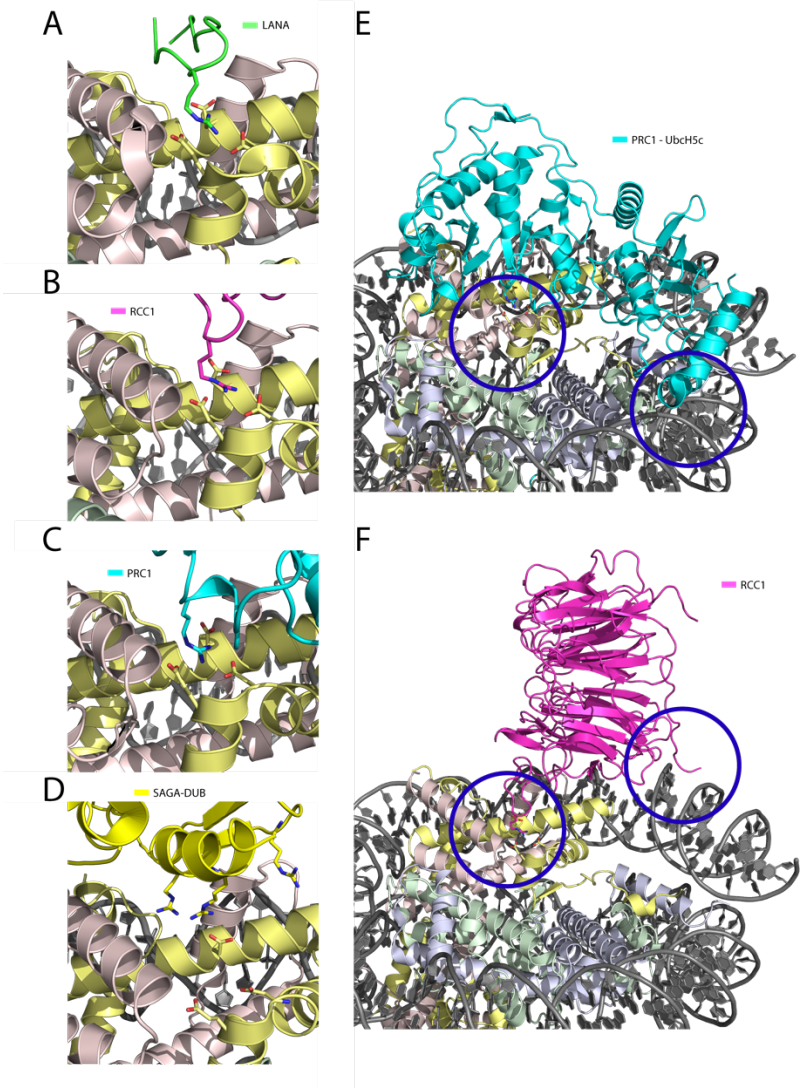


Figure 1-3. A detailed depiction of the acidic patch region and the triad of H2A that complexes the canonical arginine anchor.

(A) LANA was the first peptide model of a binding domain to be crystalized (green, pdb: 1zla). (B) The first crystalized protein domain was RCC1 (pink, pdb: 3mvd). (C) They all share the acidic patch as a common binding epitope for the arginine anchor residue, as is also the case for the PRC1 complex with its acidic patch binding RING domain (cyan, 4r8p). (D) Acidic patch binding is not necessarily limited to one single arginine anchor residue. As for the nucleosome-bound structure of the deubiquitinase complex SAGA-DUB (yellow, pdb: 4zux), three arginine residues are essential part of the acidic patch binding, of which none occupies the position in the center of the acidic triad. Combination of nucleosomal DNA and acidic patch binding shown in structure of RCC1 and PRC1. (E) Shown are PRC1 (cyan, pdb:

4r8p) that besides the acidic patch, also engages DNA with its UbcH5c subunit. (F) The same holds true for RCC1 (pink, pdb: 3mvd) that contacts to DNA with the unstructured tail region

Recently, also cryo-EM derived structures of nucleosome-protein complexes have been published. The first, solved in 2016, yielded the structure of the complex with 53BP1, a reader protein for post-translational histone modifications [40]. As for RCC1, also here a multitude of synergetic interactions were found. Subsequently, the structures of Snf2 and CENP-N were solved and published [70-72].

Since the first crystal structure two decades ago, the list of nucleosome complexes deposited in the RCSB PDB protein databank is continuously growing. Still, the fourteen high-resolution structures solved to date only encompass a fraction of all nucleosome-protein interactions. This discrepancy highlights the need for complementary techniques in chromatin structural biology.

Table 1-2. List of structures for nucleosome-bound chromatin effector proteins or domains deposited in the RCSB protein databank PDB.

Proteins						
Name	**PDB-ID**	**Role**	**Year**	**Technique**	**Reference**	**Resolution [Å]**
RCC1	3MVD	Ran recruitment	2010	X-Ray	[49]	2.9
Sir3 BAH	4JJN, 3TU4, 4LD9, 4KUD	Chromatin compaction	2011, 2011, 2013, 2013	X-Ray	[39, 50-52]	3.0/3.0/3.3/3.2
CENP-C	4X23	H3 variant binding	2013	X-Ray	[53]	3.5
Ring1B	4R8P	E3 ligase	2014	X-Ray	[54]	3.2
53BP1	5KGF	Reader	2016	EM	[40]	4.5
SAGA/DUB	4ZUX	Eraser	2016	X-Ray	[55]	3.8
Set8	5HQ2	Writer	2016	X-Ray	[56]	4.5
Chd1	5O9G, 6G0L, 6FTX	Remodeler	2017, 2018	X-Ray, EM	[73, 74]	4.8/10.0/4.5
Snf2	5X0X/Y	Remodeler	2017	EM	[70]	3.9

CENP-N	6BUZ, 6C0W	H3 variant binder	2017, 2018	EM	[71, 72]	3.9/4.0
H1	4QLC, 5NL0	Linker histone	2015, 2017	X-Ray	[45, 75]	3.5
INO80	6FML, 6ETX	Remodeling-complex	2018, 2018	EM	[58, 59]	4.3/4.8
RNA polymerase II	6A5L/O/P/R/T/U/V	DNA transcription	2018	EM	[76]	5.6/9.9/7.0/8.7/6.7/7.6/6.9/
SWR1	6GEJ/N	Remodeler	2018	EM	[77]	3.6/3.6
Dot1L	6NJ9, 6NOG, 6J99, 6NN6, 6NQA	methyltransferase	2019	EM	[78]	2.9/3.9/4.1/3.9/3.5

Peptides

Name	PDB-ID	Role	Year	Technique	Reference	Resolution [Å]
LANA	1ZLA, 5GTC	Viral protein	2006, 2017	X-Ray	[60, 79]	2.9/2.7
IE1	5E5A	Viral protein	2016	X-Ray	[80]	2.8
GAG	5MLU	Synthetic acetylation system	2017	X-Ray	[81]	2.8

DATA-DRIVEN MODELING

Next to traditional structure determination methods is the modeling of structures of complexes based on some sort of experimental information on the interaction [82, 83]. In such data-driven modeling of a complex structure, the two interaction partners are docked together, guided by the experimental data, and respecting their biophysical properties. The exact binding interface and relative orientation of the binding partners is typically refined over several steps. Prerequisite for this approach is the availability of the 3D structures of the interacting partners. Several molecular docking programs allow the incorporation and use of experimental data and so increase the accuracy of resulting structures [84]. Hence, data from diverse

biophysical techniques is translated into restraints guiding the docking process [85-87]. The type of information includes: interaction interface, distances or shape of the complex and its subunits. Techniques that can provide this information are listed in table 1-3.

Table 1-3. Biochemical and biophysical techniques that can provide information on the architecture of complexes and their subunits.

Interaction interface	Distances	Shape
Mutagenesis	XL-MS	Cryo-EM
H/D exchange	Forster resonance energy transfer (FRET)	Small angle X-ray or neutron scattering (SAXS/SANS)
NMR spectroscopy (CSP)	Electron paramagnetic resonance (EPR)	Ion-mobility mass spectrometry (IM-MS)
	NMR spectroscopy (PRE, PCS, NOE)	

Especially NMR spectroscopy can provide vital information to be implemented as restraints. It is possible to gather data on intermolecular distances by paramagnetic relaxation enhancement (PRE), pseudocontact shifts (PCS) and the nuclear Overhauser effect (NOE). Furthermore, information on binding interfaces and binding affinity are accessible through chemical shift perturbation (CSP) experiments. Here, an increasing amount of one binding partner is titrated to an isotopically labelled sample of the second binding partner. The changes in chemical environment upon complex formation causes chemical shift changes for residues in the binding interface. While CSPs can also be induced by structural changes, unrelated to direct binding, the effect proved to be very powerful to map binding interfaces with a residue resolution. Especially, the sensitivity to even low affinity binding is of particular value for interaction studies. The use of these NMR methods in docking studies is reviewed in detail elsewhere [82]. An overview of publications that used data-driven docking to investigate nucleosome-protein complexes is listed in table 1-4.

The work presented in this thesis is heavily based on data derived from NMR spectroscopy and XL-MS. Therefore, we will introduce several studies in more detail that employed particularly these two techniques.

Table 1-4. List of structural models of nucleosome-protein complexes, based on biophysical data in combination with docking studies, that were featured in peer-reviewed literature.

Protein	Role	Data source	Reference
PSIP1 - PWWP	Trimethyllysine reader H3K36	NMR	*66, 67, 88*
CoREST/LSD1	Demethylase	Crystallography/NMR	*89*
Rad6-Bre1	Ubiquitin ligase	XL-MS	*69*
LANA	Viral protein	ssNMR	*90*
NSD1	Methyltransferase H3K36	Mutagenesis	*91*
RNF169	Ubiquitin reader	NMR, SAXS	*68, 92*
H1	Linker histone	NMR	*43, 93*
ISW2	Chromatin remodeler	XL-MS	*94*
Rad18	DNA repair factor	NMR	*92*
RCC1	Ran-recruitment	Crystallography	*61*
PHF1 Tudor	Trimethyllysine reader H3K36	Crystallography/NMR	*95*
HMGN2	Chromatin decompaction	NMR	*96*

NMR SPECTROSCOPY IN DATA-DRIVEN MODELING

Over recent years, several studies have demonstrated that state-of-the-art solution NMR can offer high-resolution and site-specific characterization of the structures and dynamics of nucleosome-protein complexes. NMR has the particular advantage of its sensitivity to dynamics and the ease with which interactions can be studied, allowing detailed insights into molecular recognition processes. NMR allows studies when systems are dynamic, or (partially) disordered, while this typically hampers high-resolution structure determination by crystallography and cryo-EM. In the following, we will present several studies that involved the use of NMR data to obtain structural models of nucleosome complexes.

BRINGING DATA-DRIVEN MODELING TO NUCLEOSOME COMPLEXES

A pioneer study for data-driven modeling of a nucleosome complex was successfully applied for the lysine-specific demethylase 1 and CoREST complex (LSD1-CoREST) [89]. Both proteins cooperate in the demethylation of mono- and dimethylated H3K4. While it was possible to solve the crystal structure of LSD1-CoREST, their nucleosome-bound state remains elusive. Yang *et al.* gained insight into the molecular basis of LSD1-CoREST interaction by identifying point mutations that interfere with the LSD1-CoREST ability to demethylate a methylated peptide model of the histone H3 tail. Since it was previously shown that that LSD1 recognizes a specific stretch of the H3 tail [97], it was possible to employ modelling to identify intermolecular interactions between the peptide and both the LSD1 active site as well as the LSD1-CoREST interface (Figure 1-4B). Lastly, NMR titration experiments of the CoREST SANT2 domain with DNA revealed a DNA binding interface on SANT2. These pieces of interaction data were used to guide a docking approach resulting in a complete structural model of the LSD1-CoREST-nucleosome complex (Figure 1-4A). This is a prime example of combining crystal structures, mutagenesis and NMR data to overcome limitations of the separate techniques.

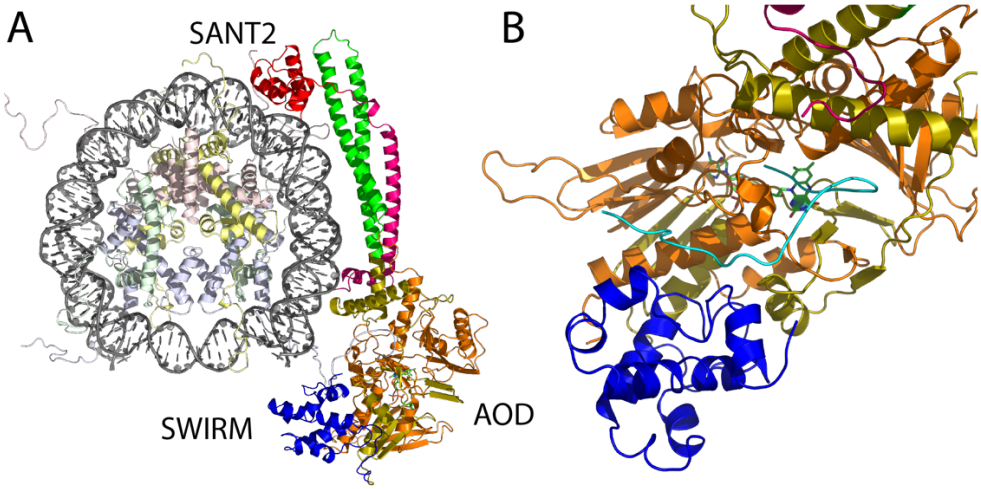


Figure 1-4. Structural model of nucleosome-bound LSD1-CoREST.

(A) Structural model of LSD1-CoREST bound to the nucleosome. The DNA binding of the SANT2 domain was elucidated by NMR spectroscopy. A previously identified binding motif in the H3 tail sequence was docked onto the interface of amine oxidase (AOD) and SWIRM domain revealing a second binding epitope. (B) The resulting model of the model peptide binding to AOD-SWIRM is shown as a close-up, highlighting how the tail is positioned on the interface of both domains. Figure generated using the author-provided PDB file [89].

NMR-BASED DATA-DRIVEN DOCKING HIGHLIGHTS THE IMPORTANCE OF THE NUCLEOSOMAL CONTEXT IN EPIGENETIC READ-OUT

The complexity of nucleosome recognition by reader proteins is well illustrated by the NMR-based studies on the recognition of H3K36me-nucleosomes by the PWWP domain of PSIP1(Ledgf). NMR studies of this reader interaction found that the PWWP domain has a binding affinity orders of magnitude lower for a H3K36me peptide compared to H3K36me3 in a nucleosomal context. Interestingly, a similar observation was made for the Tudor domain of the H3K36me reader PHF1 [88]. Here an isolated peptide model of the H3 tail showed decreased affinity as well. Due to the proximity of H3K36 to nucleosomal DNA a role of DNA binding was hypothesized for both proteins. NMR studies showed for PSIP1 and PHF1 alike a binding site for nucleosomal DNA for both reader domains, resulting in a simultaneous binding mechanism of both trimethyllysine and nucleosomal DNA.

For PHF1-Tudor, a crystal structure bound to a trimethylated H3 tail peptide was already available. The additional importance of the nucleosomal context and synergetic binding mechanism can be understood from the corresponding nucleosome-bound structural model (Figure 1-5A). In case of PSIP1-PWWP, the domain structure was solved by NMR and, together with NMR titration data, used to determine a structural model of nucleosome-bound protein (Figure 1-5B) [66, 67, 88]. The structural models of both highlighted the importance of the nucleosomal context in H3K36me3 recognition, emphasizing that complex formation critically depends on two synergetic binding processes. Firstly, the aromatic residues that form the aromatic cage bind to trimethylated lysine H3K36me3. This recognition of the PTM is crucial for the binding but only in combination with the interaction between nucleosomal DNA and positive surface residues do the readers reach their full binding affinity. This makes both studies outstanding examples of synergetic interplay of epitopes in nucleosome binding proteins (Figure 1-5C+D).

The insights on synergetic nucleosome binding derived from these structural models were used for the design of experiments to validate the structural model and may offer possible tools for further research. In case of PSIP1-PWWP, the structural model sparked efforts in the design of nucleosome-mimicking peptides to modulate the PSIP1-chromatin interaction. The design and characterization of such peptide-based substrates is subject of chapter 3 in this thesis.

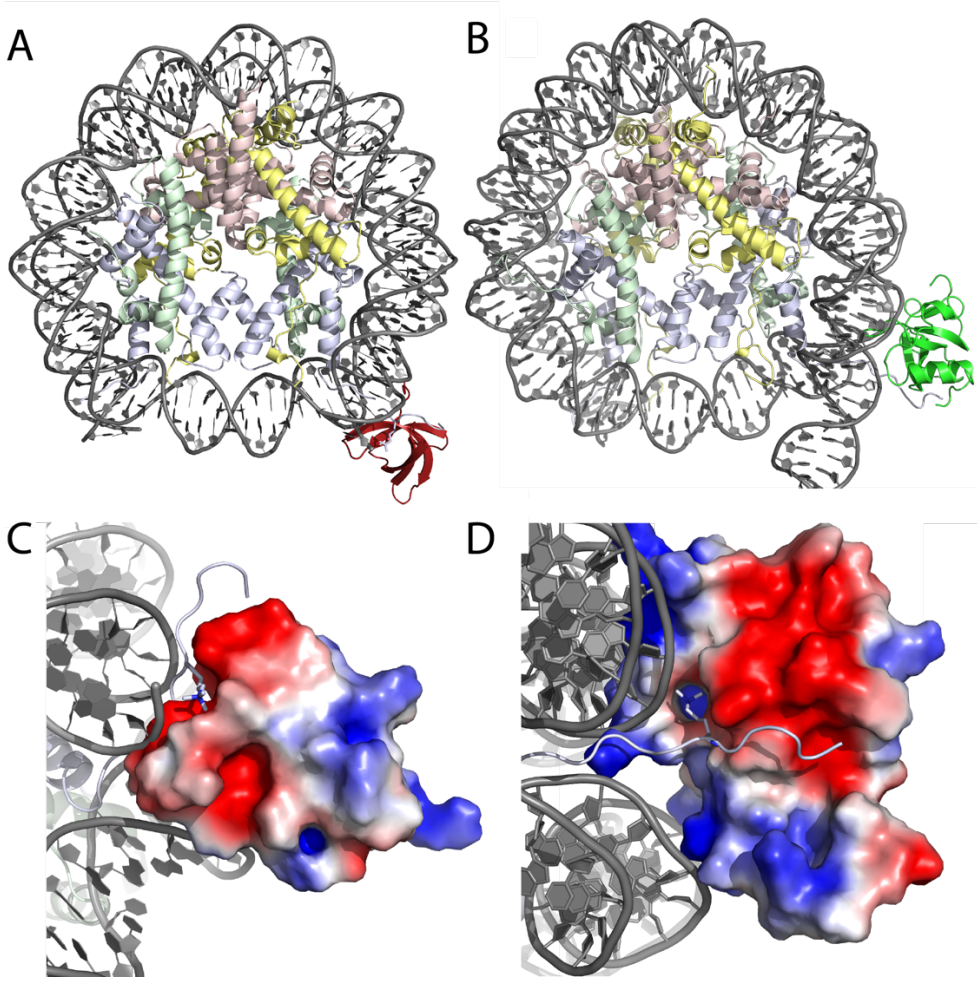


Figure 1-5. Structural model of nucleosome-bound PHF1 and PSIP1-PWWP.

Structural model of nucleosome-bound PHF1 (red) (A) and PSIP1-PWWP (green) (B). The electrostatic potential of nucleosomal DNA and the surface of PHF1 (C) and PWWP (D) respectively, act in combination with H3K36me3 recognition by the aromatic cage motif (trimethyllysine side chain shown as sticks). Figure generated using the author-provided PDB file [88].

THE STUDY OF LARGE PROTEIN COMPLEXES BY NMR

The above-mentioned studies focus on interactions of proteins with nucleosome surrogates such as peptide models or isolated DNA. The molecular size of nucleosomes, and even more so of complexes with effector proteins, poses a challenge to traditional NMR methods. However, this challenge can be overcome through the

use of methodologies designed for high-molecular weight systems. This method, methyl-group based transverse-relaxation-optimized spectroscopy (methyl-TROSY), relies on the highly sensitive observation of NMR signals of protein methyl groups [98]. Here, a specific isotope-labelling scheme is used that typically results in observation of isoleucine, leucine, valine (ILV) methyl groups. The methyl-TROSY NMR spectra can subsequently be used to delineate binding sites of effector proteins on the nucleosome surface and vice versa [67, 68, 96, 99]. Extracting more detailed structural information is possible through the use of so-called spin-labels that can generate long-range distance restraints between the interaction partners [100, 101].

Kato *et al.* were the first to use the methyl-TROSY approach for the study of nucleosome-protein interactions [96]. They reported the NMR signal assignments of the ILV-methyl groups for all histones in the nucleosomes. These assignments are essential in determining protein binding sites on the nucleosome surface. The approach was demonstrated using high mobility group nucleosomal protein 2 (HMGN2), which regulates a variety of chromatin functions. HMGN2 was found to bind both the acidic patch and nucleosomal DNA. Based on these NMR-data, supported by mutagenesis, it was possible to determine a structural model of the complex (Figure 1-6A). HMGN2 binds to the nucleosome as a staple, using two main interaction sites. On one side, HMGN2 is anchored to the acidic patch using a canonical arginine anchor in the N-terminal region of the binding domain, while the lysine rich motif in its C-terminal region binds to nucleosomal DNA (Figure 1-6B). This binding mode provided a structural basis for the antagonistic function of HMGN2 towards linker histone H1 for nucleosome binding.

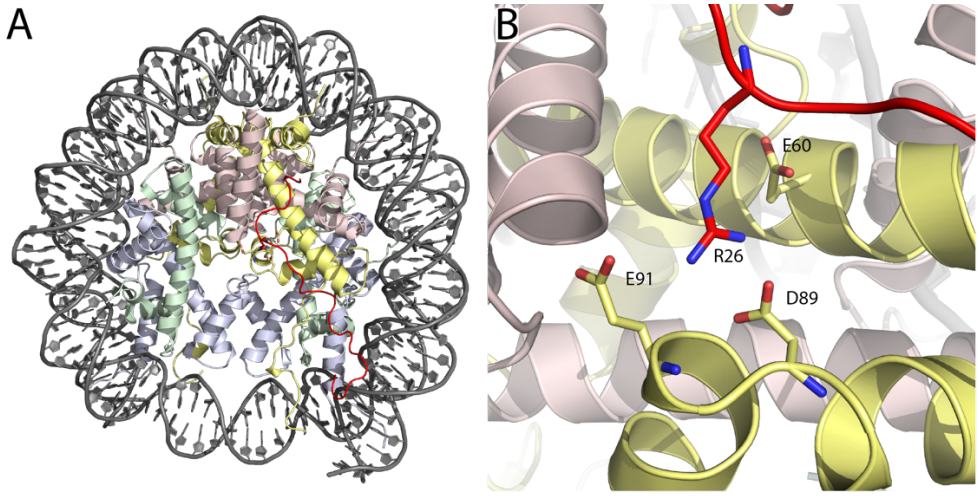


Figure 1-6. Structural model of nucleosome-bound HMGN2.

(A) Structural model of HMGN2 (red) bound to the nucleosome. The binding occurs along the nucleosome surface and is driven by interactions with the acidic patch and nucleosomal DNA, resulting in HMGN2 competing with H1 for nucleosome binding. (B) Close view on the acidic patch binding N-terminal HMGN2 region depicting the canonical arginine anchor R26 surrounded by the Glu 91, Asp 89, Glu 60 acidic triad motif of H2A. Figure generated using the author-provided PDB file [96].

LATEST APPLICATIONS OF NMR TO INVESTIGATE STRUCTURES OF NUCLEOSOME COMPLEXES

Two recent studies relied on methyl-TROSY NMR derived binding data to elucidate the recognition of ubiquitinated nucleosomes. Both focused on the interaction between ubiquitylated H2A K13/15 and the DNA repair factor RNF169. The work of Kitevski-LeBlanc *et al.* established the molecular basis of this interaction [68]. The α-helical MIU2 (motif interacting with ubiquitin) domain binds to a hydrophobic patch on the K13/15-conjugated ubiquitin while a disordered region anchors RNF169 on the nucleosome by binding to the acidic patch. They subsequently reconstructed a model structure that presents both epitopes in their nucleosome-bound state (Figure 1-7A). The work of Hu *et al.* combined traditional NOESY-based structure determination at the level of histone-dimers with interaction studies at the nucleosome level and complemented these with SAXS data into a final model [92]. The authors also extended their findings to an NMR-based structural model for the complex with DNA repair factor Rad18. Both RNF169 as well as Rad18 are known to interfere with the binding of 53BP1 to nucleosomes ubiquitinated at H2A K13/15.

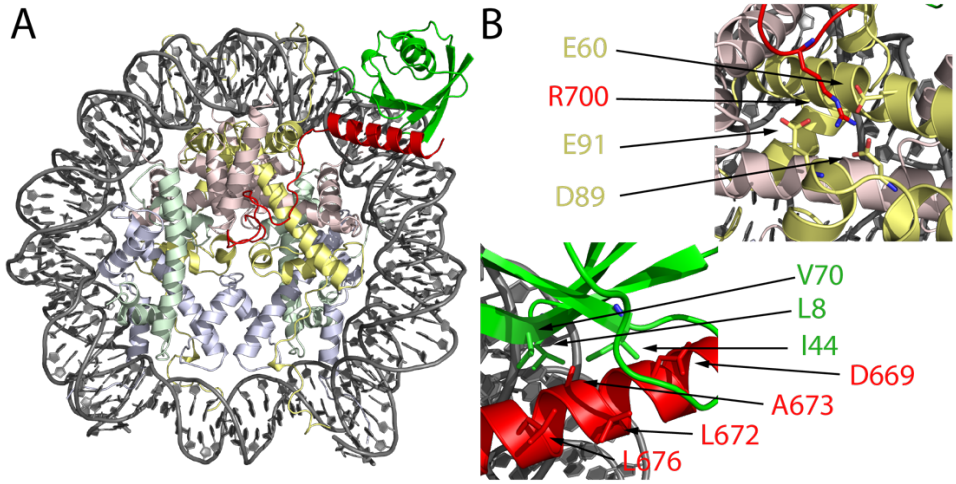


Figure 1-7. Structural model of nucleosome-bound RNF169.

(A) Structural model of nucleosome-bound RNF169 (red) and ubiquitin (green). (B, top) The proposed main acidic patch anchoring residue R700 (conserved position throughout the docking solutions) is shown in the conserved arginine anchor position between the acidic triad (Glu 60, Asp 89, Glu 91). (B, bottom) Side chain interactions between RNF169 MIU2 (red) and ubiquitin (green). Figure generated using the author-provided PDB file [68].

A STRUCTURAL MODEL DERIVED FROM SOLID STATE NMR

The studies mentioned above illustrate the potential of data-driven modelling of nucleosome-protein complexes based on state-of the-art solution NMR. Recent advances in solid-state NMR (ssNMR) have enabled the detailed investigation of large, soluble biomolecular complexes. Very recently, our lab capitalized on these advances and tailored them for application to nucleosome-protein complexes [90]. Unlike the methyl-TROSY methods, this approach allows observation of all residues, in principle allowing for a more complete mapping of binding interfaces. In this approach NMR spectra are recorded on sediments, generated by ultracentrifugation, of nucleosomes or their complexes. After assignments of NMR signals of histone H2A in the unbound nucleosome, spectra were recorded on the nucleosome complex with the LANA peptide, analogous to the LANA crystal structure (Figure 1-8A) [60, 90]. Based on the chemical shift changes the binding site of LANA could be mapped to the acidic patch and a structural model generated. The large agreement between the crystal structure and ssNMR-derived structural model (Figure 1-8B) illustrates the power of this approach. In our view ssNMR, just as the solution NMR approach, is an attractive alternative for structure determination for nucleosome-protein complexes. While its

application awaits to be extended to larger nucleosome-binding domains, we anticipate that it will be a valuable addition to the tool kit in chromatin structural biology.

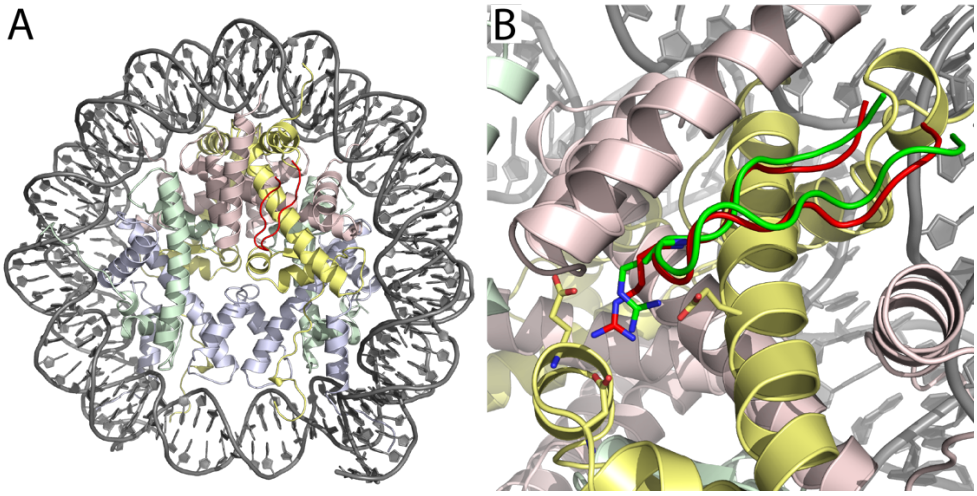


Figure 1-8. Structural model of nucleosome-bound LANA.

(A) Structural model for nucleosome-bound LANA peptide. ssNMR data derived from NMR titration experiments were used to direct the docking simulation. (B) Alignment of the ssNMR derived model for LANA (red) and the crystal structure (green, pdb: 1zla) shows remarkable accuracy of the docking derived solution. For both, the canonical arginine anchor is depicted as sticks in the typical central position between the acidic triad of H2A (yellow).

CROSS-LINKING MASS SPECTROMETRY IN DATA-DRIVEN DOCKING

Next to NMR, cross-linking mass spectrometry has found increasing application as a data source on nucleosome-protein interactions. With cross-linking, covalent intra-and intermolecular connections of proteins are established that withstand denaturing and digestion. These connections are introduced by small molecule linkers, specific for the fusion of well-defined side chains or less specific as radical-forming photo cross-linkers. Furthermore, cross-linkers possess a spacer between their terminal functional groups to define the range of cross-linking ability [102, 103]. Both characteristics can be tuned for the study of a specific system, resulting in a manifold of reported linker molecules. After cross-linking, the protein complex undergoes trypsin digestion resulting in peptide fragments of the complex. Here, covalently cross-linked fragments stay connected. An analysis of these fragments by

liquid chromatography mass spectrometry (LC-MS) enables identification of the sequence positions. Due to functionality and limited length of cross-linker molecules, the cross-linked fragments can be assigned spatial proximity and distance restraints are obtained to build a structural model of a complex [83]. In one of the earliest examples for nucleosome complexes, XL-MS was used to map the binding sites of the various nucleosome binding domains of the chromatin remodeling complex ISW2 onto the nucleosome surface [94]. These data were subsequently used to build a structural model of the ISW2-nucleosome complex. A recent case of cross-linking-based modelling in nucleosome research is the E2/E3 ubiquitin ligase complex Rad6-Bre1 (Figure 1-8A). Bre1 is known to act as a homodimer in a complex with Rad6 to specifically ubiquitinate H2B K123 [104, 105]. However, the molecular mechanism of specific ubiquitination remained unknown without any nucleosome-bound complex structure available. Gallego *et al.* addressed exactly this problem by using XL-MS data to identify the binding interface between the Bre1 RING domain and the nucleosome. Next to nucleosomal DNA binding, they observed binding of the homodimer to the acidic patch (Figure 1-8B), which was verified by LANA-induced inhibition of Bre1 RING nucleosome binding. As a first step in the modelling, the authors modelled the Rad6-Bre1 complex structures based on homology with known E2/E3 RING ligases. Importantly, the resulting model was supported by the observed cross-links. The Rad6-Bre1 model could then be docked onto the nucleosome guided by the observed cross-links. This provided the structural basis for the specificity of Bre1 towards H2B K123 ubiquitination by highlighting the interactions of acidic patch and DNA with one E3 monomer respectively, thus causing a well-defined defined orientation [69].

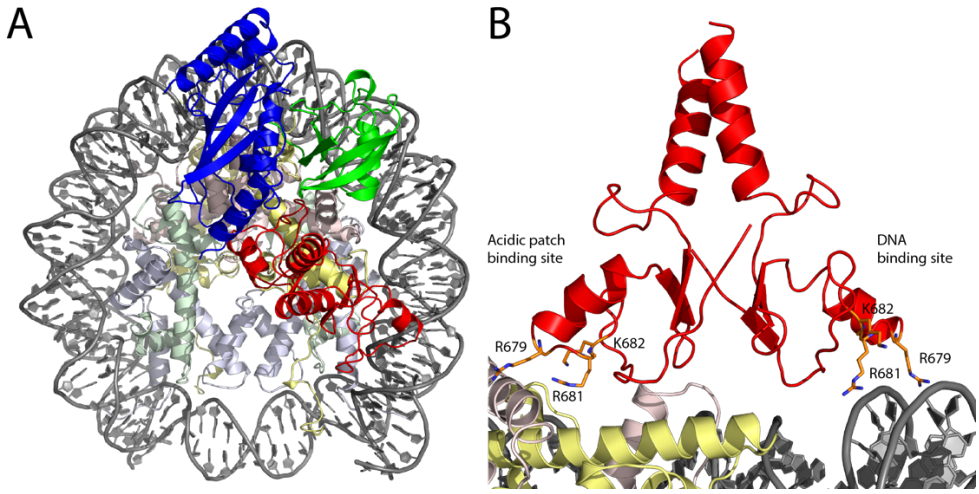


Figure 1-9. Structural model of nucleosome-bound Bre1.

(A) Structural model of homodimeric Bre1 (red) bound to the nucleosome together with the E2 ligase Rad6 (blue) with attached ubiquitin (green). The study was conducted by identifying the interactions between positive Bre1 RING residues and the acidic patch. The docking was further facilitated due to the known target lysine residue. (B) Close view on Bre1 bound to both the acidic patch and nucleosomal DNA. The homodimeric nature allows the engagement of both epitopes in simultaneous binding. Figure generated using the author-provided PDB file [69].

CONCLUSIONS

Chromatin structural biology is an equally important as demanding field. This is not only clear from the tremendous efforts necessary for the first nucleosome structure but also from the limited number of structures for nucleosome-protein complexes. While crystallography and cryo-EM resulted in various high-resolution structures, not every interaction is accessible this way due to experimental limitations, such as the need for crystallization, the fleeting nature of some complexes or the pervasive role of highly dynamic protein regions. An increasing number of studies address this by following an integrative approach, utilizing various sources of interaction data to direct sophisticated data-driven docking. This way available knowledge on a nucleosome-interacting system can be integrated into a structural model that is otherwise inaccessible. These models strongly depend on the quality and quantity of data and contain an inherent ambiguity. However, they can indicate alternative binding modes and conformational ambiguity. Additionally, crucial residues for nucleosome binding can be identified, allowing design of loss of function or loss of binding mutants to silence specific pathways. Structural models also offer the

possibility to drive the design of competing small molecule or peptide structures as potential candidates for epigenetic drugs that interfere with specific effector binding. Such developments might otherwise not occur due to the lack of a structure. Recently, a database for such structural models, akin to the RCSB protein databank, was established [106, 107]. This might prove to be essential to advance the study of chromatin effector proteins. Publicly available structures, including their data-based restraints can be used for further refinements upon availability of new, additional datasets. It also offers the possibility of negative results, otherwise rarely reported, to contribute to the validation of already reported models. As in the land of the blind, the one-eyed man is king, data-driven modelling of nucleosome-protein complexes has the potential to yield unique fundamental insights into nucleosome binding dynamics and enable advances in modulation of chromatin effector proteins, which would be otherwise inaccessible.

SCOPE OF THIS THESIS

The studies presented in this thesis utilize the advantages of integrative usage of information on protein interactions. Various biophysical techniques were combined into data-driven computational structural biology studies to gain new insights into nucleosome-protein complexes and, based on structural models, specifically modulate their interactions.

Chapter 1 offers a brief introduction to how nucleosomal DNA is organized in eukaryotic nuclei and highlights the importance of epigenetic effector proteins in virtually all DNA related processes in the cell. The reader is provided with a review of nucleosome complex structures and the corresponding biophysical techniques used to obtain them.

Chapter 2 presents a structural model of the E3 ubiquitin ligase RNF168 RING domain bound to its nucleosomal substrate, obtained via data-driven docking based on experimental data from NMR, XL-MS, mutagenesis and activity assays. This complex structure was used to design a specific loss-of-function mutant, selectively silencing the RNF168 meditated ubiquitination pathway.

Chapter 3 aims to test hypothesized role of electrostatic interactions of the PSIP1-PWWP reader domain binding to its substrate, trimethylated lysine 36 on histone H3. The underlying structural model served as basis for the design of peptide-based ligands that combine the multi-epitope principle of nucleosome binding into a single molecule.

Chapter 4 focusses on facilitating the design and application of peptide-based mimics of epigenetic reader proteins. A synthesis protocol for cysteine-based trimethyllysine mimics is developed and microscale thermophoresis (MST) is applied for screening purposes, reducing both costs and time of interaction studies, such as of PSIP1-PWWP with its target.

REFERENCES

1. Kornberg, R.D., *Chromatin structure: A repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-871.
2. Thomas, J.O. and R.D. Kornberg, *An octamer of histones in chromatin and free in solution.* Proceedings of the National Academy of Sciences, 1975. **72**(7): p. 2626.
3. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8Å resolution.* Nature, 1997. **389**: p. 251.
4. Luger, K., M.L. Dechassa, and D.J. Tremethick, *New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?* Nature Reviews Molecular Cell Biology, 2012. **13**: p. 436.
5. Sarah, H., et al., *Use of Epigenetic Drugs in Disease: An Overview.* Genetics & Epigenetics, 2014. **6**: p. GEG.S12270.
6. Gijsbers, R., et al., *Role of the PWWP Domain of Lens Epithelium-derived Growth Factor (LEDGF)/p75 Cofactor in Lentiviral Integration Targeting.* Journal of Biological Chemistry, 2011. **286**(48): p. 41812-41825.
7. Yokoyama, A. and M.L. Cleary, *Menin critically links MLL proteins with LEDGF on cancer-associated target genes.* Cancer Cell, 2008. **14**(1): p. 36-46.
8. Egger, G., et al., *Epigenetics in human disease and prospects for epigenetic therapy.* Nature, 2004. **429**: p. 457.
9. Yu, X., Z. Li, and J. Shen, *BRD7: a novel tumor suppressor gene in different cancers.* American Journal of Translational Research, 2016. **8**(2): p. 742-748.
10. Pérez-Salvia, M. and M. Esteller, *Bromodomain inhibitors and cancer therapy: From structures to applications.* Epigenetics, 2017. **12**(5): p. 323-339.
11. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
12. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
13. Rothbart, S.B. and B.D. Strahl, *Interpreting the language of histone and DNA modifications.* Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2014. **1839**(8): p. 627-643.
14. Li, Y., et al., *Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain.* Molecular Cell, 2016. **62**(2): p. 181-193.
15. Andrews, F.H., et al., *The Taf14 YEATS domain is a reader of histone crotonylation.* Nature Chemical Biology, 2016. **12**: p. 396.
16. Zhao, D., et al., *YEATS2 is a selective histone crotonylation reader.* Cell Research, 2016. **26**: p. 629.
17. Filippakopoulos, P. and S. Knapp, *Structural Genomics and Drug Discovery for Chromatin-Related Protein Complexes Involved in Histone Tail Recognition*, in *Systems Analysis of Chromatin-Related Protein Complexes in Cancer*, A. Emili, J. Greenblatt, and S. Wodak, Editors. 2014, Springer New York: New York, NY. p. 211-225.

18. Hughes, R.M., et al., *Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect.* Proceedings of the National Academy of Sciences, 2007. **104**(27): p. 11184.
19. Chen, C., et al., *Deciphering arginine methylation: Tudor tells the tale.* Nature Reviews Molecular Cell Biology, 2011. **12**: p. 629.
20. Maurer-Stroh, S., et al., *The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains.* Trends in Biochemical Sciences, 2003. **28**(2): p. 69-74.
21. Teske, K.A. and M.K. Hadden, *Methyllysine binding domains: Structural insight and small molecule probe development.* European Journal of Medicinal Chemistry, 2017. **136**: p. 14-35.
22. Gupta, A., et al., *Role of 53BP1 in the Regulation of DNA Double-Strand Break Repair Pathway Choice.* Radiation research, 2014. **181**(1): p. 1-8.
23. Yang, Y., et al., *TDRD3 Is an Effector Molecule for Arginine-Methylated Histone Marks.* Molecular Cell, 2010. **40**(6): p. 1016-1023.
24. Trojer, P., et al., *L3MBTL1, a Histone-Methylation-Dependent Chromatin Lock.* Cell, 2007. **129**(5): p. 915-928.
25. Boccuni, P., et al., *The Human L(3)MBT Polycomb Group Protein Is a Transcriptional Repressor and Interacts Physically and Functionally with TEL (ETV6).* Journal of Biological Chemistry, 2003. **278**(17): p. 15412-15420.
26. Daugaard, M., et al., *LEDGF (p75) promotes DNA-end resection and homologous recombination.* Nature Structural &Amp; Molecular Biology, 2012. **19**: p. 803.
27. Ge, H., Y. Si, and R.G. Roeder, *Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation.* The EMBO Journal, 1998. **17**(22): p. 6723-6729.
28. Flanagan, J.F., et al., *Double chromodomains cooperate to recognize the methylated histone H3 tail.* Nature, 2005. **438**: p. 1181.
29. Sims, R.J., et al., *Human but Not Yeast CHD1 Binds Directly and Selectively to Histone H3 Methylated at Lysine 4 via Its Tandem Chromodomains.* Journal of Biological Chemistry, 2005. **280**(51): p. 41789-41792.
30. Eissenberg, J.C. and S.C.R. Elgin, *The HP1 protein family: getting a grip on chromatin.* Current Opinion in Genetics & Development, 2000. **10**(2): p. 204-210.
31. Luco, R.F., et al., *Regulation of Alternative Splicing by Histone Modifications.* Science, 2010. **327**(5968): p. 996.
32. Alpatov, R., et al., *A Chromatin-Dependent Role of the Fragile X Mental Retardation Protein FMRP in the DNA Damage Response.* Cell, 2014. **157**(4): p. 869-881.
33. LeRoy, G., B. Rickards, and S.J. Flint, *The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription.* Molecular Cell, 2008. **30**(1): p. 51-60.
34. Winter, S., et al., *14-3-3 Proteins recognize a histone code at histone H3 and are required for transcriptional activation.* The EMBO Journal, 2008. **27**(1): p. 88-99.
35. Yun, M., et al., *Readers of histone modifications.* Cell Research, 2011. **21**(4): p. 564-578.

36. Barbieri, I., E. Cannizzaro, and M.A. Dawson, *Bromodomains as therapeutic targets in cancer.* Briefings in Functional Genomics, 2013. **12**(3): p. 219-230.
37. Xu, C., et al., *Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2).* Proceedings of the National Academy of Sciences, 2010. **107**(45): p. 19266.
38. McGinty, R.K. and S. Tan, *Recognition of the nucleosome by chromatin factors and enzymes.* Curr Opin Struct Biol, 2016. **37**: p. 54-61.
39. Armache, K.-J., et al., *Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution.* Science (New York, N.Y.), 2011. **334**(6058): p. 977-982.
40. Wilson, M.D., et al., *The structural basis of modified nucleosome recognition by 53BP1.* Nature, 2016. **536**: p. 100.
41. Carroll, C.W., K.J. Milks, and A.F. Straight, *Dual recognition of CENP-A nucleosomes is required for centromere assembly.* The Journal of Cell Biology, 2010. **189**(7): p. 1143.
42. Carroll, C.W., et al., *Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N.* Nat Cell Biol, 2009. **11**(7): p. 896-902.
43. Zhou, B.R., et al., *Structural insights into the histone H1-nucleosome complex.* Proc Natl Acad Sci U S A, 2013. **110**(48): p. 19390-5.
44. Fyodorov, D.V., et al., *Emerging roles of linker histones in regulating chromatin structure and function.* Nature Reviews Molecular Cell Biology, 2017. **19**: p. 192.
45. Bednar, J., et al., *Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1.* Molecular Cell, 2017. **66**(3): p. 384-397.e8.
46. Cirillo, L.A., et al., *Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4.* Molecular Cell, 2002. **9**(2): p. 279-289.
47. Soufi, A., G. Donahue, and Kenneth S. Zaret, *Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome.* Cell, 2012. **151**(5): p. 994-1004.
48. Soufi, A., et al., *Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming.* Cell, 2015. **161**(3): p. 555-568.
49. Makde, R.D., et al., *Structure of RCC1 chromatin factor bound to the nucleosome core particle.* Nature, 2010. **467**(7315): p. 562-566.
50. Wang, F., et al., *Heterochromatin protein Sir3 induces contacts between the amino terminus of histone H4 and nucleosomal DNA.* Proc Natl Acad Sci U S A, 2013. **110**(21): p. 8495-500.
51. Arnaudo, N., et al., *The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle.* Nat Struct Mol Biol, 2013. **20**(9): p. 1119-21.
52. Yang, D., et al., *Nalpha-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain.* Nat Struct Mol Biol, 2013. **20**(9): p. 1116-8.
53. Kato, H., et al., *A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C.* Science (New York, N.Y.), 2013. **340**(6136): p. 1110-1113.

54. McGinty, R.K., R.C. Henrici, and S. Tan, *Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome.* Nature, 2014. **514**(7524): p. 591-596.
55. Morgan, M.T., et al., *Structural basis for histone H2B deubiquitination by the SAGA DUB module.* Science (New York, N.Y.), 2016. **351**(6274): p. 725-728.
56. Girish, T.S., R.K. McGinty, and S. Tan, *Multivalent Interactions by the Set8 Histone Methyltransferase With Its Nucleosome Substrate.* J Mol Biol, 2016. **428**(8): p. 1531-43.
57. Pinotsis, N. and G. Waksman, *Crystal structure of the Legionella pneumophila Lpg2936 in complex with the cofactor S-adenosyl-L-methionine reveals novel insights into the mechanism of RsmE family methyltransferases.* Protein Science : A Publication of the Protein Society, 2017. **26**(12): p. 2381-2391.
58. Eustermann, S., et al., *Structural basis for ATP-dependent chromatin remodelling by the INO80 complex.* Nature, 2018. **556**(7701): p. 386-390.
59. Ayala, R., et al., *Structure and regulation of the human INO80–nucleosome complex.* Nature, 2018. **556**(7701): p. 391-395.
60. Barbera, A.J., et al., *The Nucleosomal Surface as a Docking Station for Kaposi's Sarcoma Herpesvirus LANA.* Science, 2006. **311**(5762): p. 856.
61. England, J.R., et al., *RCC1 uses a conformationally diverse loop region to interact with the nucleosome: a model for the RCC1-nucleosome complex.* Journal of molecular biology, 2010. **398**(4): p. 518-529.
62. Kan, P.-Y., T.L. Caterino, and J.J. Hayes, *The H4 Tail Domain Participates in Intra- and Internucleosome Interactions with Protein and DNA during Folding and Oligomerization of Nucleosome Arrays.* Molecular and Cellular Biology, 2009. **29**(2): p. 538-546.
63. Mattiroli, F., et al., *The nucleosome acidic patch plays a critical role in RNF168-dependent ubiquitination of histone H2A.* Nature Communications, 2014. **5**: p. 3291.
64. Clarke, P.R. and C. Zhang, *Spatial and temporal coordination of mitosis by Ran GTPase.* Nature Reviews Molecular Cell Biology, 2008. **9**: p. 464.
65. Nishitani, H., et al., *Loss of RCC1, a nuclear DNA-binding protein, uncouples the completion of DNA replication from the activation of cdc2 protein kinase and mitosis.* The EMBO Journal, 1991. **10**(6): p. 1555-1564.
66. Eidahl, J.O., et al., *Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes.* Nucleic Acids Res, 2013. **41**(6): p. 3924-36.
67. van Nuland, R., et al., *Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain.* Epigenetics & Chromatin, 2013. **6**(1): p. 12.
68. Kitevski-LeBlanc, J., et al., *The RNF168 paralog RNF169 defines a new class of ubiquitylated histone reader involved in the response to DNA damage.* Elife, 2017. **6**.
69. Gallego, L.D., et al., *Structural mechanism for the recognition and ubiquitination of a single nucleosome residue by Rad6–Bre1.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(38): p. 10553-10558.
70. Liu, X., et al., *Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure.* Nature, 2017. **544**(7651): p. 440-445.

71. Chittori, S., et al., *Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N.* Science, 2018. **359**(6373): p. 339.
72. Pentakota, S., et al., *Decoding the centromeric nucleosome through CENP-N.* Elife, 2017. **6**.
73. Farnung, L., et al., *Nucleosome-Chd1 structure and implications for chromatin remodelling.* Nature, 2017. **550**(7677): p. 539-542.
74. Sundaramoorthy, R., et al., *Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome.* eLife, 2018. **7**: p. e35720.
75. Zhou, B.-R., et al., *Structural Mechanisms of Nucleosome Recognition by Linker Histones.* Molecular Cell, 2015. **59**(4): p. 628-638.
76. Kujirai, T., et al., *Structural basis of the nucleosome transition during RNA polymerase II passage.* Science, 2018. **362**(6414): p. 595.
77. Willhoft, O., et al., *Structure and dynamics of the yeast SWR1-nucleosome complex.* Science, 2018. **362**(6411).
78. Worden, E.J., et al., *Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L.* Cell, 2019.
79. Amamoto, Y., et al., *Synthetic Posttranslational Modifications: Chemical Catalyst-Driven Regioselective Histone Acylation of Native Chromatin.* J Am Chem Soc, 2017. **139**(22): p. 7568-7576.
80. Fang, Q., et al., *Human cytomegalovirus IE1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome.* eLife, 2016. **5**: p. e11911.
81. Lesbats, P., et al., *Structural basis for spumavirus GAG tethering to chromatin.* Proceedings of the National Academy of Sciences, 2017. **114**(21): p. 5509.
82. van Ingen, H. and A.M.J.J. Bonvin, *Information-driven modeling of large macromolecular assemblies using NMR data.* Journal of Magnetic Resonance, 2014. **241**: p. 103-114.
83. Rappsilber, J., *The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes.* Journal of Structural Biology, 2011. **173**(3): p. 530-540.
84. Xue, L.C., et al., *Computational prediction of protein interfaces: A review of data driven methods.* FEBS Letters, 2015. **589**(23): p. 3516-3526.
85. Clore, G.M. and C.D. Schwieters, *Docking of Protein−Protein Complexes on the Basis of Highly Ambiguous Intermolecular Distance Restraints Derived from 1HN/15N Chemical Shift Mapping and Backbone 15N−1H Residual Dipolar Couplings Using Conjoined Rigid Body/Torsion Angle Dynamics.* Journal of the American Chemical Society, 2003. **125**(10): p. 2902-2912.
86. Dominguez, C., R. Boelens, and A.M.J.J. Bonvin, *HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information.* Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.
87. van Dijk Aalt, D.J., R. Boelens, and M.J.J. Bonvin Alexandre, *Data-driven docking for the study of biomolecular complexes.* The FEBS Journal, 2004. **272**(2): p. 293-312.
88. Musselman, C.A., et al., *Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility.* Nat Commun, 2013. **4**: p. 2969.

89. Yang, M., et al., *Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase.* Mol Cell, 2006. **23**(3): p. 377-87.
90. Xiang, S., et al., *Site-Specific Studies of Nucleosome Interactions by Solid-State NMR.* Angewandte Chemie International Edition: p. n/a-n/a.
91. Qiao, Q., et al., *The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation.* J Biol Chem, 2011. **286**(10): p. 8361-8.
92. Hu, Q., et al., *Mechanisms of Ubiquitin-Nucleosome Recognition and Regulation of 53BP1 Chromatin Recruitment by RNF168/169 and RAD18.* Mol Cell, 2017. **66**(4): p. 473-487 e9.
93. Zhou, B.R., et al., *A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome.* J Mol Biol, 2016. **428**(20): p. 3948-3959.
94. Dang, W. and B. Bartholomew, *Domain architecture of the catalytic subunit in the ISW2-nucleosome complex.* Mol Cell Biol, 2007. **27**(23): p. 8306-17.
95. Musselman, C.A., et al., *Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1.* Nature Structural &Amp; Molecular Biology, 2012. **19**: p. 1266.
96. Kato, H., et al., *Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR.* Proc Natl Acad Sci U S A, 2011. **108**(30): p. 12283-8.
97. Forneris, F., et al., *Human Histone Demethylase LSD1 Reads the Histone Code.* Journal of Biological Chemistry, 2005. **280**(50): p. 41360-41365.
98. Ollerenshaw, J.E., V. Tugarinov, and L.E. Kay, *Methyl TROSY: explanation and experimental verification.* Magnetic Resonance in Chemistry, 2003. **41**(10): p. 843-852.
99. Miller, T.C.R., et al., *A bromodomain–DNA interaction facilitates acetylation-dependent bivalent nucleosome recognition by the BET protein BRDT.* Nature Communications, 2016. **7**: p. 13855.
100. Hass, M.A.S. and M. Ubbink, *Structure determination of protein–protein complexes with long-range anisotropic paramagnetic NMR restraints.* Current Opinion in Structural Biology, 2014. **24**: p. 45-53.
101. Schilder, J., et al., *Protein docking using an ensemble of spin labels optimized by intramolecular paramagnetic relaxation enhancement.* Phys Chem Chem Phys, 2016. **18**(8): p. 5729-42.
102. Back, J.W., et al., *Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling.* Journal of Molecular Biology, 2003. **331**(2): p. 303-313.
103. Sinz, A., *Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes.* Journal of Mass Spectrometry, 2003. **38**(12): p. 1225-1237.
104. Turco, E., et al., *Monoubiquitination of Histone H2B Is Intrinsic to the Bre1 RING Domain-Rad6 Interaction and Augmented by a Second Rad6-binding Site on Bre1.* Journal of Biological Chemistry, 2015. **290**(9): p. 5298-5310.
105. Kumar, P. and C. Wolberger, *Structure of the yeast Bre1 RING domain.* Proteins, 2015. **83**(6): p. 1185-1190.

106. Burley, S.K., et al., *PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models*. Structure, 2017. **25**(9): p. 1317-1318.
107. Vallat, B., et al., *Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules*. Structure, 2018. **26**(6): p. 894-904.e2.

2

Structural Basis of Specific H2A K13/K15 Ubiquitination by RNF168

Based on the research article: **V. Horn**, M. Uckelmann, H. Zhang, J. Eerland, I. Aarsman

U. B. le Paige, C. Davidovich, T. Sixma and H. van Ingen (2019). *Structural basis of specific H2A*

K13/K15 ubiquitination by RNF168

Nature Communications 10, 1751

Ubiquitination assays, Electrophoretic mobility shift assays were performed by M. Uckelmann at the Biochemistry division, Netherlands Cancer Institute, Amsterdam (The Netherlands)

Crosslinking mass-spectrometry was performed by M. Uckelmann at the Department of Biochemistry & Molecular Biology, Monash University, Melbourne (Australia)

Expression and reconstitution of Nucleosome samples for NMR studies were performed by H. Zhang and U.B. le Paige at the Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht (The Netherlands)

Parts of NMR assignment were performed by J. Eerland at the Leiden Institute of Chemistry, Leiden University, Leiden (The Netherlands)

Expression and purification of RNF168 mutant constructs for sample quality control was performed by I. Aarsman at the Biochemistry division, Netherlands Cancer Institute, Amsterdam (The Netherlands)

ABSTRACT

Ubiquitination of chromatin by modification of histone H2A is a critical step in both regulation of DNA repair and regulation of cell fate. These very different outcomes depend on the selective modification of distinct lysine residues in H2A, each by a specific E3 ligase. While polycomb PRC1 complexes modify K119, resulting in gene silencing, the E3 ligase RNF168 modifies K13/15, which is a key event in the response to DNA double-strand breaks. The molecular origin of ubiquitination site specificity by these related E3 enzymes is one of the intriguing open questions in the field.

Using a combination of NMR spectroscopy, crosslinking mass-spectrometry, mutagenesis and data-driven modelling, we show that RNF168 binds the acidic patch on the nucleosome surface, directing the E2 to the target lysine. The structural model highlights the role of E3 and nucleosome in promoting the ubiquitination and provides a basis for understanding and engineering of chromatin ubiquitination specificity.

INTRODUCTION

The covalent attachment of ubiquitin to target proteins is a key post-translational modification in almost any eukaryotic cellular pathway. The process relies on an E1-E2-E3 conjugation machinery, where the ubiquitin E3 ligases catalyze the discharge of ubiquitin from a charged E2 to a target lysine[1,2]. The E2 discharge is facilitated through conformational selection, where the E3 engages in specific interactions with E2 and ubiquitin. This stabilizes the ubiquitin-charged E2 in a conformation that leaves the active site prone to aminolysis[3-6].

How target lysines are selected and how selective this process is, varies. While most E3 ligases likely ubiquitinate target proteins rather unselectively[7,8], some E3 ligases show a remarkable specificity for a certain group of lysines[2,9-11]. Lysine selection here is achieved through specific contacts between the E3 and the target protein. This interaction orients the E2 active site towards the appropriate target lysine[1,2,12]. The E3-substrate interactions that guide lysine selection are likely unique for each E3-substrate pair. An interesting model system to study target lysine selection by the E3 ligase is ubiquitination of nucleosomes, especially of histone H2A where multiple E3 ligases modify distinct sites.

Three separate E3 ligases, RNF168, RING1B/BMI1 and BRCA1/BARD1, each ubiquitinate a unique site on H2A, each with a distinct biological outcome[9-11]. Ubiquitination at lysine 118/119 (K119) by PRC1 polycomb complexes, such as RING1B/BMI1 is important for transcriptional silencing[13] and may play a role in the regulation of DNA double strand break (DSB) repair[14-16]. Ubiquitination at lysine 13/15 (K15) by RNF168 and at lysines 125/127/129 (K129) by BRCA1/BARD1 both are crucial in the DNA damage response. The K15 mark acts as a binding platform for specific recruitment of 53BP1[17,18], a protein responsible for establishing a DNA-end resection block[19,20]. The BRCA1 mark at K129 leads to recruitment of SMARCAD1, a chromatin remodeler that facilitates extended DNA-end resection[21]. The balance between K15 and K129 ubiquitination is thought to determine the choice between DSB repair by either non-homologous end joining or one of the homology repair pathways[14,21,22].

Initial understanding how these RING-type E3 ligases can target a nucleosome has come from the spectacular crystal structure of the complex of the E2, the RING domains of RING1B/BMI1 and the nucleosomal core particle. In this structure RING1B/BMI1 interacts with a cluster of acidic residues on the nucleosome surface

to orient UbcH5C (Ube2D3), the cognate E2, towards K119[12]. Based on biochemical characterizations a similar mode of interaction was proposed for BRCA1/BARD1, though direct structural evidence is missing in this case[12]. For RNF168, mutagenesis studies have shown that part of the acidic patch is crucial for RNF168 activity but not for direct recruitment[23,24] and that a basic residue on RNF168 (R57) is essential for binding the nucleosome[9,23]. Strikingly, the presence of the nucleosome was observed to promote the rate of ubiquitination reaction, indicating it contributes to E2 activation. While the structure of RING1B/BMI1 in complex with the nucleosome provides valuable insight towards the mechanism of interaction, for RNF168 and BRCA1/BARD1 it is not yet clear how, on a molecular level, the acidic patch is employed to govern unique lysine selectivity, neither how the nucleosome contributes to E2 activation.

Here we provide mechanistic analysis of specific ubiquitination of H2A at K15 by RNF168. Using NMR interaction studies on both histone complexes and nucleosome substrates, we show that the RING domain of RNF168 binds the nucleosome acidic patch. Through ubiquitination assays, we identify multiple critical arginine residues in RNF168. Based on the combination of NMR, mutational analysis and crosslinking mass-spectrometry data, we present a data-driven integrative structure of the E3-substrate complex highlighting how the RING domain RNF168, while bound to the nucleosome, orients the E2 towards the target lysine. The structure suggests that the nucleosome surface contributes to E2 activation by promoting the formation of the closed, activated ubiquitin-charged E2 conformation. Using structure-guided mutagenesis, we identified a single mutation at the periphery of the binding site and distant from the target lysine that is sufficient to specifically interfere with RNF168 ubiquitination, while retaining RING1B activity. Our findings underscore the crucial role of the shape and electrostatic properties of acidic patch binding proteins in determining their precise binding modes and point to the intricate contribution of both E3 ligase as well as substrate in controlling activity and specificity of ubiquitination.

RESULTS

RNF168-RING BINDS THE ACIDIC PATCH USING AN ARG-RICH HELIX

RNF168 harbors multiple domains that are critical for its function in the DNA damage response (Figure 2-1A). It contains two ubiquitin interacting motifs (Ub-dependent DSB recruitment modules, UDMs): one is necessary for recruitment of RNF168 to the DNA damage site by binding ubiquitinated linker histone H1[25], while the second motif binds to ubiquitinated H2A, the product of RNF168 activity[26-28]. Its monomeric RING domain is required and sufficient to direct ubiquitination by E2 UbcH5c to H2A residues K13/K15[9,23]. To understand the substrate recognition and ubiquitination specificity of RNF168, we set out to identify how its RING domain ($RNF168^{RING}$) interacts with the nucleosome. Despite extensive efforts, structural analysis of the $RNF168^{RING}$-nucleosome interaction through crystallography was unsuccessful. Over recent years NMR spectroscopy has proven to be an attractive approach to map out the binding interfaces involved in nucleosome-protein interactions[29-34]. Since RNF168-UbcH5c can specifically ubiquitinate both H2A-H2B dimers (25 kDa) and nucleosomes (200 kDa), we mapped the interaction surface of the $RNF168^{RING}$ domain on histone dimers using traditional amide-backbone based NMR, and on the nucleosome, using methyl-group based NMR for high-molecular weight systems.

Titration of the $RNF168^{RING}$ domain to H2A/H2B dimers with either ^{15}N-labeled H2A or H2B caused specific changes in the NMR signals of H2A and H2B residues (Figure 2-1B and Supplementary Figure 2-1). Progressive addition of RING domain also caused increasing levels of precipitation, making it impossible to reach a fully bound state of the dimer. Poor solubility of the $RNF168^{RING}$-H2A-H2B complex was observed for a range of buffer conditions tested. Under the sub-stoichiometric conditions of the titration experiments, the dimer remains predominantly unbound such that the effect of binding is mostly visible through a peak intensity decrease for specific amide groups (Figure 2-1C). Resonances with most significant changes include residues L22, Q23, E63, E90, E91 and L92 of H2A and a region around E102 and T116 in H2B (*Dm.* histone residue numbering). These residues cluster in and around the acidic patch on the surface of the H2A/H2B dimer (Figure 2-1D). In addition, peak intensity changes and small chemical shift changes were observed for regions of the H2A/H2B surface that are occluded in the context of the nucleosome, for example the DNA-binding region around H2A H30 and the DNA and chaperone binding region

around H2B I51 (see Figure 2-1C,D and Supplementary Figure 2-1). Reasoning that specific binding should result in larger chemical shift differences and slower dissociation rates than unspecific binding, we used CPMG relaxation dispersion experiments to probe these parameters via their effect on the ^{15}N transverse relaxation rate, $R_{2,eff}$. Addition of RNF168^{RING} caused large dispersion in $R_{2,eff}$ values for V45, L103 and T116 that are located in the acidic patch area, but not for any of the resonances in other interfaces, such as I51 (Figure 2-1E and Supplementary Figure 2-1). Together with the absence of dispersion effects in the unbound state, this indicates that V45, L103 and T116 experience larger changes in chemical shift and/or have a higher population of the bound state than resonances such as I51. We thus conclude that RNF168^{RING} binds the H2A/H2B acidic patch, and that in context of the dimer additional, non-specific binding modes are possible.

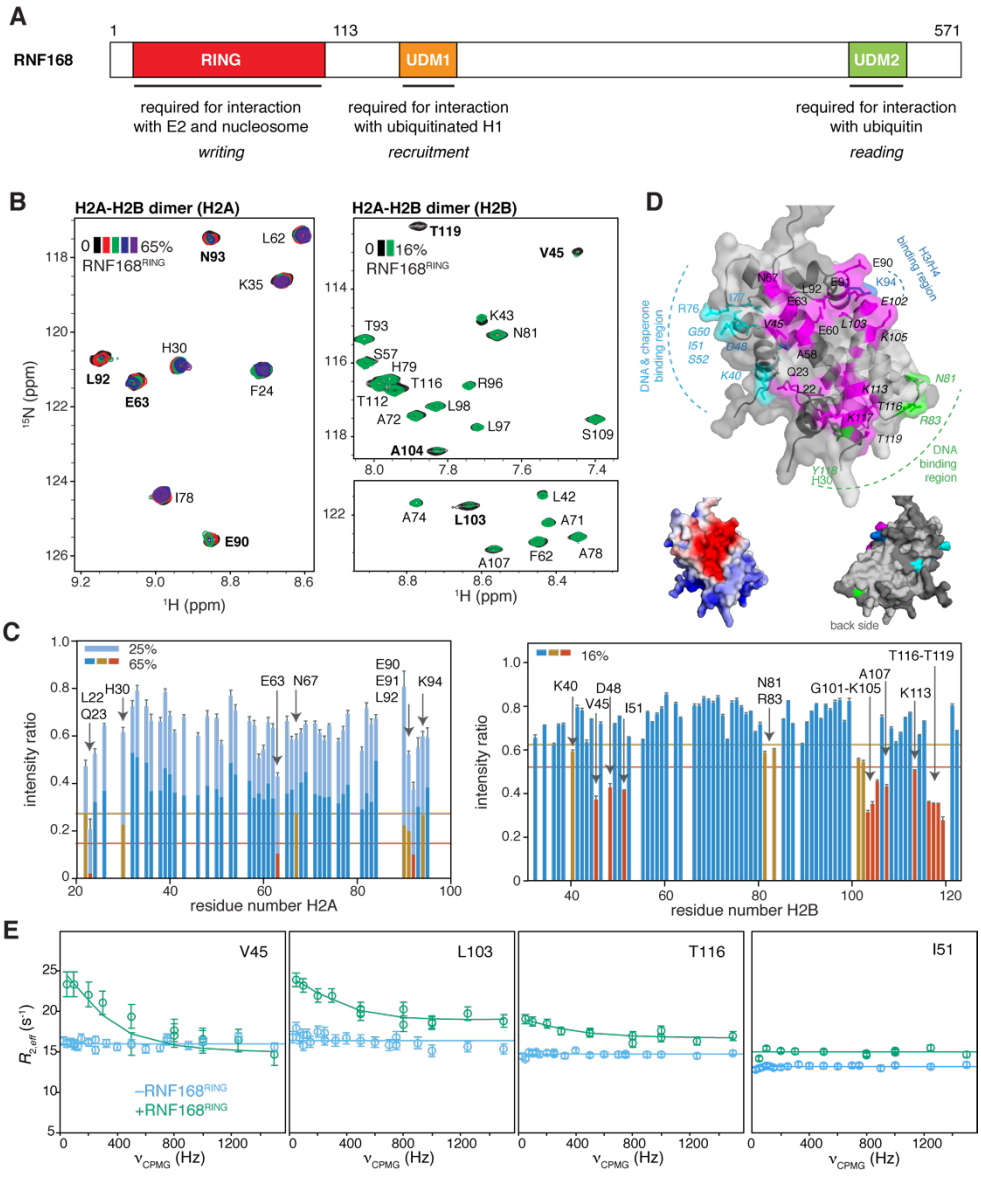


Figure 2-1. RNF168^{RING} binds to the acidic patch of H2A/H2B dimers.

(A) Domain architecture of RNF168 with domain function indicated. (B) Sections of the 2D ^{1}H-^{15}N correlation spectra of H2A (left) and H2B (right) within the H2A-H2B dimer with increasing amounts of RNF168^{RING} added. Color coding indicated. Residues strongly affected by RNF168 binding are labeled in bold. (C) Normalized peak intensity ratios for H2A (left) and H2B (right) upon addition of sub-stoichiometric amounts of RNF168^{RING} (indicated in figure). Residues with intensities that are one (two) standard deviation lower than the 10% trimmed mean are color coded in yellow (red) and labeled. (D) Residues with significant intensity/chemical shift changes (shown as sticks) cluster predominantly in and around the H2A-H2B acidic patch (magenta), with additional effects observed for the regions otherwise occluded

in the context of the nucleosome (indicated in blue, cyan and green). Residues in light/dark grey have no significant changes/no data available. Small insets show the electrostatic surface potential of the H2A/H2B dimer and the observed changes on the back side of the dimer. (E) Addition of RNF168^{RING} results in clear exchange-induced line broadening for H2B residues in and around the acidic patch (V45, L103, T116) but not for residues in non-specific binding regions (I51) as evidenced by ^{15}N CPMG relaxation dispersion. Color coding indicated in figure. Best-fits (solid lines) for the acidic patch cluster are consistent with low micromolar dissociation constants and ~1 ppm chemical shift changes. Error bars are s.d. based on noise levels in panel C and based on three replicate data points in panel E.

Next, these results were verified and extended by studying the interaction of the RNF168^{RING} domain with the nucleosome, using methyl-based transverse relaxation optimized NMR (methyl-TROSY)[35]. Unlabeled RING domain was added to milligram quantities of nucleosomes, reconstituted from 601-DNA and fully deuterated histones from *Drosophila melanogaster* (*Dm.*) in which the methyl groups of Ile, Leu and Val in H2A and H2B were ^{1}H,^{13}C-labeled (ILV-labeling). The high solubility of the RNF168^{RING}-nucleosome complex allowed to perform the titration up to 2:1 molar ratio of RING domain to nucleosome. Significant and specific chemical shift perturbations (CSPs) can be observed for the methyl group resonances of L64 and L92 in H2A and V45 and L103 in H2B (Figure 2-2A,B). These residues are part of or surround the acidic patch and thus reinforce the result from the dimer titration that the acidic patch is the specific binding site of the RNF168 RING domain (Figure 2-2C). The shifting resonances are in the fast-exchange regime indicating a dynamic interaction with a ~1 ms upper limit for the life-time of the complex.

Previously, it was shown that the region encompassing a basic helix (residues 57-72) is required for binding and effective ubiquitination of H2A/H2B dimers and nucleosomes[9,23]. In particular, an R57D mutation was shown to interfere with nucleosome binding and ubiquitination of RNF168^{RING}. Interestingly, the basic-helix and its C-terminal loop contain additional basic residues R56, R63, R67 and R68 that could potentially mediate interaction with the acidic patch (see Figure 2-2D) by acting as an arginine-anchor to bind the acidic patch[36]. Because of its central position in the putative nucleosome binding region, we tested nucleosome binding and ubiquitination activity of an RNF168 R63A mutant RING. While the mutant protein is folded, electrophoretic mobility shift assays (EMSA) indicate it binds nucleosomes with reduced affinity and in an altered binding mode (Supplementary Figure 2-2). Compared to wild-type RNF168^{RING}, the R63A mutant conjugated only trace levels of fluorescently labeled ubiquitin to nucleosomes (Figure 2-2E). Notably, we can rule out that the mutation caused ubiquitination of other histones since the readout in this experiment is the fluorescent ubiquitin itself.

To gain further insight in the nucleosome binding mode of RNF168^{RING}, we subjected a mixture of RNF168^{RING} and nucleosomes to crosslinking mass-spectrometry (XL-

MS). The RING domain of RNF168 contains five lysine residues, of which two (K37 and K46) are relatively close to the Arg-rich helix (Figure 2-2D). Using the amine-reactive bis(sulfosuccinimidyl)suberate as a cross-linking agent, a total of 62 unique crosslinks were obtained, mostly from lysines in the flexible histone tails (Figure 2-2F and Supplementary Figure 2-3). In total 21 crosslinks were found for the rigid histone core, the far majority in excellent agreement with the nucleosome crystal structure. Importantly, one intermolecular crosslink between RNF168 and the nucleosome was observed in all three replicate measurements. This crosslink links RNF168 K46, flanking the Arg-rich helix, to H2B K105, which lines the acidic patch (see also Figure 2-2C). Together, mutagenesis, NMR and XL-MS data indicate that RNF168^{RING} domain binds to the acidic patch using its arginine-rich helix.

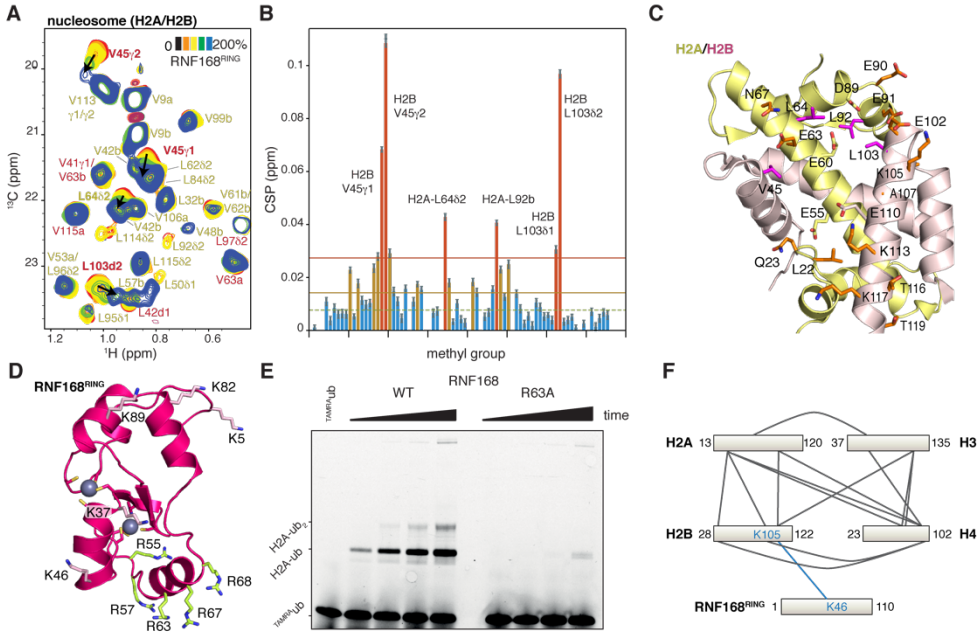


Figure 2-2. RNF168^{RING} binds the nucleosome acidic patch using its Arg-rich helix.

(A) Section of the 2D ^{1}H-^{13}C methyl-TROSY spectrum of the nucleosome with ILV-labeled H2A/H2B with increasing amounts of RNF168^{RING} added. Color coding indicated. Labels a/b refer to either of the δ1/δ2 Leu or γ1/γ2 Val methyl groups. (B) Weighted chemical shift perturbation (CSPs) between the spectra of 1:2 and 1:0 nucleosome:RNF168^{RING}. Residues with CSPs that are one (two) standard deviations larger than the 10% trimmed mean are highlighted in yellow (red) and labelled. (C) The RNF168 binding surface as defined by NMR (shown as sticks, color coded in orange/magenta for dimer/nucleosome data). Side chains of other acidic patch residues are shown as sticks, color coded in yellow for H2A and light red for H2B. (D) Crystal structure of the RNF168^{RING} domain with Arg residues of the basic-helix highlighted as yellow sticks. Lysine residues are shown as sticks and labelled, Zn atoms shown as spheres, coordinating residues as sticks. (E) Nucleosome ubiquitination by either wild-type (WT) or R63A mutant RNF168^{RING}. Fluorescent TAMRA-labeled ubiquitin was used to detect ubiquitinated species. Samples were resolved using SDS-PAGE. For both wild-type and mutant time points were taken at 5, 15, 30 and 60 min of incubation of the reaction

mixture. Molecular size of mono- and di-ubiquitinated H2A is indicated. (F) Schematic representation of inter-histone crosslinks (gray; for rigid histone core only) and intermolecular RNF168-nucleosome crosslinks (blue) obtained from XL-MS.

INTEGRATIVE STRUCTURE OF THE RNF168-NUCLEOSOME COMPLEX

To achieve a deeper understanding of the underlying molecular mechanism of RNF168 mediated H2A ubiquitination, we sought to construct a structural model of the E3-nucleosome and E2-E3-nucleosome complex. On the basis of the experimental NMR, XL-MS and mutagenesis data, we first docked $RNF168^{RING}$ to the nucleosome using the data-driven docking program HADDOCK[37,38]. The crosslink was implemented as an unambiguous distance restraint between the Cα atoms of the crosslinked residues with 28 Å upper limit[39]. Since the docking calculation imposes this restraint as a Euclidian distance rather than a surface accessible distance (SASD), solutions with SASD > 35 Å were rejected. The final ensemble consists of one large and one smaller cluster of solutions (see Supplementary Figure 2-4). The dominant cluster, shown in Figure 2-3A, contains the overall best scoring solution and has more favorable physio-chemical scores. In this structure, the $RNF168^{RING}$ basic helix is wedged between the H2A α2 and H2B αC helix and forms an extensive network of hydrogen bonding interactions with the acidic patch (Figure 2-3B). RNF168 residue R63 is hydrogen bonded to H2A E60, D89 and E91, thus fulfilling the role of the canonical arginine-anchor[40]. R57 hydrogen-bonds to E60 and E63 and R67 provides additional interactions to D89 and E91. In addition, there are favorable interactions involving R56 and S60 in most structures, either by electrostatic interactions or hydrogen-bonding.

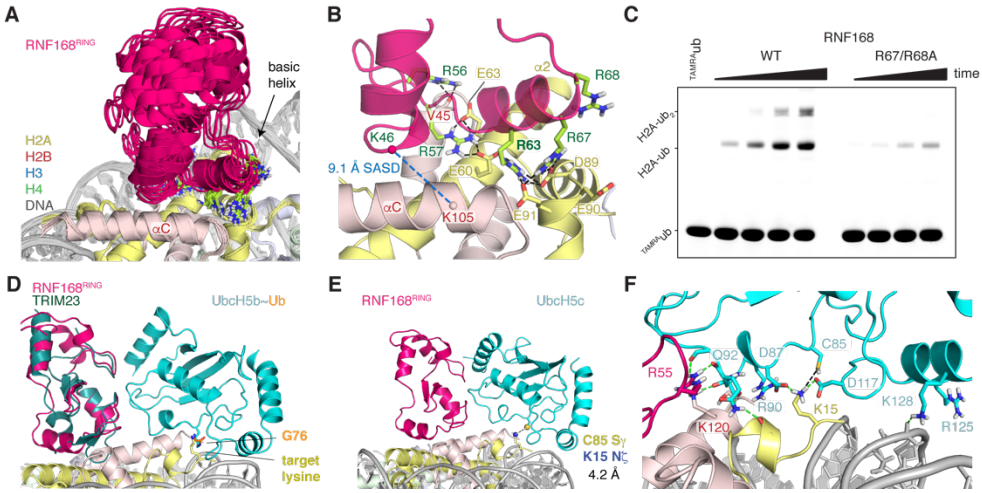


Figure 2-3. Structure of the RNF168^{RING}–nucleosome complex and implications for E2 positioning.

(A) Superposition of the 10 best scoring solutions of cluster 1 (containing 187 out of 200 solutions) calculated using HADDOCK. Color coding indicated. Docking statistics are reported in Supplementary Figure 2-4. (B) Zoom on the binding interface for the best scoring solutions of cluster 1, with key hydrogen-bond interactions involving the Arg-rich helix indicated. The Cα atoms and SASD of the crosslinked residues is indicated. (C) Nucleosome ubiquitination by either wild-type (WT) or R67/R68A mutant RNF168^{RING}. Fluorescent TAMRA-labeled ubiquitin was used to detect ubiquitinated species. Samples were resolved using SDS-PAGE. For both wild-type and mutant time points were taken at 2, 6, 18 and 40 min of incubation of the reaction mixture. Molecular size of mono- and di-ubiquitinated H2A is indicated. (D) Superposition of the TRIM23 E3 complex with Ub-conjugated E2 (PDB-ID 5VZW) and the RNF168^{RING}-nucleosome structure, showing close approximation of the Ub C-terminal residue G76 in the active site to the target lysine H2A K14. (E) Structural model of the ternary UbcH5c-RNF168^{RING}-nucleosome complex calculated using HADDOCK. The best scoring solution is shown, distance between the E2 active site and target lysine is indicated. (F) Zoom on the UbcH5c-nucleosome interface showing hydrogen bonds with green dashes. Panels e and f show models with human histones, residues are numbered accordingly.

The Cα-Cα SASD between the crosslinked residues is 9.1 Å for the best scoring solution and 8.7 ± 1.1 Å for the best 10 structures, well within the typical limit of 30 Å[39]. We further validated the model by predicting all possible crosslinks (SASD < 30 Å) between RNF168^{RING} and the nucleosome based on the 10 best structures (see Supplementary Figure 2-4). This yields five possible crosslinks, of which four pairs involve RNF168 K46. Experimentally, we only observed a single crosslink from RNF168 K46, to H2B K105. Since the SASD for this pair is much shorter than the other three, it suggests the observed crosslink is more readily formed, thus precluding the observation of the other predicted links. A fifth predicted crosslink is expected between RNF168 K37 and H2B K105 but this is not found. Again, H2B K105 may more readily crosslink to RNF K46 which is much closer in the structure (on average 8.7 vs. 22 Å SASD). We thus conclude that the model is in good agreement with experimental crosslinking data.

Since our structure predicts a direct role of RNF168 R67 in the interface, we performed an additional mutation study with an R67/68A double mutant to verify these results. Here, R68 was included in the mutation because of its proximity to R67. While the mutant protein binds with comparable affinity to nucleosomes as wild-type, the altered band-shift pattern indicates a change in binding mode (Supplementary Figure 2-2). The mutant exhibits strongly decreased ubiquitination activity (Figure 2-3C), suggesting the mutant has reduced ability to bind in a productive binding mode. As for the R63A mutant, we can exclude that ubiquitination activity is shifted to other histones than H2A. The residual activity observable is in contrast to the near inactive R57A[23]and R63A mutant (Figure 2-2E). This correlates well with the more extensive hydrogen bonding by R57 and R63 compared to R67.

Crucially, ubiquitination target specificity is determined by the interplay between E2, Ub-conjugated E2 (E2~Ub) and the substrate. We thus selected the E3-E2~Ub complex structure that most closely resembled RNF168^{RING} E3 structure and superimposed this on our RNF168^{RING}–nucleosome complex structure (Figure 2-3D). Strikingly, this results in the placement of the E2 active site with the conjugated C-terminus of Ub directly above the target lysines. Since the superposition introduces a few steric clashes between the E2 and the nucleosome surface, we next used HADDOCK to model the ternary complex between E3 RNF168^{RING} domain, E2 enzyme and the nucleosome (containing human histones). While the cognate E2 of RNF168 is yet unknown, we selected UbcH5c as it can function as such in vitro[9]. Restraints for the modelling were based on structurally conserved E3-E2 interactions, including the so-called 'linchpin' hydrogen bond, which is crucial for E2 activation[6] (see Supplementary Figure 2-5). In RNF168 this involves hydrogen-bond between R55 and the backbone of E2-Q92. Importantly, proximity of the E2 to the target lysines is not enforced during docking. Since the HADDOCK protocol allows for limited flexibility, clashes in the intermolecular interfaces can be removed and their physio-chemical quality optimized.

The resulting best scoring model places the E2 catalytic center in close proximity to the H2A K15 (4.2 Å Sγ-Nζ distance) without introducing steric clashes (Figure 2-3E). Within the twenty best solutions, the closest distance is 3.3 Å. The second lysine ubiquitinated by RNF168, H2A K13 is not present in the model. Both from the absence of electron density in many crystal structures and from recent NMR studies on the nucleosome[34], it is clear that this region of the H2A N-terminal tail is dynamic. Superposition with nucleosome structures that include this side chain shows that K13 is still reasonably close such that with some backbone and sidechain reorientation it can get close enough to the active site to be ubiquitinated (Supplementary Figure 2-6).

A lysine in the H2B αC-helix, Hs. H2B K120 that is ubiquitinated by E3 ligase Bre1 (RNF20/40)[41,42], is relatively close to the H2A K15 site. The distance from this lysine to the UbcH5c catalytic center in our structure is however 15.7 Å, thus explaining the preferential ubiquitination of the H2A N-terminus by RNF168. This specificity is well maintained over the best 20 solutions, with no solutions that place H2B K120 within 9 Å. The C-terminal H2B residue, Hs. H2B K125, is similarly placed at 14.8 Å in the best solution with no solutions within 6 Å.

Analysis of the E2–nucleosome interface in the 20 best scoring solutions shows that there are few intermolecular interactions between the E2 and the nucleosome, primarily via a hydrogen-bond between H2A K15 and D117, the so-called gateway residue[43] (Figure 2-3F). In addition, most solutions show an interaction to H2B K120 and in some solutions the E2 interacts with the nucleosomal DNA through K128 or R125. An NMR titration of UbcH5c to H2A/H2B dimers in presence of sub-stoichiometric amounts of RNF168 did not reveal a clear binding interface for the E2 (Supplementary Figure 2-7). E2 binding may be stabilized in the nucleosomal context, which is also suggested by the higher activity of RNF168/UbcH5c on nucleosomes compared to dimers[23].

Overall, our data-driven integrative structure of the E3-nucleosome complex and structural model of the E2-E3-nucleosome complex indicate that RNF168 relies on multiple critical arginine-acidic patch interactions to bind the nucleosome in a productive conformation, thereby directing the E2 towards the target lysine.

UNCOUPLING OF PRC1 AND RNF168 UBIQUITINATION

We find that the RING domain of RNF168 binds to the acidic patch on the nucleosome surface, the very same binding epitope as was found for another E3 RING domain, the heterodimeric RING1B/BMI1[12]. These two E3 have their respective target lysines, K13/15 for RNF168 and K118/119 for RING1B/BMI1, at opposite sides of the nucleosome with respect to the acidic patch. The opposite position of the target lysines correlates directly with the approximate 180° turn of the E2 between both complexes, caused by a rotation of the E3 RING domain on the nucleosome surface (Figure 2-4A,B). The opposite N-to-C terminal direction of the nucleosome binding motifs of $RNF168^{RING}$ and RING1B is emphasized in Figure 2-4B. The different binding modes of RNF168 and RING1B/BMI1 on the nucleosome suggest the possibility to selectively manipulate their interactions and subsequent ubiquitination activity. Since the binding site of the two E3 proteins overlap extensively in the anchor

region, we hypothesized that mutation of histone residues in the periphery of the RNF168 interface that are distant from the RING1B interface could be effective in selective silencing of RNF168-mediated ubiquitination. In addition, the mutagenesis experiments suggest that interference with the electrostatic environment of RNF168 could be effective, as productive RNF168 binding proved highly sensitive to mutations of arginines in and around the basic helix. We thus chose H2B E110 (E113 in humans) as a feasible mutation site as it is in the periphery of the basic-helix interaction site, in the proximity of R57, and neither interacts with RING1B nor is required for the structural integrity of the H2A/H2B-dimer (Figure 2-4B). Indeed, when wild-type nucleosome or nucleosomes reconstituted from a H2B E110A mutant are interrogated for both types of ubiquitination, H2A ubiquitination mediated by RING1B is unaffected while RNF168-dependent ubiquitination is very strongly decreased (Figure 2-4C). These results demonstrate that it is possible to selectively interfere in ubiquitination pathways by mutations far from the target lysine.

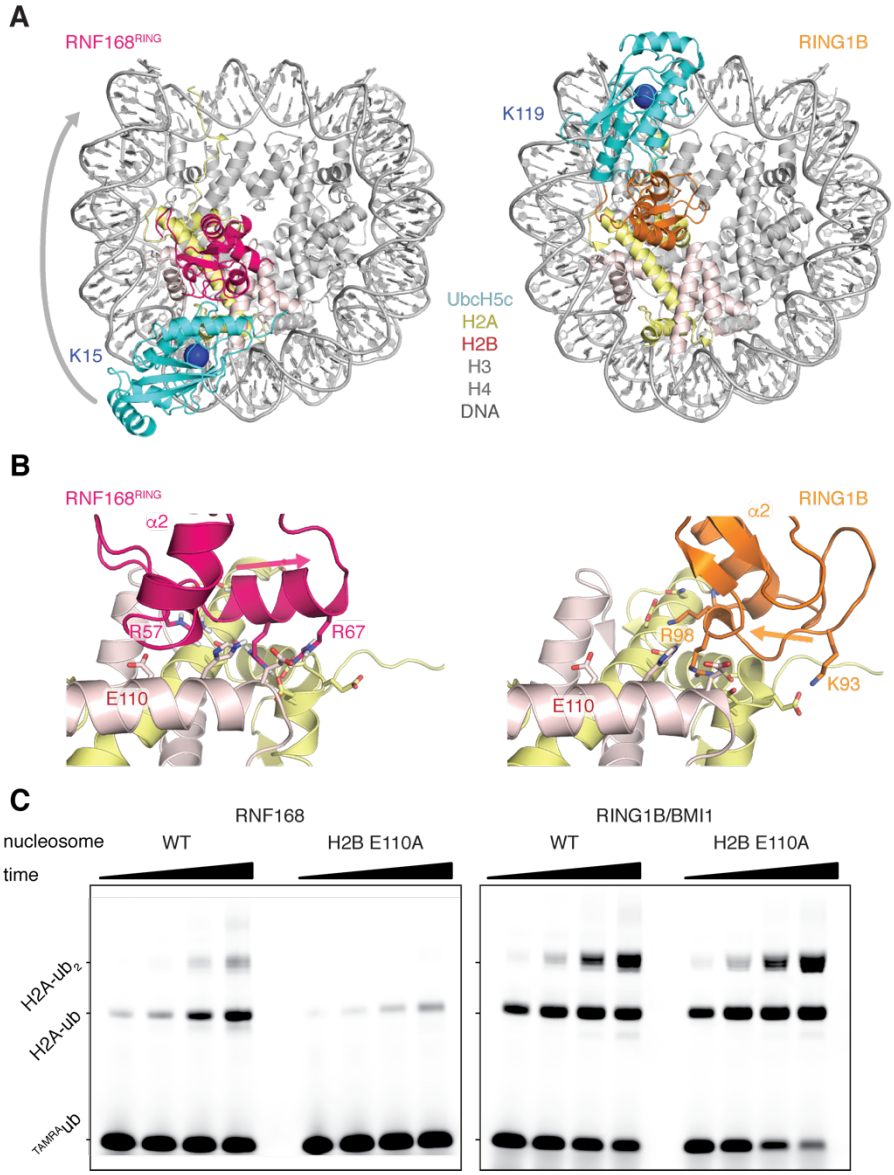


Figure 2-4. Uncoupling of PRC1 and RNF168 ubiquitination.

(A) Top view on the nucleosome highlighting the 180° rotated position of UbcH5c in the RNF168 (left) compared to RING1B nucleosome complex (right). Color coding indicated in the figure, position of the target lysines is indicated with the blue sphere. (B) Side-by-side comparison of the RING domain positions on the acidic patch, RNF168 (right) and RING1B (left). Secondary structure elements are labeled to highlight the approximate 180° turn of the RING domain, also indicated by the arrow. The H2B E110 (E113 in humans) side chain is close to the E3 RING surface for RNF168 but far from the surface of RING1B. Selected side chains are shown as sticks. (C) Nucleosome ubiquitination by either RNF168^{RING} or RING1B/BMI1 on wild-type (WT) or H2B E110A nucleosomes. Fluorescent TAMRA-

labeled ubiquitin was used to detect ubiquitinated species. Samples were resolved using SDS-PAGE. For all samples time points were taken at 2, 6, 18 and 40 min of incubation of the reaction mixture. Molecular size of mono- and di-ubiquitinated H2A is indicated.

DISCUSSION

RNF168 is an important driver of the DNA damage response, an activity that depends on its RING domain-mediated ubiquitination of H2A K13/K15. Here, we used a combination of methyl-based and amide backbone-based NMR to show that RNF168^{RING} binds the H2A/H2B acidic patch on the nucleosomal surface. Based on NMR, XL-MS and mutagenesis data, we derived a data-driven structural model of the E3-substrate complex formed by RNF168^{RING} and the nucleosome detailing the interaction between the Arg-rich helix of RNF168 and the acidic patch. Recent studies have shown that one of the UDM motifs of RNF168 also bind the acidic patch to aid recognition of the H2A-K15Ub mark[26,27], making RNF168 the first example of a histone modifier in which both writing and reading modules bind the same epitope on the nucleosome, yet distant from the modification itself.

Our structure shows how the binding mode of the E3 directs and positions the E2 towards the target lysine, thereby explaining the observed specificity of H2A ubiquitination. The observed crosslink between RNF168 and the histone surface has been crucial in defining the E3 binding mode. NMR and mutagenesis data identify binding surfaces only and thus do not directly encode in which orientation these surfaces interact. The observed crosslink established unambiguously the orientation of RNF168 on the nucleosome surface, thus resulting in a single binding mode. Importantly, the structure is validated by additional mutagenesis, and provided a structural guide to design a nucleosome mutant that selectively interferes with K13/15 but not K118/119 ubiquitination.

We find that the RNF168^{RING} basic helix is anchored by three arginine residues, R57, R63 and R67, to the acidic patch, while previous work had failed to identify a direct interaction to the acidic patch[23]. The histone mutant used in this study included a mutation for acidic patch residue E90, which is pointing away from the interface in our model, but not for D89 and E91, which are bound to both R63 and R67, and thus likely failed to abolish the interaction.

The same work by Mattiroli *et al.* established that the nucleosome promotes the rate of ubiquitination discharge[23], thus uncovering a catalytic role of the nucleosome substrate.

To shed light on the structural underpinnings of this striking observation, we compared our structural model of the E2-E3-nucleosome complex with recent E3-E2~Ub structures. These structures have highlighted the importance of Ub–E2 and Ub–E3 interactions in stabilizing the 'closed', active state of the complex. In this state the Ub C-terminal tail is bound by the E2 and thought to be strained, thus activating the thioester bond between C-terminus of Ub and the catalytic cysteine of the E2[4-6]. Alternatively, the complex may be in an open state, in which the Ub is dynamically sampling a large conformational space. Superposition of such open E3-E2~Ub structures onto our model of the E2-E3-nucleosome complex shows that these states are incompatible with nucleosome binding due to severe clashes of Ub with the nucleosome (Figure 2-5A). Superposition of E3-E2~Ub complexes in the closed state with our model shows that our structure is compatible with such closed states and the formation of E2-Ub contacts (Figure 2-5B). Notably, the orientation of RNF168^{RING} on the nucleosome surface is fully compatible with the formation of E3-Ub contacts that are observed in the structure of related RING domains bound to E2~Ub (Supplementary Figure 2-6). We thus suggest that the nucleosome itself, by steric occlusion, constrains the conformational space of Ub in the E3-E2~Ub complex and thus promotes the formation of closed complexes.

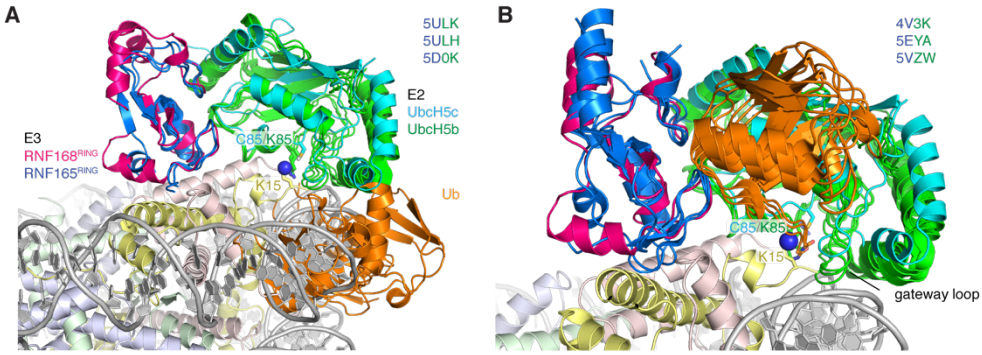


Figure 2-5. The nucleosome promotes active, closed E3-E2~Ub conformations.

(A) Superposition of E3-E2~Ub structures in the open state with E3 RNF165 and the E2-E3-nucleosome model for RNF168 indicates that nucleosome surface clashes with such open states. (B) Superposition of RING E3-E2~Ub structures and the E2-E3-nucleosome model for RNF168 shows that the model is compatible with the closed state of the E3-E2~Ub complex. There are no direct interactions between the nucleosome surface and Ub. Target lysine, here labeled as K15, catalytic center, and gateway loop are indicated. Color coding indicated in the figure.

Compared to the PRC1 machinery responsible for H2A K119 ubiquitination, we find that the E3 RNF168^{RING} domain is rotated ~180° on the nucleosome surface, explaining the difference in target specificity (see Figure 2-4). This raises the question as to the molecular origin of this this difference in binding mode. Simple considerations

underline the evident need for a specific binding mode for each E3. First, the helical structure of this element in RNF168^{RING} would cause steric clashes with the nucleosome surface when adopting the exact same binding mode as RING1B (Figure 2-6A). Second, the basic surfaces of two RING domains are oriented differently with respect to the E2 binding site and have different shapes and charge distributions (Figure 2-6B). In particular, the area around the linchpin Arg, R91 in RING1B and R55 in RNF168, is much less electropositive in RING1B than in RNF168 due to the presence of two glutamic acid residues at the surface. At the opposite side of the molecule, RNF168 is less electropositive, mainly due to the presence of E45. The overall result is that the basic surface of RNF168 is rod-shaped while that of RNF168 is rather square shaped.

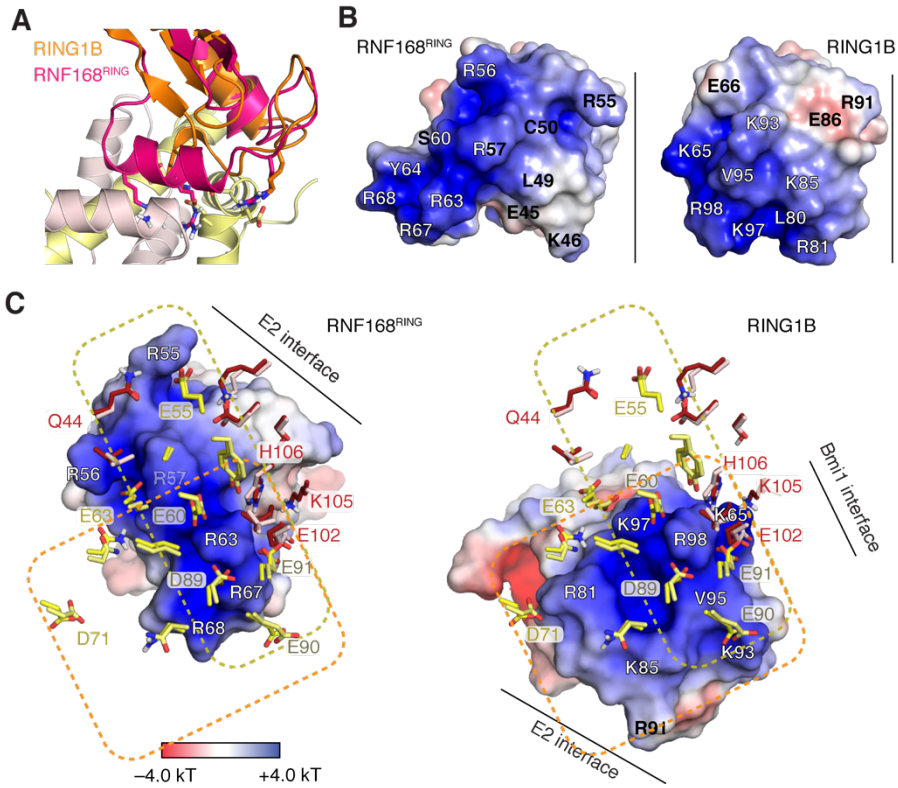


Figure 2-6. Comparison of acidic patch binding mode of RNF168 and RING1B.

(A) Superposition of RNF168^{RING} on the RING1B structure results in severe clashes of the Arg-rich helix with the H2B αC helix. (B) Molecular surface of the bound states of RNF168 and RING1B RING domains with color coded electrostatic potential, calculated for the solvent accessible surface. Both RING domains are shown looking down on their nucleosome binding surface in an aligned orientation. Location of the E2 interface is indicated with a solid line. (C) Comparison of the acidic patch binding modes of RNF168^{RING} and RING1B^{RING}, both shown as molecular surface

with color coded electrostatic potential, calculated for the solvent accessible surface, and aligned on the nucleosome. The yellow and orange dashed boxes delineate the main interaction sites of RNF168 and RING1B. Selected residues are labeled.

Examination of the two complexes shows how these differences correlate with the observed binding mode (Figure 2-6C). The positive extension of RNF168 interacts with region around H2A E55, thus filling the negatively charged groove formed by the H2A α2 and H2B αC helix completely. In RING1B the area around H2A E55 is not contacted, instead the unique extension of RING1B, R81, one of the drivers of nucleosome binding[12], interacts outside the α2-αC cleft with H2A D72. In RNF168, the presence of E45 around this position favors interaction with H2B K105 on the opposite side of the cleft. It is thus tempting to speculate that this combination of changes, i.e. a linear vs. square shaped extension of a 'core' basic surface around the arginine anchor, is responsible for the change in binding mode. Experiments in which the nucleosome-binding elements of the two E3 RING domain were swapped resulted in inactive protein. A further complicating factor is that RING1B binds as a heterodimer with BMI1 to the nucleosome surface, with BMI1 predominantly contacting H3/H4[12]. Nevertheless, the differences in RING-nucleosome interaction between RING1B and RNF168 highlight the importance of residues surrounding the arginine anchor in determining the exact binding mode.

The RNF168^{RING}–nucleosome structure was used to design an H2B E110A nucleosome mutant to selectively suppress RNF168 activity without interfering with RING1B mediated ubiquitination. The successful design further validates the structural model. Since RNF168 still shows low residual activity on nucleosomes carrying the H2B E110A mutant, a charge reversal at this position may be needed to completely and selectively silence RNF168. This H2B E110 mutant (E113 in humans) may be an excellent tool to investigate both RNF168 and PRC1 pathways without interfering with the target lysine residue.

Overall, our results contribute to a better understanding of the molecular mechanism of nucleosome ubiquitination, underscoring the crucial and diverse roles that RING type E3 ligase and their substrates play in ubiquitination. Our data also highlight how chromatin factors can exploit the nucleosome acidic patch in very different ways to control chromatin biology and cell fate.

MATERIALS AND METHODS

PROTEIN EXPRESSION AND PURIFICATION

For NMR studies, *Drosophila melanogaster* histones were expressed in BL21 Rosetta2 (DE3) cells (Novagen) and purified as described in ref.[44]. For the ubiquitination, gel-shift and crosslinking studies *Xenopus leavis* histones were used. Histones used for NMR studies were produced in M9 minimal medium containing desired isotopes, histones used for ubiquitination assays were expressed in LB medium. Methyl-labeling of Ile-δ1-[$^{13}CH_3$] and Val/Leu-[$^{13}CH_3$, $^{12}CD_3$] (ILV-labeling) followed the procedure of Tugarinov[35].

The RNF168-RING domain construct (residues 1-113) was cloned into petNKI-His-SUMO2-kan vector and purified using a His-SUMO tag. The protein was expressed in *E.coli*. Cells were grown to an OD of 0.6 and induced with 200 μM IPTG. The protein was expressed overnight at 16 °C. The cells were harvested in lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 μM $ZnCl_2$, 1 mM TCEP, 2 mM imidazole) including complete EDTA-free protease inhibitor. Lysate was cleared by centrifuging at 21000 g and the supernatant was loaded on chelating sepharose beads (GE healthcare) charged with Ni^{2+}. The beads were washed with lysis buffer containing 20 mM imidazole and the protein was eluted using 350 mM imidazole. The His-SUMO tag was cleaved overnight using His-tagged SENP2 protease while dialyzing against 50 mM HEPES pH 7.5, 150 mM NaCl, 1 μM $ZnCl_2$, 1 mM TCEP. Uncleaved protein and SENP2 protease were removed using chelating sepharose beads charged with Ni^{2+}. The sample was then diluted to 50 mM NaCl and loaded onto a Heparin column (GE healthcare). The protein was eluted using a salt gradient ranging from 50 mM to 1 M NaCl over 12 column volumes. Fractions containing RNF168 were combined and gel filtered on a Superdex 75 16/60 column (GE healthcare) in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 μM $ZnCl_2$, 1 mM TCEP. Fractions containing RNF168 were combined, snap frozen in liquid nitrogen and stored at – 80 °C.

HISTONE REFOLDING AND NUCLEOSOME RECONSTITUTION

H2A/H2B dimers or histone octamers were refolded from equimolar mixes of denatured purified histones by dialysis to 2M NaCl and subsequent purification using size-exclusion chromatography[44]. The 601-DNA 'Widom' template was produced from a pUC19 plasmid containing 12 copies of the 601 fragment amplified in *E. coli*

DH5α and extracted by alkaline lysis followed by isopropanol and ethanol precipitations as described in Dyer *et al.* [45]. The pellet was dissolved in TE buffer (20 mM Tris-HCl pH 7.5, 5mM EDTA, 100mM NaCl) and purified by anion exchange chromatography. The purified plasmid was restricted with ScaI (ThermoFisher) and purified by anion exchange chromatography. Nucleosomes were reconstituted from the purified 167bp 601 DNA and refolded and purified histone octamers according to the salt-gradient deposition method described by Dyer *et al.*[45]

FLUORESCENT LABELING OF UBIQUITIN

Ubiquitin carrying a cysteine as residue 2 was cloned into the petNKI-His-SUMO2-kan vector and expressed in *E.coli*. Cells were grown in LB until an OD of 0.8 was reached and then induced with 200 μM IPTG. The protein was expressed for 4 hours at 37 °C and cells were harvested in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP and 5 mM imidazol). Cells were lysed by sonication and the lysate was cleared by centrifuging at 21000 g. The supernatant was loaded on chelating sepharose beads charged with Ni^{2+} and washed with lysis buffer containing 20 mM imidazol. The protein was eluted in lysis buffer containing 350 mM imidazol. The His-SUMO tag was cleaved over night with SENP2 protease while dialyzing against 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP. Protease, uncleaved protein and His-SUMO tag was removed using chelating sepharose charged with Ni^{2+}. Perchloric acid was added dropwise to the sample while stirring on ice until a final concentration of 2 % v/v. The sample was centrifuged at 21000 g and the supernatant was collected and dialyzed against 50 mM ammonium acetate pH 4.5 overnight. The sample was loaded on a SP HP column (GE healthcare) and eluted using a linear salt gradient ranging from 0 to 500 mM NaCl. The sample was then gel-filtered in 50 mM Tris pH 8, 100 mM NaCl, 5 mM DTT and stored at – 80 °C. Before labeling the required amount of ubiquitin was dialyzed against at least 3 changes (one step overnight) of 2 L labeling buffer (50 mM Tris, pH 7.5, 150 mM NaCl) to remove the DTT. The ubiquitin was then labeled by adding a 5-fold molar excess of TAMRA-maleimide (Setareh Biotech). Samples were incubated for 2 hours at room temperature, followed by 14 hours at 4 °C. The labeling reaction was quenched by adding 5 mM DTT and excess label was removed through size exclusion chromatography using a Superdex S75 column (GE healthcare) in 50 mM Tris pH 7.5, 150 mM NaCl. The labeled protein was stored at – 80 °C

UBIQUITINATION ASSAYS

All ubiquitination assays were done in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, and 5 mM DTT. Nucleosomes were ubiquitinated using 1 μM hUBA1, 0.5 μM UBE2D3, 1 μM RNF168 (wild type or the respective mutant), 7 μM recombinant nucleosome core particles containing *Xenopus laevis* histones, 15 μM TAMRA labeled ubiquitin at 30 °C. The reaction was started by adding 3 mM ATP and stopped by mixing with SDS-loading dye at the desired time point. Ubiquitinated histones were resolved using SDS-PAGE. Bands were visualized using fluorescence of TAMRA-labeled ubiquitin.

ELECTROPHORETIC MOBILITY SHIFT ASSAYS

Recombinant nucleosome core particles containing *Xenopus laevis* histones (250 nM) were mixed with varying concentrations of $RNF168^{1-113}$ (wild-type or the respective mutants) in a buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM DTT. Samples were incubated at room temperature for 20 minutes and then separated using a 0.2 x TBE, 6 % Acrylamide native gel run at 120 V at 4 °C for 75 minutes. The gel was pre-run prior to the experiment for 90 minutes at 4 °C at 120 V. The gel was then stained using SYBR-safe stain (ThermoFisher) and Coomassie Brilliant Blue.

NMR SPECTROSCOPY

All NMR experiments were carried out at 293K on a Bruker Avance III HD spectrometer operating at 850 MHz ^1H Larmor frequency equipped with a TCI cryo-probe, unless noted otherwise. Processing was done using the NMRPipe package[46] or Bruker's TopSpin. Spectra were analyzed using Sparky (Goddard and Kneller, UCSF).

Samples for assignment of H2A contained ~300 μM [U-^2H/^{13}C/^{15}N]-H2A-H2B or [U-^{13}C/^{15}N]-H2A-H2B in 95/5% $H2O/D_2O$ in NMR buffer (20 mM NaPi, pH 6.5, 50 mM NaCl, 5% D2O, 0.02% NaN_3, complete EDTA-free protease inhibitor cocktail (Roche)). Backbone assignments were based on TROSY-based HNCACB, HN(CO)CACB, HNCA, HN(CO)CA, HNCB, HN(CO)CB, HNCO, and HN(CA)CO spectra, recorded at 308 K. Overall assignment completeness was 97.1% for all backbone atoms. Assignments are deposited in the BMRB databank under accession code 27547.

NMR titration experiments of H2A-H2B histone dimer with $RNF168^{RING}$ were performed using ~110 μM [U-^{1}H,^{15}N]-H2A-H2B for the H2A-observed titration and~200 μM [U-^{2}H]-H2A-[U-^{2}H,^{15}N]-H2B for the H2B-observed titration and ~460 μM unlabeled $RNF168^{RING}$ stock. Both proteins were extensively dialyzed to 90/10% H_2O/D_2O with 25 mM Tris pH 8.0, 100 mM NaCl, 10 μM $ZnCl_2$ and 3 mM DTT. Nucleosome interaction studies were done at 298K, using ~25 μM mononucleosomes reconstituted with ILV-H2A, ILV-H2B / ^{2}D-H3 & ^{2}D-H4 labelling scheme. The buffer conditions were 30 mM Tris pH* 7.3 in 100% D_2O with 100 mM NaCl, 1 mM TCEP and 1 μM $ZnCl_2$. Methyl group assignments were transferred from the original assignment condition[29] to the interaction buffer using a buffer and temperature titration.

The titration and CPMG relaxation dispersion study of H2B-labeled H2A/H2B dimers were carried on a Bruker Avance III HD spectrometer operating at 950 MHz ^{1}H Larmor frequency equipped with a TCI cryo-probe. Relaxation dispersion experiments were recorded on the ^{15}N TROSY coherence[47] using a constant-time CPMG relaxation delay of 40(20) ms and 22(13) ν_{CPMG} values, including three replicates for error estimation, in the range of 25(50)-1500 Hz for the unbound(bound) dimer. Peak volumes were determined using Fuda[48] and best-fits were obtained using Catia[49].

CROSSLINKING MASS-SPECTROMETRY

Nucleosome core particles (NCP), reconstituted from *Xenopus laevis* histones and 601-DNA, and $RNF168^{1-113}$ were crosslinked at final concentrations of 15 μM RNF168 and 5 μM NCP in 10 μL volume using a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl. The samples were incubated for 10 minutes at room temperature. A 500-fold molar excess of BS3 (with respect to NCP concentration) was added and crosslinking proceeded for 20 minutes at room temperature. The reaction was stopped by addition of Tris pH 8 to a final concentration of 125 mM. The samples were then diluted to a volume of 100 μL using a buffer containing 50 mM Tris pH 8 and 150 mM NaCl. TCEP was added to a final concentration of 10 mM and the samples were incubated at 60 °C for 30 minutes. Chloroacetamide was added to a final concentration of 40 mM and the samples were incubated in the dark for 20 minutes. The samples were the digested using trypsin at 37 °C overnight. The digest was stopped by adding formic acid to a final concentration of 1 %. The digested samples were purified using 100 μl ZipTip pipette tips (Merck) according to the manufacturer's instructions. Samples were then concentrated to ~ 5 μL using a SpeedVac vacuum centrifuge and diluted with 20 μL Buffer A (0.1% formic acid).

The peptides were analyzed by online nano-high-pressure liquid chromatography (UPLC) electrospray ionization-tandem mass spectrometry (MS/MS) on an Q Exactive Plus Instrument connected to an Ultimate 3000 UPLC (Thermo-Fisher Scientific). Peptides reconstituted in 0.1% formic acid were loaded onto a trap column (Acclaim C18 PepMap 100 nano Trap, 2 cmx 100 µm I.D., 5-µm particle size and 300-Å pore size; Thermo-Fisher Scientific) at 15 µL/min for 3 min before switching the precolumn in line with the analytical column (Acclaim C18 PepMap RSLC nanocolumn, 75 µm ID × 50 cm, 3-µm particle size, 100-Å pore size; Thermo-Fisher Scientific). The separation of peptides was performed at 250 nL/min using a non-linear ACN gradient of buffer A (0.1% formic acid) and buffer B (0.1% formic acid, 80% ACN), starting at 2.5% buffer B to 42.5% over 95 min. Data were collected in positive mode using a Data Dependent Acquisition m/z of 375–2000 as the scan range, and higher-energy collisional dissociation (HCD) for MS/MS of the 12 most intense ions with z 2-5. Other instrument parameters were: MS1 scan at 70 000 resolution, MS maximum injection time 118 ms, AGC target 3E6, ion intensity threshold of 4.2e4 and dynamic exclusion set to 15 s. MS/MS resolution of 35000 at Orbitrap with the maximum injection time of 118 ms, AGC of 5e5 and HCD with Collision Energy=27%.

For the data analysis, Thermo raw files were analyzed using the pLink 2.3.4 search engine[50], searching against the sequences of $RNF168^{1-113}$, Histone H2A, H3, H4 and H2B in FASTA format. The default settings for searches were used. N-terminal acetylation and methionine oxidation were used as variable modifications and carbamidomethyl on cysteines as a fixed modification. False discovery rates of 1% for peptide spectrum match level were applied by searching a reverse database. Reproducible crosslinks were identified from three replicate experiments and manually verified. Crosslinks within the histone core were analyzed for compatibility with nucleosome structure by calculating the solvent accessible surface distance using Jwalk[39].

STRUCTURAL MODELLING

The structural model of the $RNF168^{RING}$-nucleosome complex was determined using the experimental NMR, XL-MS and mutagenesis data in the data-driven docking software HADDOCK[37]. Briefly, the $RNF168^{RING}$ domain (PDB-id 4GB0[51]) was docked to the nucleosome surface (PDB-id 2PYO[52]) using 12 ambiguous interaction restraints (AIR) corresponding to mapped binding interface and one 1 unambiguous distance restraint corresponding to the observed crosslink. Active residues on the nucleosome were defined based on significant intensity and/or chemical shift

perturbations according to Figure 2-2C. Active residues for RNF168^{RING} were defined based on the mutagenesis experiments (R57 and R63). Passive residues were defined automatically by HADDOCK. Because of the small number of restraints for RNF168 the default random exclusion of 50% of the AIRs was switched off. The XL-MS restraint was defined as 28 Å Cα-Cα distance between RNF168 K46 and H2B K105. At each stage of the docking process all solutions were filtered based on their SASD for the crosslinked residues. Solution with SASD > 35 Å were rejected. Rejected solutions had on average 38.2 Å SASD, while accepted structures had SASD of 11.1 Å on average. Of the 200 water-refined solutions, 99.5% were clustered based on fraction of common contacts in two conformations (see Supplementary Figure 2-4 for the statistics).

To model the E2-E3-nucleosome complex, we first extracted 12 E2-E3 structures from the PDB database with either high structural similarity (PDB-id 4S30 and 4R8P) or high sequence similarity of the E3 to RNF168^{RING} (PDB-id 2YHO, 3EB6, 3HCT , 3RPG, 4A49, 4AP4, 4QPL, 4TKP, 5FER, 5VNZ) using either the PDBeXplore or PDBeFOLD web-service. These twelve structures were superimposed and analyzed for conserved E2-E3 interactions (see Supplementary Figure 2-5). The 13 conserved interactions were subsequently used as AIRs, specifying only the pairwise interactions at the residue-level, to dock the E2 UbcH5c (PDB-id 1X23) to RNF168. In addition, the linch-pin hydrogen bond between the RNF168 R55 side chain and E2-Q92 backbone was enforced. The E2 was docked to the ensemble of 10 best scoring solutions for the RNF168^{RING}-nucleosome complex, in which *Dm.* H2A N18 and *Dm.* H2B S121 were mutated in silico to match the Hs. H2A and H2B sequence. All solutions cluster in a single conformation of the E2-E3-nucleosome ternary complex. In all cases, the 20 best scoring solutions were analyzed for conserved intermolecular interactions.

MOLECULAR GRAPHICS

All molecular graphics were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC). Electrostatic surfaces were calculated using the adaptive Poisson-Boltzman solver[53] and the AMBER force field.

REFERENCES

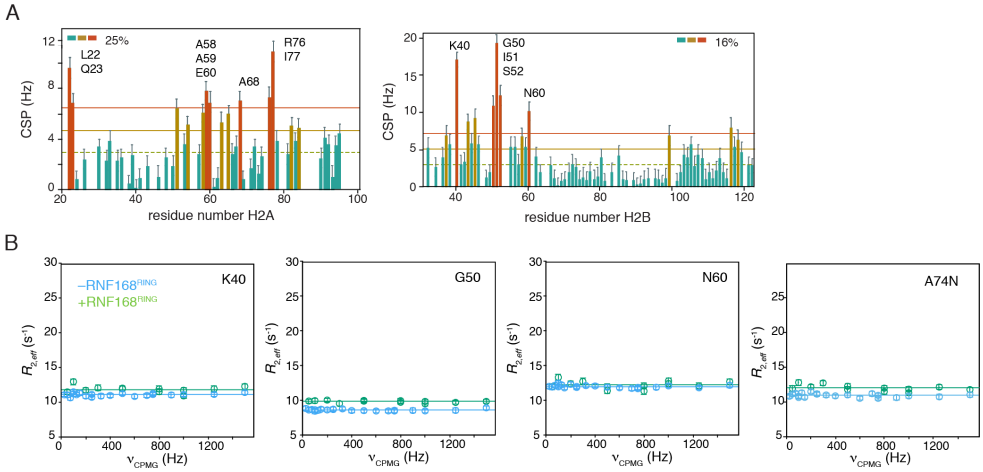
1. Deshaies, R. J. & Joazeiro, C. A. P. RING Domain E3 Ubiquitin Ligases. *Annu Rev Biochem* **78,** 399–434 (2009).
2. Mattiroli, F. & Sixma, T. K. Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nat Struct Mol Biol* **21,** 308–316 (2014).
3. Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* **21,** 301–307 (2014).
4. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat Struct Mol Biol* **19,** 876–883 (2012).
5. Plechanovová, A., Jaffray, E., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489,** 115–120 (2012).
6. Pruneda, J. N. *et al.* Structure of an E3:E2~Ub Complex Reveals an Allosteric Mechanism Shared among RING/U-box Ligases. *Mol. Cell* **47,** 933–942 (2012).
7. Bloom, J., Amador, V., Bartolini, F., DeMartino, G. & Pagano, M. Proteasome-Mediated Degradation of p21 via N-Terminal Ubiquitinylation. *Cell* **115,** 71–82 (2003).
8. Fischer, E. S. *et al.* The Molecular Basis of CRL4DDB2/CSA Ubiquitin Ligase Architecture, Targeting, and Activation. *Cell* **147,** 1024–1039 (2011).
9. Mattiroli, F. *et al.* RNF168 Ubiquitinates K13-15 on H2A/H2AX to Drive DNA Damage Signaling. *Cell* **150,** 1182–1195 (2012).
10. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431,** 873–878 (2004).
11. Kalb, R., Mallery, D. L., Larkin, C., Huang, J. T. & Hiom, K. BRCA1 Is a Histone-H2A-Specific Ubiquitin Ligase. *Cell Rep* **8,** 999–1005 (2014).
12. McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514,** 591–596 (2014).
13. Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell* **49,** 808–824 (2013).
14. Uckelmann, M. & Sixma, T. K. Histone ubiquitination in the DNA damage response. *DNA Repair (Amst)* **56,** 92–101 (2017).
15. Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M. & Greenberg, R. A. An ATM-Dependent Transcriptional Silencing Program is Transmitted Through Chromatin in Cis to DNA Double Strand Breaks. *Cell* **141,** 970–981 (2010).
16. Kakarougkas, A. *et al.* Requirement for PBAF in Transcriptional Repression and Repair at DNA Breaks in Actively Transcribed Regions of Chromatin. *Mol. Cell* **55,** 723–732 (2014).

17. Wilson, M. D. *et al.* The structural basis of modified nucleosome recognition by 53BP1. *Nature* **536,** 100–103 (2016).
18. Fradet-Turcotte, A. *et al.* 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* **499,** 50–54 (2013).
19. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell* **47,** 497–510 (2012).
20. Zimmermann, M. & de Lange, T. 53BP1: pro choice in DNA repair. *Trends Cell Biol* **24,** 108–117 (2014).
21. Densham, R. M. *et al.* Human BRCA1–BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. *Nat Struct Mol Biol* **23,** 647–655 (2016).
22. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* **26,** 52–64 (2016).
23. Mattiroli, F., Uckelmann, M., Sahtoe, D. D., van Dijk, W. J. & Sixma, T. K. The nucleosome acidic patch plays a critical role in RNF168-dependent ubiquitination of histone H2A. *Nat Commun* **5,** 3291 (2014).
24. Leung, J. W. *et al.* Nucleosome Acidic Patch Promotes RNF168- and RING1B/BMI1-Dependent H2AX and H2A Ubiquitination and DNA Damage Signaling. *Plos Genetics* **10,** e1004178 (2014).
25. Thorslund, T. *et al.* Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature* **527,** 389–393 (2015).
26. Kitevski-LeBlanc, J. *et al.* The RNF168 paralog RNF169 defines a new class of ubiquitylated histone reader involved in the response to DNA damage. *Elife* **6,** 215 (2017).
27. Hu, Q., Botuyan, M. V., Cui, G., Zhao, D. & Mer, G. Mechanisms of Ubiquitin-Nucleosome Recognition and Regulation of 53BP1 Chromatin Recruitment by RNF168/169 and RAD18. *Mol. Cell* **66,** 473–487.e9 (2017).
28. Takahashi, T. S. *et al.* Structural insights into two distinct binding modules for Lys63-linked polyubiquitin chains in RNF168. *Nat Commun* **9,** 170 (2018).
29. Kato, H. *et al.* Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci USA* **108,** 12283–12288 (2011).
30. van Nuland, R. *et al.* Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics & Chromatin* **6,** 12 (2013).
31. Musselman, C. A. *et al.* Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. *Nat Commun* **4,** 2969 (2013).
32. Zhou, B.-R. *et al.* Structural insights into the histone H1-nucleosome complex. *Proc Natl Acad Sci USA* **110,** 19390–19395 (2013).
33. Zhou, B.-R. *et al.* Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol. Cell* **59,** 628–638 (2015).
34. Xiang, S. *et al.* Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **57,** 4571–4575 (2018).

35. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **125,** (2003).
36. McGinty, R. K. & Tan, S. Nucleosome Structure and Function. *Chemical Reviews* **115,** 2255–2273 (2015).
37. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428,** 720–725 (2016).
38. Wassenaar, T. A. *et al.* WeNMR: Structural Biology on the Grid. *Journal of Grid Computing* **10,** 743–767 (2012).
39. Matthew Allen Bullock, J., Schwab, J., Thalassinos, K. & Topf, M. The Importance of Non-accessible Crosslinks and Solvent Accessible Surface Distance in Modeling Proteins with Restraints From Crosslinking Mass Spectrometry. *Mol Cell Proteomics* **15,** 2491–2500 (2016).
40. McGinty, R. K. & Tan, S. Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol* **37,** 54–61 (2016).
41. Wood, A. *et al.* Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol. Cell* **11,** 267–274 (2003).
42. Gallego, L. D. *et al.* Structural mechanism for the recognition and ubiquitination of a single nucleosome residue by Rad6–Bre1. *Proc Natl Acad Sci USA* **113,** 10553–10558 (2016).
43. Stewart, M. D., Ritterhoff, T., Klevit, R. E. & Brzovic, P. S. E2 enzymes: more than just middle men. *Cell Res* **26,** 423–440 (2016).
44. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Meth Enzymol* **304,** 3–19 (1999).
45. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth Enzymol* **375,** 23–44 (2004).
46. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. **6,** 277–293 (1995).
47. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E. & Kay, L. E. Measurement of bond vector orientations in invisible excited states of proteins. *Proc Natl Acad Sci USA* **104,** 18473–18477 (2007).
48. Hansen, D. F. www.biochem.ucl.ac.uk/hansen/fuda/. Available at: (Accessed: 9 January 2018)
49. Hansen, D. F. http://www.ucl.ac.uk/hansen-lab/catia/. Available at: (Accessed: 6 December 2018)
50. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat Methods* **9,** 904–906 (2012).
51. Zhang, X. *et al.* Structural basis for role of ring finger protein RNF168 RING domain. *Cell Cycle* **12,** 312–321 (2013).
52. Clapier, C. R. *et al.* Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **71,** 1–7 (2008).

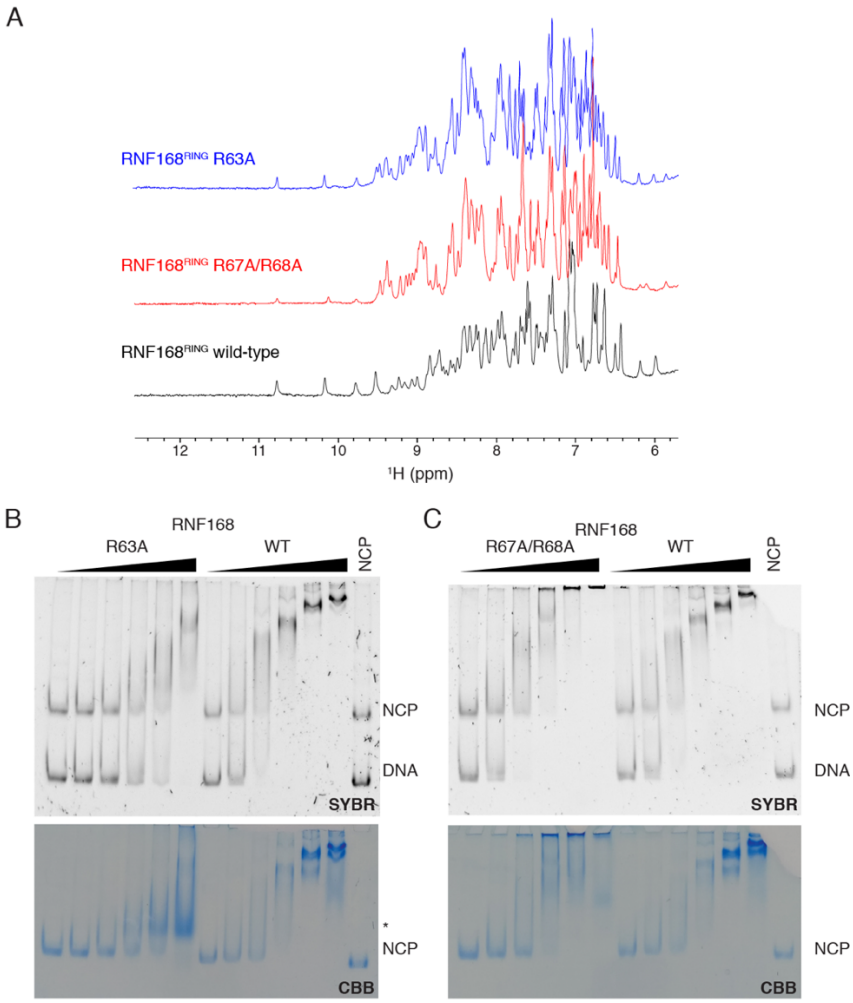
53. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98,** 10037–10041 (2001).

SUPPLEMENTARY INFORMATION



Supplementary Figure 2-1. Non-specific binding of RNF168^{RING} to the DNA binding regions of the H2A/H2B dimer.

(A) Chemical shift perturbations (CSPs) for H2A (left) and H2B (right) within H2A/H2B dimers upon addition of sub-stoichiometric amounts of RNF168^{RING} (indicated in figure). Residues with CSPs higher that are one (two) standard deviation higher than the 10% trimmed mean are color coded yellow (red) and labeled. (B) Addition of RNF168^{RING} does not induce dispersion of effective ^{15}N transverse relaxation rates, $R_{2,eff}$ for H2B residues K40, G50 and N60 that are part of or close to the DNA binding regions of the H2A/H2B dimer. Residue A47 is not affected by addition of RNF168-RING and is shown as control. Exchange with the RNF168^{RING}-H2A-H2B complex causes a slight increase in $R_{2,eff}$. Color coding indicated in figure. Best-fit straight solid lines are shown to guide the eye. Error bars are s.d. based on 1/10 ppb errors in ^{1}H/^{15}N peak position in panel A and based on three replicate data points in panel B.



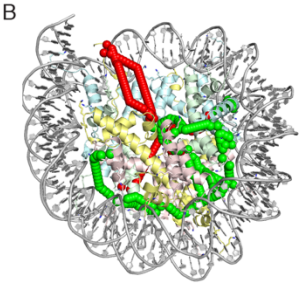
Supplementary Figure 2-2. RNF168^{RING} R63A and R67/R68A mutants are folded and have altered nucleosome binding.

(A) ^{1}H NMR spectra of wild-type, R63A, and R67/R68A mutant RNF168^{RING} showing the amide-region. The high chemical shift dispersion and the similarity of the peak pattern between mutant and wild-type indicate that the mutants retain the native fold. (B,C) Native PAGE electrophoretic mobility shift assay (EMSA) of wild-type (WT)), R63A (B), and R67/R68A mutant (C) RNF168^{RING} binding to nucleosomes; gel is stained using SYBR DNA-stain at top, and using Coomassie Brilliant Blue (CBB) at bottom. RNF168 concentrations used are 50 μM, 25 μM, 12.5 μM, 6.25 μM, 3.13 μM, and 1.56 μM. (B) The R63A mutant shows a markedly reduced in affinity compared to wild-type as judged from the disappearance of the nucleosome band (labeled NCP) upon increasing RNF168 protein concentration. In addition, a protein-DNA complex at relatively high mobility is formed (labeled with *) indicating an altered binding mode. (C) Judging from the disappearance of the nucleosome band (labeled NCP) upon increasing RNF168 protein concentration, the R67A/R68A mutant protein has comparable affinity as wild-type for nucleosome binding. The absence of a well-defined band as for the wild-type suggest decreased stability of the complex and/or an altered binding mode.

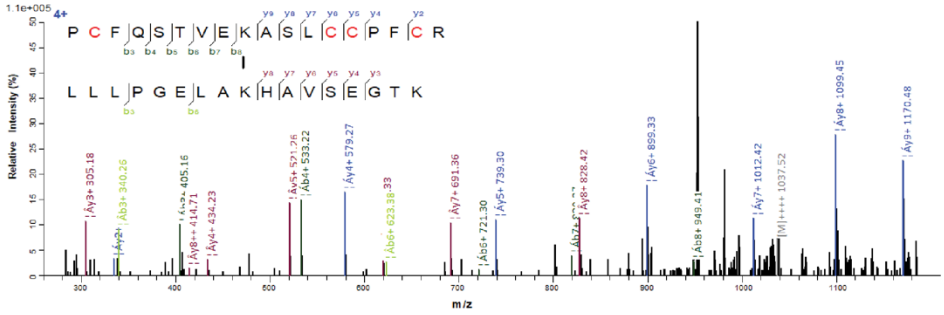
A

	# / crosslink	SASD		# / crosslink	SASD
Total reproducible crosslinked peptides	65			H2A-K36 / H2B-K82	11.5
Total unqiue reproducible crosslinks	62			H2A-K36 / H4-K77	29.0
Crosslinks involving tails	40			H2A-K36 / H4-K91	24.2
Crosslinks within/between tails	16			H2A-K95 / H2B-K105	22.0
Crosslinks within histone core	21			H2A-K95 / H4-K91	21.5
intramolecular	6			H2B-K31 / H2B-K82	14.6
	H2A-K95 / H2A-K118	40.9*		H2B-K31 / H4-K91	24.3
	H2A-K95 / H2A-K119	42.4*		H3-K56 / H2A-K74	18.6
	H2B-K31 / H2B-K43	31.1		H3-K56 / H4-K91	49.5*
	H3-K56 / H3-K64	26.9		H3-K79 / H2B-K105	17.0
	H3-K56 / H3-K79	55.8*		H3-K79 / H4-K77	12.3
	H4-K77 / H4-K79	8.6		H3-K79 / H4-K79	9.9
intermolecular	15			H4-K91 / H2B-K82	23.6
	H2A-K36 / H2A-K36	17.7	**Intermolecular crosslinks RNF168-histones**	1	
	H2A-K36 / H2B-K31	17.2		H2B-K105 / RNF168-K46	8.7 ± 1.1

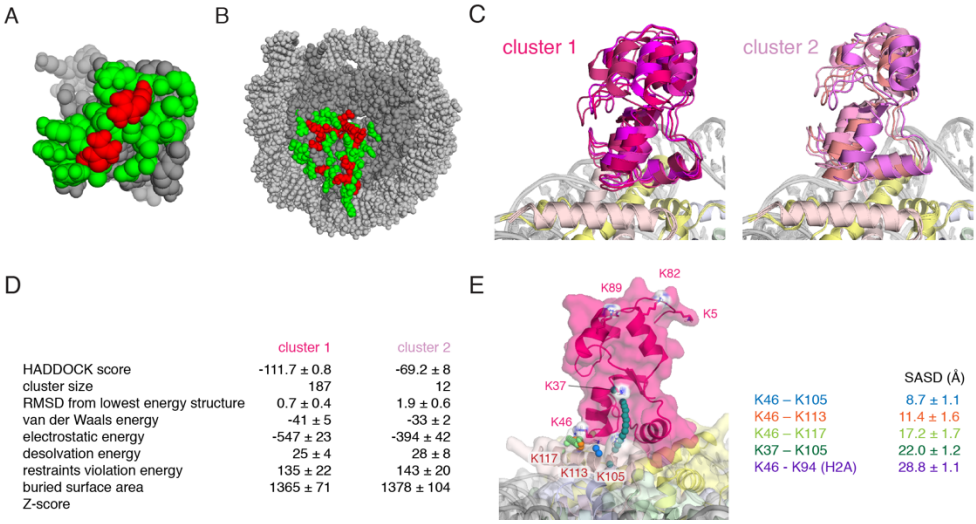
B



C



Supplementary Figure 2-3. Crosslinking mass-spectrometry results for the RNF168^{RING}–nucleosome complex. (A) Overview of all crosslinks obtained that were reproduced over three replicate experiments. The number of crosslinks obtained in various categories are indicated, as well as the identity of crosslinks that do not involve lysines in the flexible histone tails. For each listed crosslink, the solvent accessible surface distance (SASD) between Cα-atoms was calculated using Jwalk[1] based on the nucleosome crystal structure (1KX5). For the intermolecular crosslinks between RNF168^{RING} and the nucleosome, the SASD represents the average distance and standard deviation in the 10 best scoring structures of cluster 1. Out of 21 histone-histone crosslinks, 17 have SASD values that are compatible with cross-linking, while four have SASDs more than 40 Å (marked with *). For two of these, the cross-link involved lysines in the H2A C-terminal tail (K118/K119) which likely experiences dynamics and can thus transiently visit conformations that bring it close to H2A K95. The other two incompatible cross-links involve H3K56. These may be the result of transiently captured nucleosome-nucleosome contacts. (B) The solvent accessible surface distance (SASD) for the crosslinks involving the histone core are shown as spheres on the nucleosome structure. Incompatible crosslinks are shown in red, otherwise in green. (C) Mass spectrum and fragmentation pattern for the intermolecular crosslinked peptide as interpreted by the pLink software[2]. The peptide sequence is shown in the top, the top part of the sequence refers to RNF168, the bottom part to H2B.

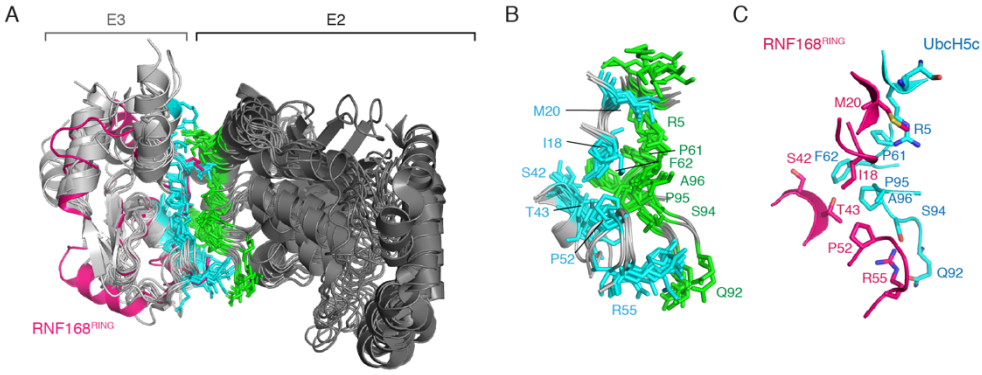


	cluster 1	cluster 2
HADDOCK score	-111.7 ± 0.8	-69.2 ± 8
cluster size	187	12
RMSD from lowest energy structure	0.7 ± 0.4	1.9 ± 0.6
van der Waals energy	-41 ± 5	-33 ± 2
electrostatic energy	-547 ± 23	-394 ± 42
desolvation energy	25 ± 4	28 ± 8
restraints violation energy	135 ± 22	143 ± 20
buried surface area	1365 ± 71	1378 ± 104
Z-score		

	SASD (Å)
K46 – K105	8.7 ± 1.1
K46 – K113	11.4 ± 1.6
K46 – K117	17.2 ± 1.7
K37 – K105	22.0 ± 1.2
K46 - K94 (H2A)	28.8 ± 1.1

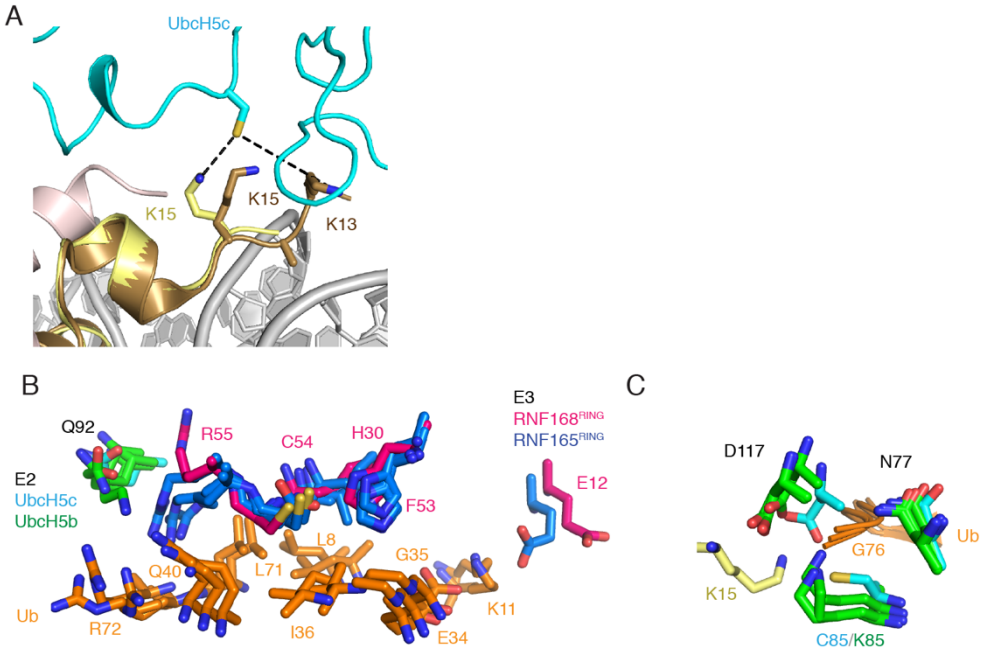
Supplementary Figure 2-4. Overview of the docking solutions for the RNF168^{RING}-nucleosome complex.

(A,B) Definition of active and passive residues as input for the HADDOCK docking protocol. Views on the mapped binding interfaces of RNF168 (A, PDB-4GB0) and the nucleosome (B, PDB 2PYO) from mutagenesis and NMR data. Active residues are shown in red, passive residues in green and all others in gray. DNA is shown in light gray. (C,D) Superposition of (C) and statistics for (D) the best four structures of each cluster obtained by NMR, XL-MS and mutagenesis data driven docking of RNF168^{RING} to the nucleosome. Cluster 2 is a slightly tilted solution compared to cluster 1 and would position the E2 active site at ~13 Å from the target lysine. Error bars are s.d. over the four best structures. (E) Overview of all expected intermolecular crosslinks between RNF168^{RING} and the nucleosome based on the best ten solutions in cluster 1. Only crosslinks with solvent accessible surface distance (SASD) < 30 Å in all solutions were considered. All crosslinks are to lysines of H2B unless noted otherwise. Of the five expected crosslinks, four involve RNF168 K46. The experimentally observed crosslink also involves K46 and has the shortest predicted SASD. The SASD of K46-K94 (H2A) partially overlaps with the SASD path of the experimentally observed crosslink and is not shown for clarity. Error bars are s.d. over the best ten structures



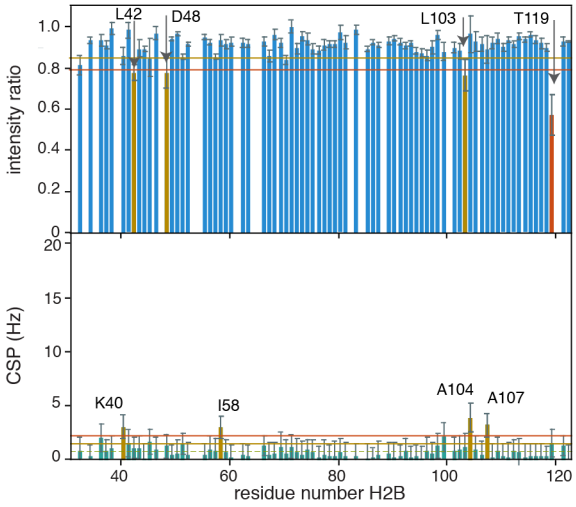
Supplementary Figure 2-5. Modelling of the E2-E3 interface.

(A) Superposition of twelve E2-E3 structures with either high structural or sequence similarity to RNF168^{RING}. The interaction surfaces are color coded in blue for the E3 and green for the E2. Structures were superimposed on the E3. (B) Zoom on the intermolecular E2-E3 interface for the UbcH5 family, with conserved interactions shown in stick representation. Contacts between the labeled residues were imposed during docking of the ternary UbcH5c-RNF168^{RING}-nucleosome complex. In addition, the 'linchpin' hydrogen bond characteristic of active E3-E2~Ub complexes[3] was enforced between RNF168 R55 and UbcH5c Q92. (C) Zoom on the intermolecular E2-E3 interface in the best scoring solution for the UbcH5c-RNF168^{RING}-nucleosome complex.



Supplementary Figure 2-6. Implications for Ub-transfer.

(A) Superposition of *Hs.* nucleosome structure (PDB-ID 2CV5) with the RNF168-UbcH5c-nucleosome model. Side chains of H2A K13 and K15 are shown as brown sticks. In our model the Sγ-Nζ distance is 3.3 Å for K15. The distance to K13 is 10.8 Å, suggesting that ubiquitination of this residues involves a conformational change in this region of the dynamic H2A tail. (B) Detailed view on the conserved E3-Ub interface, including a potential salt-bridge or hydrogen interaction between Ub-K11 and RNF168-E12, a hydrogen bond between backbone of Ub-G35 and RNF168-H30, and a hydrophobic interaction between Ub residues L71, L8, I36 with RNF168 residues C54, and F53. (C) Detailed view on the catalytic center in the closed-state E3-E2~Ub structures superposed on the E3-E2-nucleosome model for RNF168. Crucial side chains are properly aligned to enable Ub transfer, further validating our model. D117, which has been termed the gateway residue, is hydrogen-bonding to the target lysine in our model. This may promote activation of the target lysine by stabilizing its deprotonation. Structures in panels (A) are superimposed on the RING domain of the E3, in panel (B) on the E2. Color coding indicated in the figure.



Supplementary Figure 2-7. Test of E2 histone-dimer interaction.

Normalized intensity ratios (top) and chemical shift perturbations (CSPs, bottom) observed for H2B within H2A/H2B dimers upon addition of 0.05 molar equivalents Ubch5c in the presence of 0.16 equivalents of RNF168. Residues with CSPs that are one (two) standard deviation higher than mean, or with intensities one (two) s.d. lower than the mean, are color coded yellow (red) and labeled. The small effects are observed for the same set of residues as for RNF168 addition (see main text Figure 2-1 and Supplementary Figure 2-1). This suggest the ternary complex is not stable enough to be observed under these conditions. Addition of more equivalents of E2 resulted in problems with complex stability.

SUPPLEMENTARY REFERENCES

1. Matthew Allen Bullock, J., Schwab, J., Thalassinos, K. & Topf, M. The Importance of Non-accessible Crosslinks and Solvent Accessible Surface Distance in Modeling Proteins with Restraints From Crosslinking Mass Spectrometry. *Mol Cell Proteomics* **15,** 2491–2500 (2016).

2. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat Methods* **9,** 904–906 (2012).

3. Pruneda, J. N. *et al.* Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47,** 933–942 (2012).

3

Mimicking the Nucleosomal Context in a Peptide: Increasing the Binding Affinity Towards the Epigenetic Reader PSIP1

Based on the research article: **V. Horn**, S. Jongkees and H. van Ingen (2019). *Mimicking the nucleosomal context in a peptide: Increasing the binding affinity towards the epigenetic reader PSIP1*

Manuscript in preparation

ABSTRACT

The growing understanding of the connection between DNA regulation at the chromatin level and a manifold of pathologies sparked the search for epigenetic drugs. A promising approach is the modulation of interactions between epigenetic modifications on nucleosomes and their respective reader proteins. However, this approach proved to be a challenging task for methyllysine readers because they recognize their substrate with shallow epitopes. One of these reader proteins is PSIP1-PWWP that complexes trimethyllysine K36 on the histone H3 tail. In addition, PSIP1-PWWP relies on interactions with nucleosomal DNA for its binding which explains why isolated H3 tail peptide models show almost no binding. With NMR spectroscopy, we investigate the influence of different DNA mimicking approaches on the PSIP1-PWWP affinity and incorporate them into novel H3 tail model peptides to establish a rational for alternative targeting of shallow binding epitopes. We find that the addition of negative charges to our peptides correlates with an increase in affinity while decreasing the specificity of the binding.

INTRODUCTION

In eukaryotic cells, DNA is hosted in the small compartment of the cell nucleus by being organized as chromatin. The basic unit of chromatin is the nucleosome, the assembly formed by DNA wrapped around an octamer of histone proteins [1-3]. Post translational modifications (PTMs) of histone proteins play a key role in regulating access to genomic DNA for either DNA repair, replication, or transcription. Histone proteins can be modified in various ways, including methylation, acetylation, phosphorylation, ubiquitination and crotonylation of specific residues. These modifications are installed by so-called writers, interpreted by readers and removed by eraser proteins [4-6]. The modifications and associated effector proteins are generally considered part of the epigenetic landscape, controlling gene function. Due to their prominent role, histone modifications are associated with a wide range of pathologies [7]. This makes the interaction of epigenetic effector proteins with histones, and thus nucleosomes, an important and promising therapeutic target, resulting in a growing number of small-molecule and peptide based epigenetic drugs [8, 9].

Particularly promising are inhibitors targeting reader proteins of specific post translational modifications. This allows specific interference with one biological pathway regulated by the respective reader protein without general inhibition of the modification. In addition, inhibition of reader proteins allows to specifically silence the PTM-recognition of an effector protein while leaving additional biological functions of the protein untouched [10]. Most advanced are inhibitors that target bromo domains which recognize acetylated lysines. For lysine methylation, despite its importance in a variety of pathologies, the number of promising inhibitors is significantly lower [11]. A major difficulty in the development of inhibitors for methyllysine readers is the molecular nature of methyl side chain recognition. Here several aromatic residues form a so-called aromatic cage and complex the methylated lysine head-group via cation-π interactions. This cage typically forms a deep cavity in case of monomethylated lysine binding. For trimethylated lysines, however, the cage forms a shallow pocket to allow binding of the bulky trimethylammonium group. The lack of a pronounced binding cavity makes trimethyllysine readers an intrinsically challenging target for small-molecule inhibition.

Development of peptide-based inhibitors may benefit from studying the native binding mode of reader proteins to the modified histone. Since histone PTMs predominantly occur on the flexible N-terminal tails that protrude from the nucleosome core, many such studies have focused on histone peptides as models for the binding epitope [12-16].

However, for several modifications that are on or close to the nucleosome core, recognition of the modification is not an isolated process but rather a synergetic interplay of binding between multiple components on both the nucleosome and the reader protein [17, 18]. A striking example for synergetic binding is the recognition of trimethylated lysine 36, which sits very close to the nucleosomal DNA on the histone H3 tail (H3K36me3) and occurs in the gene body of actively transcribed genes. The eader proteins of H3K36me3 include Royal family domains that feature a shallow aromatic cage for specific methyllysine recognition. Many reader proteins of H3K36me3 show very low affinities in the millimolar range for histone tail peptides [19-21] The transcriptional co-activator PSIP1 [22, 23], which has been linked to HIV integration, autoimmune disease and cancer [24-28], has a particularly low affinity. Its PWWP domain binds a H3K36me3 peptide with a K_D of ~17 mM [17]. In the nucleosomal context, the interaction is several orders of magnitude tighter (K_D in the low micromolar range), due to an additional simultaneous interaction between the DNA and an extended positively charged surface on the reader protein (Figure 3-1A) [17, 29]. Such synergistic binding from a combined interaction with both DNA and trimethyllysine is observed also for other K36me readers [18] and is thought to be a general feature of these proteins.

Here, we aim to derive design rules for peptide-based inhibitors of K36me3 reader proteins. We use the PSIP1-PWWP domain as a model system since it represents the worst-case-scenario of K36me3 readers. To overcome the challenge posed by the shallow trimethyllysine recognition site and the extremely low peptide affinity, we aimed to incorporate the binding synergy present in the native nucleosomal context into a peptide model. We designed a series of linear and branched peptides that include both the trimethyllysine and negative charges to mimic the nucleosomal context. With detailed interaction studies using NMR spectroscopy, we show that by increasing the negative charge of the peptide, binding affinity can be boosted up to ~50-fold. Peptides with increased electrostatic interaction to the protein had reduced occupancy of the aromatic cage with the trimethyllysine side chain. This indicates a steric incompatibility between the two binding modes in our peptide models, highlighting the need to carefully optimize the fit of both binding epitopes to the protein surface. All the same, given the large increase in affinity obtained, the established framework holds potential to develop more potent inhibitors of the important, yet hard to target class of methyllysine reader domains.

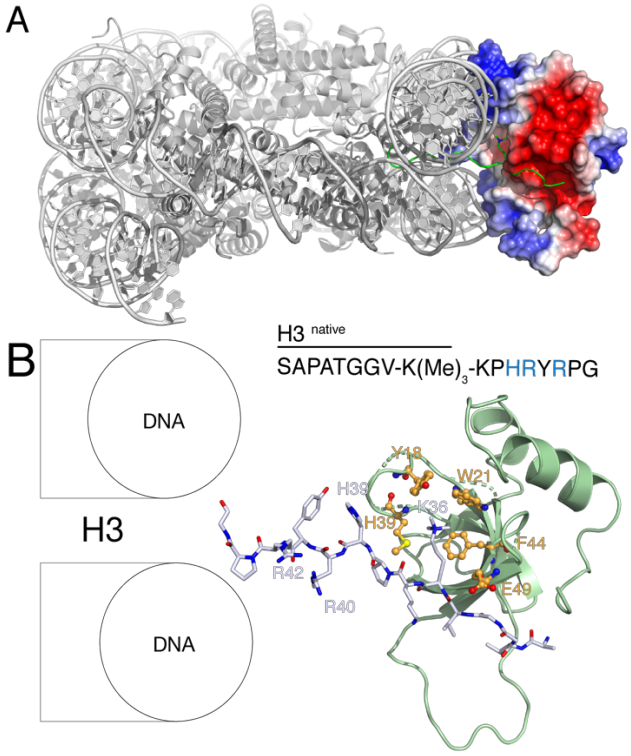


Figure 3-1. The synergetic nature of PSIP1-PWWP binding to nucleosomes.

(A) Structural model of PSIP1-PWWP bound to nucleosomes with trimethylated H3 tail residue K36 [17]. The model illustrates the electrostatic interactions with nucleosomal DNA by positively charged surface residues. (B) Schematic depiction of the H3 tail emerging between the dyads of DNA. The peptide sequence of the N-terminal H3 region is shown as sticks and annotated sequence. Positively charged residues in the H3 tail sequence are highlighted.

RESULTS

ELECTROSTATIC REPULSION DECREASES THE AFFINITY OF H3 PEPTIDES FOR PSIP1-PWWP

As a starting point for designing nucleosome-mimicking peptides for PSIP1-PWWP, we first examined the role of three positively charged residues in the H3 tail, H39, R40 and R42, that are in close proximity to the K36 modification site (Figure 3-1B). In the nucleosome, these are located in-between the two DNA gyres and thus their positive charges are essentially masked. However, when studying the binding of PSIP1-PWWP to a H3 tail peptide, these charges will no longer be masked, but fully exposed to the protein. We therefore hypothesized that the non-native electrostatic nature of a peptide-based H3 tail model peptide leads to unfavorable repulsion with positive surface residues of PSIP1-PWWP.

To dissect the influence of these residues, we mutated H39, R40 and R42 to alanine and studied the binding of this partly neutralized H3 tail peptide ($H3^{neutral}$) to the PSIP1-PWWP domain using NMR spectroscopy. NMR is particularly well suited in this case because it allows studying the low affinity interactions as well as obtaining detailed insights into the binding mode of the peptide.

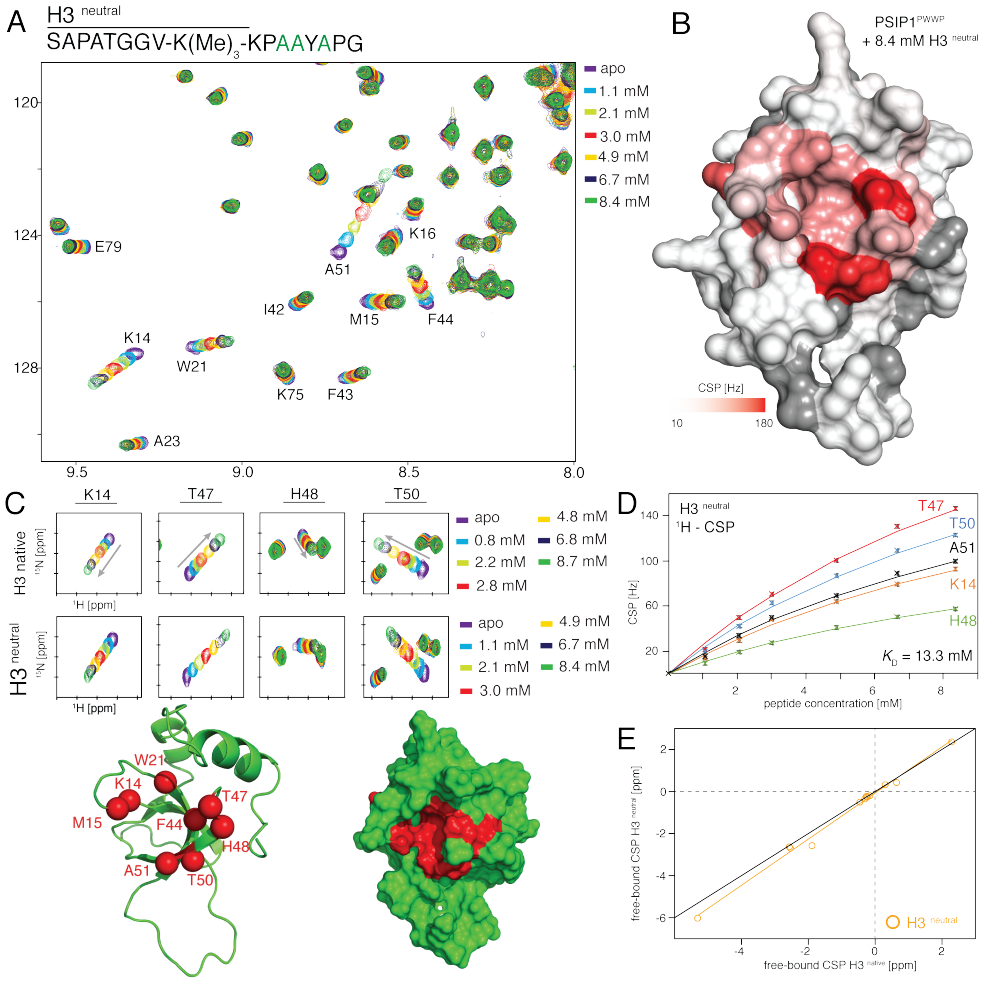


Figure 3-2. Removal of positive charges in H3 tail modestly boosts PSIP1-PWWP binding.

(A) 15N-HSQC spectrum of H3^{neutral}. NMR titration experiments show significant chemical shift perturbations for residues interacting with the N-terminus of H3^{neutral}. (B) CSPs upon binding of H3^{neutral} plotted and color coded on the PSIP1-PWWP surface. Peptide concentration is indicated. Largest changes are observed around the aromatic cage region. (C) Zoom on peak trajectories for selected residues with significant CSPs in the H3^{native} and H3^{neutral} titration. While peak trajectories have similar directions, the CSPs are significantly larger at comparable peptide concentrations for H3^{neutral}. Tic marks are placed each 0.1/0.5 ppm in the ^{1}H/^{15}N dimension. (Top). The locations of selected reporter residues on the PSIP1-PWWP surface cluster in and around the aromatic cage (residues shown in panel C + D highlighted in red) (Bottom). (D) CSP-derived binding curves with best fits (solid lines) for selected residues. Color coding and best-fit K_D indicated in the Figure. (E) Comparison of the free-bound chemical shift difference for binding H3^{native} and H3^{neutral}, based on best-fits of the titration data. Black line indicates perfect correlation.

Addition of the mutant H3-tail peptide to ^{15}N-labelled PSIP1-PWWP domain resulted in clear and specific chemical shift perturbations (CSPs). We observed pronounced

CSPs for several residues including T47, T50, A51 that are part of the β4-strand and known to form a beta-sheet with H3 tail residues 34-36 (Figure 3-2A). Comparison with titration data of the native H3 tail peptide [17] shows that the affected residues for both peptides are clustered in two main binding sequences (BS), K14 – A23 (BS I) and F44 - A51 (BS II), resulting in the same mapped binding site in and around the aromatic cage (Figure 3-2B and Figure S3-1). Furthermore, comparison of the $H3^{native}$ and $H3^{neutral}$ titrations shows that the peak trajectories have nearly identical directions, indicating comparable bound-state chemical shifts and thus a comparable bound-state conformation for the mutant complex (Figure 3-2C). The observed CSPs are however significantly increased in the $H3^{neutral}$ titration with respect to $H3^{native}$ at an equivalent peptide concentration (~3 mM) , as shown in Figure 3-2C. A global fit of the complete titration data (up to ~8 mM peptide) yields a best-fit K_D value of 13.3 mM for $H3^{neutral}$ (Figure 3-2D). The low affinity hampers precise determination of the K_D, which is reflected in the rather wide 95% probability limits for the K_D, with lower limit of 9.3 and upper limit of 21.1 mM. These values are overall lower than the K_D of 17.3 mM (13.9-22.1 mM 95% probability limits) for the $H3^{native}$ peptide. The fits further allow to compare the binding modes of the two peptides based on the fitted bound-state chemical shifts in the $H3^{native}$ and $H3^{neutral}$ titrations. We selected residues that are situated in and directly around the aromatic cage and report on binding of the trimethyllysine by formation of the intermolecular beta-sheet (see Figure 3-2C, bottom). Strikingly, the two datasets independently arrive at very similar chemical shifts for the bound-state for the reporter residues (Figure 3-2E). Thus, our data indicate that the $H3^{neutral}$ peptide binds in a native binding mode to PSIP1-PWWP with a modestly increased affinity. We conclude that charge repulsion between H39, R40, R42 and the positive PWWP surface is contributing to the low affinity of PSIP1-PWWP towards H3 tail peptides, indicating a need to introduce additional attractive interactions to boost affinity.

INTRODUCTION OF NEGATIVE CHARGES ENHANCES PEPTIDE BINDING TO PSIP1-PWWP

We next designed two H3 tail peptide models that incorporate negatively charged residues to mimic the additional attractive electrostatic forces present in a nucleosomal context. In the first sequence, the three positively charged residues H39, R40 and R42 are mutated to glutamate (denoted as $H3^{E3}$). Since in the native binding mode the C-terminal part of the H3 tail is orientated towards the DNA-binding surface of PSIP1, we designed a second sequence, in which all seven residues on the C-terminal side of Kme3 are substituted by glutamate to create a highly negatively charged H3 tail model ($H3^{E7}$). By comparing the effects of these two Glu-rich peptides, the impact of

a step-wise increase in negative charge can be monitored. NMR titration experiments with both H3^{E3} and H3^{E7} resulted in clear chemical shift changes for PSIP1-PWWP residues in the two binding sequences BS I and BS II, that are clustered in and around the aromatic cage (Figure 3-3A and Figure S3-1). As for H3^{neutral}, most residues show similar direction of CSPs, pointing to a similar bound-state conformation (Figure 3-3B and see Figure 3-2C). At similar concentrations of ligand, the CSPs observed for H3^{E3} and H3^{E7} are much larger compared to the native H3 tail peptide, indicating that H3^{E3} and H3^{E7} bind with significantly increased affinity to PSIP1-PWWP (Figure 3-3C). Global fits of the NMR titration profiles result in a best-fit K_D of 8.5 mM (95% probability limits 5.9-13.0 mM) for binding to the H3^{E3} peptide. Strikingly, the affinity of H3^{E7} is strongly increased, with a best-fit K_D value of 1.2 mM (95% probability limits 0.97-1.5 mM). These values correspond to a 2- and 14-fold increase in affinity over the native H3 tail peptide.

Interestingly, comparison of the bound-state chemical shifts derived from the fits show that binding of both Glu-rich peptides results in nearly identical chemical shifts for the trimethyllysine and beta-sheet reporter residues (Figure 3-3D). At the same time, the bound-state shifts are all smaller than observed for the native peptide. Strikingly, all bound-state shifts are uniformly scaled to ~55% of their values for the native peptide.

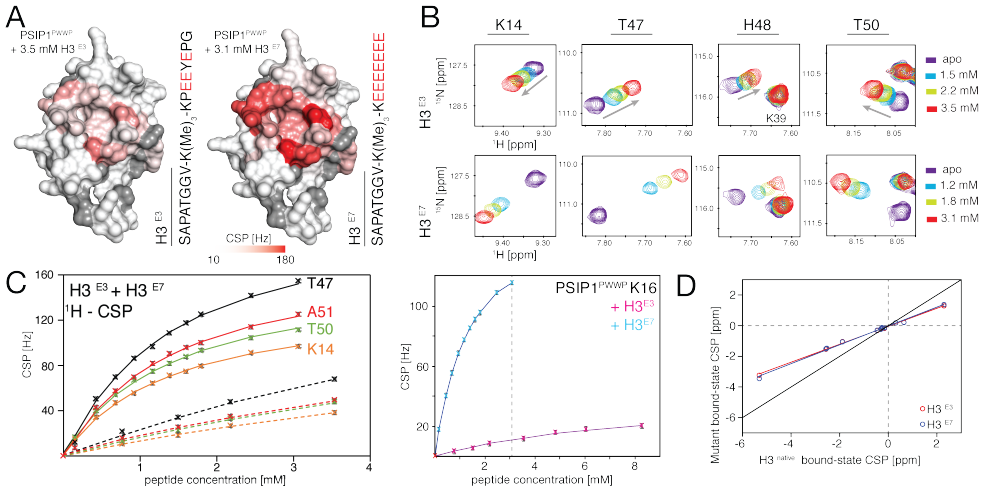


Figure 3-3. Glu-rich H3 peptides have increased affinity for PSIP1-PWWP with altered methyllysine binding.

(A) CSPs upon binding of H3^{E3} and H3^{E7} plotted and color coded on the PSIP1-PWWP surface. Peptide concentration is indicated. Largest changes are observed around the aromatic cage region. (B) Zoom on peak trajectories for selected reporter residues around the aromatic cage in the H3^{E3} and H3^{E7} titration. While peak trajectories have similar directions, the CSPs are significantly larger at comparable peptide concentrations for H3^{E7}. Tic marks are placed each 0.05/0.5 ppm in the ^{1}H/^{15}N dimension. (C) CSP-derived binding curves with best fits (solid lines for H3^{E7}; dashed lines for H3^{E3}; dotted lines for H3^{native}) for selected residues. Color coding and best-fit K_D with 95% probability limits indicated

in the Figure. Fits for $H3^{E3}$ are based on titration up to 8.28 mM peptide; full data is shown in Figure S3-2. (D) Comparison of the free-bound chemical shift difference for binding $H3^{native}$ and $H3^{E3}$ (red) or $H3^{E7}$(blue), based on best-fits of the titration data. Black line indicates perfect correlation.

ROLE OF A SPECIFIC INTERACTION BETWEEN PWWP-K16 AND $H3^{E7}$

The addition of four more negative charges in the $H3^{E7}$ peptide resulted in a strong increase in binding affinity. Comparison of the observed CSPs highlights one residue, PWWP K16, with significantly larger shift in the titration with $H3^{E7}$ compared to $H3^{native}$, $H3^{neutral}$ and $H3^{E3}$ (Figure 3-4A and Figure S3-1). While in the latter three titrations this residue experiences highly similar CSPs indicating a comparable chemical environment, K16 exhibits a much more pronounced perturbation for its amide proton in the $H3^{E7}$ titration. This points, together with the increased affinity, towards a crucial effect of K16 on substrate binding for $H3^{E7}$, possibly by engaging in an intermolecular interaction with the peptide. In order to get further insights into the influence of K16 on PSIP1-PWWP binding, we used the NMR-derived data on the protein interaction interface in a data-driven docking effort to obtain a molecular model for the PWWP–$H3^{E7}$ complex using the HADDOCK program [30, 31]. Based on the bound-state chemical shift analysis described above, we enforced a native-like orientation of the methyllysine and the formation of the intermolecular beta-sheet. Residues in the Glu-rich tail were defined as fully flexible, thus allowing any of these residues to be close to the CSP-identified binding interface of PWWP. As a result, the best-scoring solutions show large variation in orientation of the Glu-rich tail (Figure 3-4B). Nevertheless, most top scoring structures feature a hydrogen-bond between PSIP1-PWWP K16 and $H3^{E7}$ residue E38. The latter residue is only present in $H3^{E7}$ while all other peptide models contain the native proline on that position. To test whether this interaction is responsible for the large increase in binding affinity for $H3^{E7}$, we tested binding of a mutant $H3^{E7}$ peptide retaining the native P38 ($H3^{PE6}$). The overall analysis of shifting residues shows that for $H3^{PE6}$ both BS I and BS II are present as may be expected (Figure S3-1). The direction and extent of CSPs for the key shifting residues (K14, T47, H48, T50 and A51) are conserved to previous experiments, confirming a native-like methyllysine interaction for $H3^{PE6}$ (Figure S3-3). Notably, the CSP for K16 is very similar to that seen for $H3^{native}$, $H3^{neutral}$ and $H3^{E3}$ that also contain the native P38 in their sequence (Figure 3-4C). When put in the context of the surrounding residues in BS I, it can be seen that the CSP for K16 is comparable to that for K14 and G17 in $H3^{E7}$. However, for $H3^{PE6}$ the CSP of K16 is decreased to the level of M15 as is the case for $H3^{E3}$, without affecting the CSPs of other residues (Figure 3-4D). These data thus validate the presence of a specific interaction between K16 and E38 as seen in the model for the $H3^{E7}$ complex. Of note, the CSP for K14 is increased in both $H3^{PE6}$ and

H3^{E7} compared to H3^{E3}, indicating additional electrostatic interaction for these Glu-extended peptides.

We next analyzed the H3^{PE6} titration data quantitatively to determine the contribution of E38/K16 interactions to the binding affinity. Best-fits of the CSP-derived binding curves resulted in a K_D of 1.74 mM (95% probability limits 1.64 – 1.87 mM) (Figure S3-2). Surprisingly, the affinity of H3^{PE6} is only modestly decreased compared to H3^{E7} (1.2 mM) and far below the K_D for H3^{E3} (8.5 mM). Notably, the 95% confidence intervals of K_D values for H3^{PE6} and H3^{E7} do not overlap indicating their difference is significant. The K16/E38 interaction is thus not the sole critical determinant of the affinity gain for H3^{E7} over H3^{E3}. This implies that both peptides rely on electrostatic interactions with charged patches rather than well-defined interactions on both the peptide and protein side.

Overall, the data for these linear Glu-rich peptides as straightforward mimics of the nucleosomal context of the K36me3 modification emphasize the potential of electrostatic interactions between PSIP1-PWWP and peptide substrates to boost affinity and highlight the challenge in engineering specific electrostatic interactions and balancing these against the interaction with the methyllysine-binding pocket.

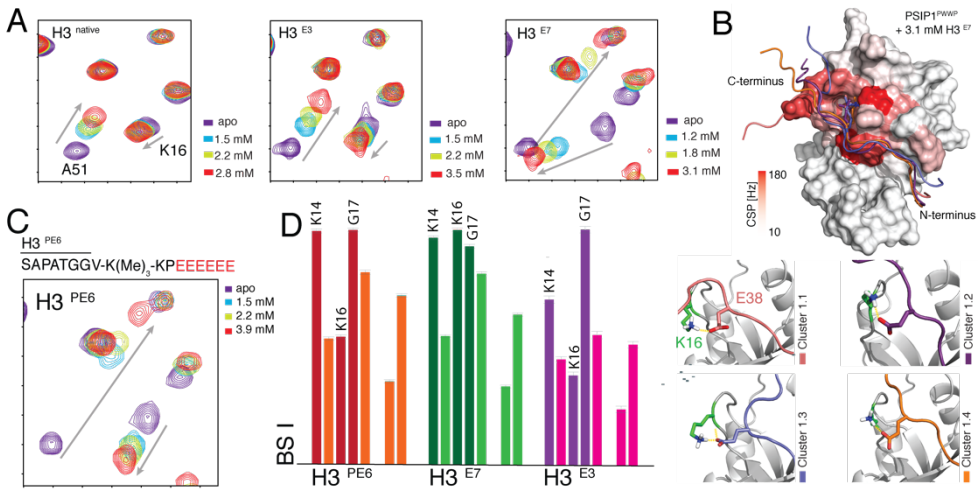


Figure 3-4. Identification of a specific interaction between PWWP-K16 and H3-E38 in H3^{E7}.

(A) Zoom on peak trajectories for A51 and K16 in the H3^{native}, H3^{E3} and H3^{E7} titrations. The direction and extent of peak displacement for K16 differs between H3^{E7} and H3^{native}/H3^{E3}. (B) Structural model generated by data-driven docking of H3^{E7} bound to PSIP1-PWWP. Depicted are the best four structures of the top scoring cluster, containing 172/198 structures (1: salmon, 2: purple, 3: blue, 4: orange) (Top). Detailed view on the conserved hydrogen bond between residues K16 (PSIP1-PWWP) and E38 (H3^{E7}) four the four best scoring structures (cluster1.1 – cluster1.4) (Bottom). (C) Zoom on peak trajectories for A51 and K16 in the H3^{PE6} titration. The direction and extent of peak

displacement corresponds well that for H3native and H3E3, shown in panel (A). (D) Observed CSPs upon binding of indicated peptide for PSIP1-PWWP residues in the binding sequence I (BS I). CSPs are calculated from apo spectra and with 5.47 mM $H3^{PE6}$, 3.5 mM $H3^{E3}$ and 3.1 mM $H3^{E7}$ (For comparison scaled to the same bar height for G17). Tic marks in panels (A) and (C) are placed each 0.05/0.5 ppm in the $^{1}H/^{15}N$ dimension.

INFLUENCE OF NUCLEOSOME-LIKE CHARGE DISTRIBUTION ON H3 PEPTIDE COMPLEXES

We next sought to improve further on the linear Glu-rich H3 peptides by not only mimicking part of the electrostatic landscape in the nucleosomal context, but also the steric organization of charges. When binding K36me3-modified nucleosomes, two strands of DNA encompass the H3 tail that emerges between them. Therefore, we designed and synthesized a branched peptide structure that consists of two negatively charged extensions to the central methyllysine containing sequence ($H3^{branched}$) (Figure 3-5A). This central sequence (Kme3-VGGTAPAS) is connected via two glycine residues to the side chain amino group of lysine. This lysine forms the branching point and is extended on both its N- and C-terminus by glycine-linked glutamate-rich tails. Since we observed that substitution of the native sequence with Glu-residues altered the extent of trimethyllysine interactions with the aromatic cage, we here introduced glycine residues as flexible linkers between the trimethyllysine part and the Glu-rich tails. Such design potentially allows to simultaneously engage the aromatic cage and a large part of the PSIP1-PWWP DNA-binding surface.

NMR titration experiments with the $H3^{branched}$ peptide showed the expected CSPs for PWWP residues in BS I and II that are in or around the aromatic cage moiety (Figure 3-5B and Figure S3-1). Furthermore, the direction of peak displacement for these residues is similar to the linear peptide models (Figure 3-5B). Visualization of the CSPs on the PSIP1-PWWP surface shows that the interaction surface is expanded with respect to the linear peptides (Figure 3-5C). Residues K73, R74 and K75 experience a significant perturbation of their chemical environment upon addition of $H3^{branched}$. This additional binding sequence (BS III) did not exhibit pronounced CSPs previously (Figure S3-1). The expanded interaction surface of $H3^{branched}$ suggests that the two Glu-rich tails are interacting with the PWWP surface as intended.

Markedly, saturation of binding can already be observed upon addition of 1.4 mM peptide (Figure S3-2). In addition, residues with larger CSPs, such as A51, show significant resonance broadening during the titration, indicating a fast-to-intermediate exchange regime between free and bound states of the protein (see Figure 3-5B). Correspondingly, a global fit of the CSP-derived binding curves arrives at a significantly

improved K_D of 360 μM (95% probability limits 190-700 μM). This constitutes a significant improvement in affinity over the linear peptides and a nearly 50-fold increase over the native peptide.

Surprisingly, even if the protein is nearly saturated with peptide at the end of the titration, the CSPs observed are much smaller in magnitude than for the linear peptides. Comparison of the bound-state chemical shifts for the aromatic cage reporter residues shows that for H3^{branched} the bound-state shifts are scaled to ~20% of their native values (Figure 3-5D and E). It is worth emphasizing that the direction of the CSPs is the same as for the native and linear peptides, indicative a conserved binding mode. Yet, the uniform scaling of the bound-state shifts strongly suggests a global change in the aromatic cage-trimethyllysine interaction for the H3^{branched} peptide.

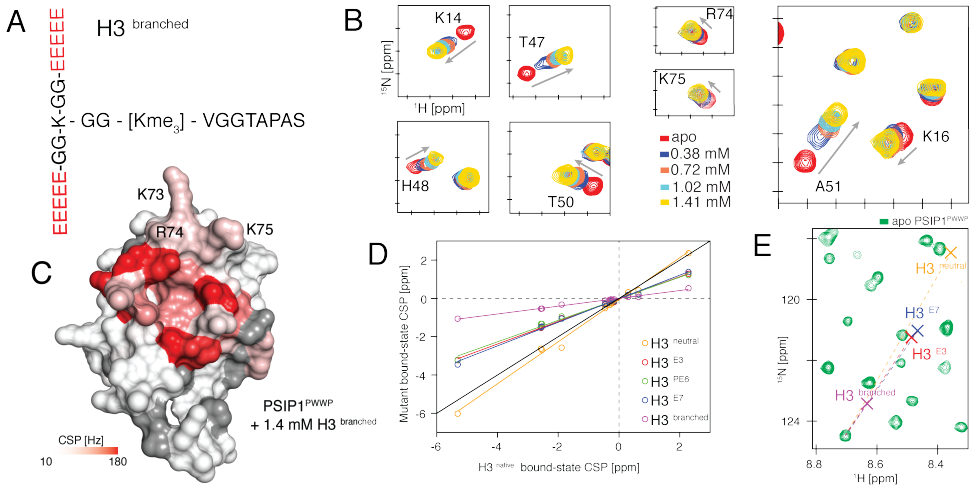


Figure 3-5. An extended and high-affinity interaction of a branched nucleosome mimicking peptide with PSIP1-PWWP.

(A) Sequence of H3^{branched} peptide. (B) Zoom on peak trajectories for selected residues in the H3^{branched}. Tic marks are placed each 0.05/0.5 ppm in the $^{1}H/^{15}N$ dimension. (C) CSPs upon binding of H3^{branched} plotted and color-coded on the PSIP1-PWWP surface, highlighting the extended interaction surface compared to the linear peptides. Peptide concentration is indicated. (D) Comparison of the free-bound chemical shift difference for binding H3^{native} and all peptide-models of this study, based on best-fits of the titration data. Black line indicates perfect correlation. Color coding indicating in the figure. (E) Section of the ^{15}N-HSQC spectrum showing the peak trajectory for A51 in the H3^{branched} apo spectrum with bound-state peak positions as derived from the corresponding fit for all peptide substrates indicated.

DISCUSSION

In this work we studied the interaction of various peptides with the PSIP1-PWWP domain, a reader of the H3K36me3 modification. Like other reader domains that specifically interact with trimethylated lysines, PSIP1-PWWP features a shallow binding pocket that is generally hard to target for inhibitors. Since the PSIP1-PWWP domain binds extremely weakly to H3-tail peptides, but with high affinity to modified nucleosomes thanks to additional DNA binding, we explored the potential of including such additional binding mode into a peptide model. From a series of NMR titration experiments with trimethyllysine peptides ($H3^{neutral}$, $H3^{E3}$, $H3^{PE6}$, and $H3^{E7}$) we found that binding affinity indeed increases in step-wise manner with electronegativity (Figure 3-6A). For the best linear peptide, containing a Glu-rich tail of seven residues after the modified lysine affinity was increased ~ten-fold over the native peptide. We additionally designed a branched peptide $H3^{branched}$ with two Glu-rich tails to mimic the structural arrangement of the negative charges in the nucleosomal context. The affinity of this peptide for PSIP1-PWWP was in the micromolar range, a five-fold increase compared to the best linear model, thus a ~50-fold improvement over the native H3-tail peptide.

For the two highest affinity peptides, $H3^{E7}$ and $H3^{branched}$, the NMR data evidenced additional interactions between peptide and protein compared to the native peptide. From a comparison of chemical shift perturbations in $H3^{PE6}$ and $H3^{E7}$, we could identify a specific interaction between $H3^{E7}$ E38 and PSIP1-PWWP K16. Removal of this interaction in $H3^{PE6}$ reduced affinity only slightly, indicating overall electrostatic interactions drive complex formation rather than specific interactions. Perturbations upon binding of $H3^{branched}$ included a larger part of the DNA-binding surface of PSIP1-PWWP, validating our hypothesis that inclusion of the DNA-mimicking Glu-rich tails boosts affinity by engaging with the positively charged PSIP1-PWWP surface.

Strikingly, our data indicate that increased electrostatic binding alters the interaction of the trimethyllysine and the aromatic cage in a subtle way. In all peptide titrations, the direction of the peak trajectory from the free to the bound state is highly similar for the reporter residues in and around the aromatic cage, strongly suggesting that the binding mode of the trimethyllysine is conserved. This also fits with the fact that the Glu-mutations are all distant from the trimethyllysine. The magnitude of the CSPs to the bound state, however, is strongly reduced for the Glu-rich linear and even more so for the branched peptide. Since the bound-state of all reporter residues are scaled uniformly, the aromatic cage seems to experience a global change in the binding mode.

This indicates that at the point where the protein is saturated with peptide, the trimethyllysine binding pocket is only fractionally occupied (Figure 3-6B). In this scenario, the chemical shifts of the fully bound aromatic cage, $\Delta\omega_{aro,bound}$, are scaled by the fraction in which the cage is bound by the trimethyllysine, $f_{\text{aro,bound}}$, resulting in the chemical shifts observed at end of the titration with one the peptide substrates, $\Delta\omega_{sat,peptide}$:

$$\Delta\omega_{sat,peptide} = f_{\text{aro,bound}} * \Delta\omega_{aro,bound}$$

Notably, it is implicit here that the intrinsic, microscopic affinity for the trimethyllysine interaction is sufficiently low to result in a high dissociation rate and thus fast-exchange between free and bound states. Such low intrinsic affinity is consistent with the low affinity for the native peptide.

Mechanistically, we interpret the increase in affinity and decrease in occupancy of the aromatic cage upon increasing negative charge as an altered balance in a two-step binding process, as shown in Figure 3-6B. The charged peptides first bind electrostatically using their Glu-rich tail, after which the trimethyllysine can dock into the aromatic cage. The electrostatic complex formed in the intermediate state is akin to an encounter state, such as shown in the formation of various protein-protein complexes [32, 33]. It was shown that proteins that engage in an encounter complex based on electrostatics with no specific orientation show small CSPs upon binding. It was hypothesized that the interaction interface during purely electrostatic encounters is still solvated and therefore does not influence the chemical environment of the observed backbone amides [34, 35]. These encounter states can be observed in paramagnetic NMR experiments, but are otherwise hardly detectable [36]. Accordingly, throughout our titration experiments we observed that CSPs for PSIP1-PWWP residues engaging in electrostatic interactions are very limited (see also Chapter 4 Figure 4-7). We additionally did not observe a pronounced effect for the deletion of E38 in H3^{PE6} besides its indicated involvement in H3E7 binding. This supports that what we observe is a highly dynamic electrostatic interaction that relies on interactions between charged patches rather than well-defined side chain interactions. Our peptides thus stabilized primarily the encounter state, at expense of the trimethyllysine-cage interaction. We thus find that the electrostatic interactions compete with effective binding of the trimethyllysine, "pulling" it out of the aromatic cage. Interestingly, our data indicate the aromatic cage is bound to a similar extent (~55%) for all three linear Glu-rich peptides, while for double-tailed branched peptide, despite featuring a double glycine linker, the fraction bound drops to ~20% at saturation. This continuous decrease highlights the impact of the electrostatic interactions we introduced and the need to carefully

optimize the steric fit between negatively charged moieties and the native trimethyllysine binding mode. At the same time, successful optimization will further boost affinity, thus leaving ample room for future improvement.

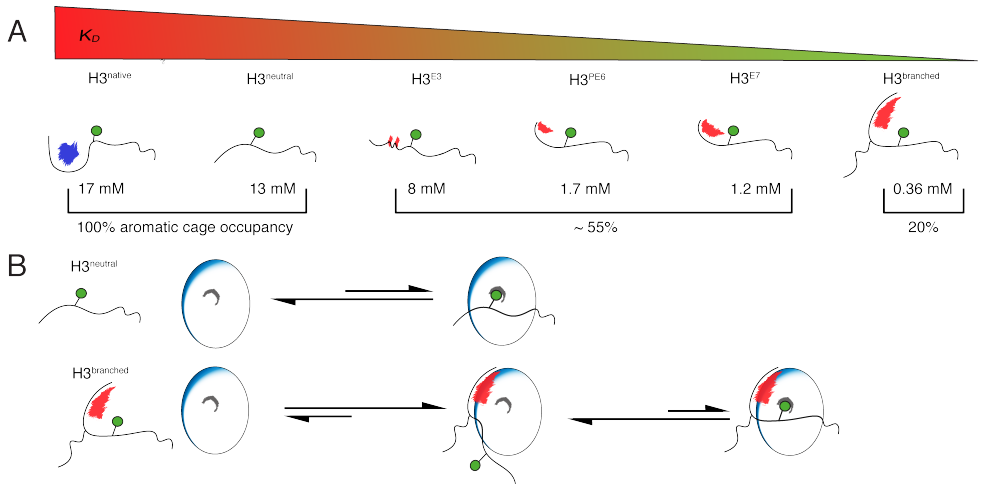


Figure 3-6. Possible binding mechanism for the studied peptide models and their respective binding affinities.

(A) A schematic illustration of the peptide models used in this study. A continuous manipulation of the peptide sequence towards the incorporation of increasing negative charges results in a consistent boost of binding affinity ($H3^{neutral}$ < $H3^{E3}$ < $H3^{PE6}$ < $H3^{E7}$). This can be further optimized by introducing structural motifs that resemble the electrostatic surrounding of the H3 tail in a nucleosomal context ($H3^{branched}$). However, a decrease in the fraction of occupied aromatic cage suggests that the improved binding is due to electrostatic interactions. (B) Possible binding mechanism, explaining the observed effects for mutant peptide binding. H3native/H3neutral engage in a PSIP1-PWWP binding that does not include electrostatic interactions and has Kme3 complexation as sole driving force. Therefore, 100% of the observed binding correlates to an occupied aromatic cage. In case of negatively charged mutants and especially H3branched, the peptide engages in an encounter complex facilitated by electrostatic interactions. Of this complex, only a fraction of Kme3 (20%) is additionally complexed by the aromatic cage (reaction rates indicated with arrow length).

CONCLUSIONS

We here presented rationally designed peptide-based binders of the histone trimethyllysine reader domain of transcriptional co-regulator PSIP1. Inclusion of negative charges to mimic the nucleosomal context of the native binding mode resulted in a ~50-fold improvement in binding affinity over a simple histone peptide. The introduced charges stabilize an encounter complex, resulting in decreased trimethyllysine binding. We envisage that further screening of variant peptides for

structural optimization of both binding modes will result in further improved binding. The framework presented here to use electrostatic interactions to position the peptide in close proximity to the trimethyllysine binding pocket may be generally applicable to K36 trimethyllysine reader proteins. Combined targeting of both the electrostatic surface and the shallow Kme3 binding epitope of these proteins may be a promising strategy in the development towards potent chemical probes and inhibitory compounds.

MATERIALS AND METHODS

PROTEIN EXPRESSION

For NMR studies, PSIP1-PWWP (AA 2-100, Uniprot-ID: O75475) was expressed as a GST-fusion protein in BL21 Rosetta2 (DE3) cells (Novagen). PSIP1-PWWP was produced in M9 minimal medium containing $^{15}NH_4Cl$. Cells were grown to an OD of 0.6 and induced with 500 µM IPTG. The protein was expressed for 3 h at 37 °C. The cells were harvested in lysis buffer (50 mM Tris pH 8.0, 300 mM KCl, 0.5 mM PMSF, 1 mM DTT, 3.3 mM TritonX-100, complete EDTA-free protease inhibitor (Roche) and lysed by French-press. The soluble fraction was collected after ultracentrifugation at 25000 rpm and 4 °C for 40 min, and subsequently loaded on a 5 mL GSTrap HP column (GE Healthcare). The column was washed with wash buffer I (50 mM Tris pH 8.0, 300 mM KCl, 1 mM DTT, complete EDTA-free protease inhibitor (Roche)) and wash buffer II (50 mM KPi pH 7.0, 100 mM KCl, 1 mM DTT, complete EDTA-free protease inhibitor (Roche)). The GST-fusion protein was eluted using elution buffer (50 mM KPi pH 7.0, 100 mM KCl, 50 mM reduced glutathione, 1 mM DTT, complete EDTA-free protease inhibitor (Roche)). Thrombin cleavage of the GST-tag was carried out for 3h at 37 °C after adding 10U of thrombin to the combined eluted fractions, then quenched by addition of 1 mM PMSF and incubation for 15 min at 37 °C. The PSIP1-PWWP domain was then purified using ion-exchange chromatography over a HiTrap SP HP column (GE Healthcare). A gradient from IEX buffer (20 mM KPi pH 7.0, 0.2 mM PMSF, 0.5 mM DTT) w/o KCl to IEX buffer with 1 M KCl was used to elute PSIP1-PWWP from the column. The PWWP-containing fractions were then combined and dialyzed to NMR buffer (20 mM NaPi (pH = 7.0), 150 mM NaCl, cOmplete protease inhibitor, 10 %(V/V) D_2O)

PEPTIDES

$H3^{neutral}$ (Ac-SAPATGGV[Kme_3]KPAAYAPG-NH_2),

$H3^{E3}$ (Ac-SAPATGGV[Kme3]KPEEYEPG-NH_2),

$H3^{E7}$ (Ac-SAPATGGV[Kme3]KEEEEEEE-NH_2)

and $H3^{PE6}$ (Ac-SAPATGGV[Kme_3]KPEEEEEE-NH_2)

were supplied by Biomatik with a >95% purity as TFA salt. All peptides were dissolved in NMR buffer (20 mM NaPi (pH = 7.0), 150 mM NaCl, cOmplete protease inhibitor, 10 %(V/V) D_2O). The pH for all peptide stocks was adjusted to 7.0 using NH_4OH. Determination of peptide stock concentrations was done using 1D NMR spectra with TSP as reference compound (Figure S3-4). The integral of the isolated singlet arising from TSP methyl groups (0.0 ppm) was put in relation to the peak for Val35 methyl groups in the peptides (~0.8 ppm) to back-calculate the effective peptide concentrations. The Val signal was verified by 2D NMR spectra (TOCSY, ^{13}C-HSQC, see Figure S3-4). The stock concentrations were 70.54 mM ($H3^{neutral}$), 50.98 mM ($H3^{E3}$), 29.74 mM ($H3^{E7}$), 11.85 mM ($H3^{branched}$) and 26.8 mM ($H3^{PE6}$).

SYNTHESIS $H3^{BRANCHED}$

$H3^{branched}$ was synthesized manually, using Fmoc-based solid phase peptide synthesis. The backbone sequence (EEEEEGGKGGEEEEE) was synthesized on a Rink amine resin (Rapp polymere TentaGel S RAM) with the lysine side chain amine protected with an Mmt group. The coupling steps were performed as double couplings for 40 min each with 4 eq. amino acid together with HBTU/HOBt activation and DIPEA as base. After every coupling step and at the end of the synthesis, free backbone amide was inactivated by acetylation. For branching, Lys(Mmt) was selectively deprotected on-resin with a cleavage solution of 2%(V/V) TFA, 98%(V/V) DCM. Completion was ensured by monitoring the yellow color of the trityl-cation. The synthesis of the branch was propagated via the e-N of the Lys side chain. All following amino acids were coupled using the DIC/HOAt strategy with 10 eq. amino acid and double coupling. The crude product was cleaved off the solid support using a cleavage solution of 95%(V/V) TFA, 2.5%(V/V) TIPS and 2.5%(V/V) H2O. Purification of the peptide was done using LC-MS, isolating a peak with molecular mass of 1349.8/1350.2 g/mol (m/z = 2) $+Na^+$ (theoretical: 2674.7 g/mol). The product was dissolved in H_2O and evaporated three times to co-evaporate residual TFA. Afterwards the product was

lyophilized and dissolved in NMR buffer. The stock concentration was found to be 11.85 mM at pH 6.97 using the method described above.

NMR SPECTROSCOPY

All NMR experiments were carried out at 298 K on a Bruker Avance spectrometer operating at 750 MHz ^{1}H Larmor frequency with room-temperature probe. Processing was done using the Bruker's TopSpin. Spectra were analyzed using Sparky (Goddard and Kneller, UCSF). Samples for the titration experiments contained ~300 μM ^{15}N-PSIP1-PWWP in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)).

TITRATION DATA ANALYSIS

Titration data were fitted using MatLAB scripts using the fast-exchange assumption. Briefly, titration profiles from non-overlapping residues with CSPs larger than 50 Hz ($H3^{neutral}$, $H3^{E7}$, $H3^{PE6}$), 20 Hz ($H3^{E3}$) or 10 Hz ($H3^{branched}$) were selected to be included in a simultaneous fit of titration curves to determine the K_D. Profiles were selected either form the 1H or 15N dimension. A 1:1 binding model was assumed. Fitting parameters are residue-specific values for an offset term to compensate for inaccuracies in the initial peak position (in all cases < 10 Hz), the residue-specific values for the free-bound chemical shift differences, and the global K_D value.

$$\chi^2_{CS} = \sum_j^M \sum_i^N \left(\frac{\delta_{calc}\left(\kappa_D, k_{off}, offset\right) - \delta_{obs}}{\sigma_{nucleus}} \right)^2$$

$$DOF_{CS} = NM - M - 3 - \theta_m - \varphi_m$$

where δ is the peak position, *nucleus* = ^{15}N or ^{1}H, σ_{1H} = 1.5 Hz and σ_{15N} =1.5 Hz, *i* runs over all *N* titration points and *j* runs over all *M* selected resonances. The degrees of freedom are equal to total number of points *NM* compensated for the number of global variables, the *M* offset-variables and *M* chemical shift differences. To obtain the 95% confidence limits in the K_D, a one-dimensional grid search was performed either systematically increasing or decreasing the K_D until the reduced χ^2 exceeded the 95% critical value according to an F-test.

Analysis of the bound-state chemical shifts was based on Selected residues K14 ($^{1}H/^{15}N$), M15 (^{1}H), W21 (^{1}H), F44 (^{15}N) T47 ($^{1}H/^{15}N$), H48 (^{1}H), T50 ($^{1}H/^{15}N$) and A51 ($^{1}H/^{15}N$) due to their significant shift in all experiments.

DATA-DRIVEN DOCKING

The structural model of the PSIP1-PWWP/$H3^{E7}$ complexes was determined using the experimental NMR data in the data-driven docking software HADDOCK. The $H3^{native}$ structure [17] was used as a template to build an atomic model for the $H3^{E7}$ peptide including amidation (C-terminus) and acetylation (N-terminus). This was docked to the PSIP1-PWWP surface from the nucleosome-bound structural model [17]. Insertion into the aromatic cage was enforced by unambiguous restraints between K36me3 and PSIP1-PWWP residues M15, Y18, W21 and F44. Kme3 was set was an active residue. The H3 tail sequence part of the peptide was directed into a native confirmation by enforcing O-N hydrogen bonds between PSIP1-PWWP T50(O) and K36me3 (N) as well as T50 (N) and G35 (O). Glutamate residues on the peptide were defined as fully flexible segment. 13 ambiguous restraints were included for PSIP1-PWWP residues with observed CSP and neighboring solvent accessible residues (see Figure 3-3B and Figure S3-5). Docking was done with increased MD steps for rigid body high temperature TAD (2000), first rigid body cooling stage (2000), second cooling stage with flexible side-chains at interface (4000) and third cooling stage with fully flexible interface (4000) [37]. Of the 200 water-refined solutions, 198 were clustered in four clusters (Figure S3-6).

REFERENCES

1. Kornberg, R.D., *Chromatin structure: A repeating unit of histones and DNA*. Science, 1974. **184**(4139): p. 868-871.
2. Thomas, J.O. and R.D. Kornberg, *An octamer of histones in chromatin and free in solution*. Proceedings of the National Academy of Sciences, 1975. **72**(7): p. 2626.
3. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**: p. 251.
4. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
5. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy*. Cell, 2012. **150**(1): p. 12-27.
6. Speranzini, V., et al., *Touch, act and go: landing and operating on nucleosomes*. EMBO J, 2016. **35**(4): p. 376-88.
7. Sarah, H., et al., *Use of Epigenetic Drugs in Disease: An Overview*. Genetics & Epigenetics, 2014. **6**: p. GEG.S12270.
8. Pérez-Salvia, M. and M. Esteller, *Bromodomain inhibitors and cancer therapy: From structures to applications*. Epigenetics, 2017. **12**(5): p. 323-339.
9. Arrowsmith, C.H., et al., *Epigenetic protein families: a new frontier for drug discovery*. Nature Reviews Drug Discovery, 2012. **11**: p. 384.
10. Greschik, H., R. Schüle, and T. Günther, *Selective targeting of epigenetic reader domains*. Expert Opinion on Drug Discovery, 2017. **12**(5): p. 449-463.
11. Hauser, A.-T., D. Robaa, and M. Jung, *Epigenetic small molecule modulators of histone and DNA methylation*. Current Opinion in Chemical Biology, 2018. **45**: p. 73-85.
12. Barbera, A.J., et al., *The Nucleosomal Surface as a Docking Station for Kaposi's Sarcoma Herpesvirus LANA*. Science, 2006. **311**(5762): p. 856.
13. Xiang, S., et al., *Site-Specific Studies of Nucleosome Interactions by Solid-State NMR*. Angewandte Chemie International Edition: p. n/a-n/a.
14. Musselman, C.A., et al., *Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1*. Nature Structural &Amp; Molecular Biology, 2012. **19**: p. 1266.
15. van Ingen, H., et al., *Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3*. Structure, 2008. **16**(8): p. 1245-56.
16. Forneris, F., et al., *Human Histone Demethylase LSD1 Reads the Histone Code*. Journal of Biological Chemistry, 2005. **280**(50): p. 41360-41365.
17. van Nuland, R., et al., *Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain*. Epigenetics & Chromatin, 2013. **6**(1): p. 12.
18. Musselman, C.A., et al., *Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility*. Nat Commun, 2013. **4**: p. 2969.

19. Wu, H., et al., *Structural and histone binding ability characterizations of human PWWP domains.* PLoS One, 2011. **6**(6): p. e18919.
20. Vezzoli, A., et al., *Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1.* Nat Struct Mol Biol, 2010. **17**(5): p. 617-9.
21. Xu, C., et al., *Structural Basis for the Recognition of Methylated Histone H3K36 by the Eaf3 Subunit of Histone Deacetylase Complex Rpd3S.* Structure, 2008. **16**(11): p. 1740-1750.
22. Barski, A., et al., *High-Resolution Profiling of Histone Methylations in the Human Genome.* Cell, 2007. **129**(4): p. 823-837.
23. Vermeulen, M., et al., *Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers.* Cell, 2010. **142**(6): p. 967-980.
24. Yokoyama, A. and M.L. Cleary, *Menin critically links MLL proteins with LEDGF on cancer-associated target genes.* Cancer Cell, 2008. **14**(1): p. 36-46.
25. Ciuffi, A. and F.D. Bushman, *Retroviral DNA integration: HIV and the role of LEDGF/p75.* Trends in Genetics, 2006. **22**(7): p. 388-395.
26. Ganapathy, V. and C.A. Casiano, *Autoimmunity to the nuclear autoantigen DFS70 (LEDGF): What exactly are the autoantibodies trying to tell us?* Arthritis & Rheumatism, 2004. **50**(3): p. 684-688.
27. Daugaard, M., et al., *Lens Epithelium-Derived Growth Factor Is an Hsp70-2 Regulated Guardian of Lysosomal Stability in Human Cancer.* Cancer Research, 2007. **67**(6): p. 2559.
28. Huang, T.-s., et al., *LEDGF/p75 has increased expression in blasts from chemotherapy-resistant human acute myelogenic leukemia patients and protects leukemia cells from apoptosis in vitro.* Molecular cancer, 2007. **6**: p. 31-31.
29. Crowe, B.L., et al., *Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes.* Nucleic Acids Research, 2013. **41**(6): p. 3924-3936.
30. van Zundert, G.C.P., et al., *The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.* Journal of Molecular Biology, 2016. **428**(4): p. 720-725.
31. Wassenaar, T.A., et al., *WeNMR: Structural Biology on the Grid.* Journal of Grid Computing, 2012. **10**(4): p. 743-767.
32. Ubbink, M., *The courtship of proteins: Understanding the encounter complex.* FEBS Letters, 2009. **583**(7): p. 1060-1066.
33. Bashir, Q., et al., *Visualization of the Encounter Ensemble of the Transient Electron Transfer Complex of Cytochrome c and Cytochrome c Peroxidase.* Journal of the American Chemical Society, 2010. **132**(1): p. 241-247.
34. Worrall, J.A.R., et al., *Transient Protein Interactions Studied by NMR Spectroscopy: The Case of Cytochrome c and Adrenodoxin.* Biochemistry, 2003. **42**(23): p. 7068-7076.
35. Schilder, J. and M. Ubbink, *Formation of transient protein complexes.* Current Opinion in Structural Biology, 2013. **23**(6): p. 911-918.
36. Clore, G.M. and C.D. Schwieters, *Docking of Protein−Protein Complexes on the Basis of Highly Ambiguous Intermolecular Distance Restraints Derived from 1HN/15N Chemical Shift Mapping and Backbone 15N−1H Residual Dipolar Couplings Using*

Conjoined Rigid Body/Torsion Angle Dynamics. Journal of the American Chemical Society, 2003. **125**(10): p. 2902-2912.

37. Trellet, M., A.S.J. Melquiond, and A.M.J.J. Bonvin, *A Unified Conformational Selection and Induced Fit Approach to Protein-Peptide Docking*. PLOS ONE, 2013. **8**(3): p. e58769.

SUPPORTING INFORMATION

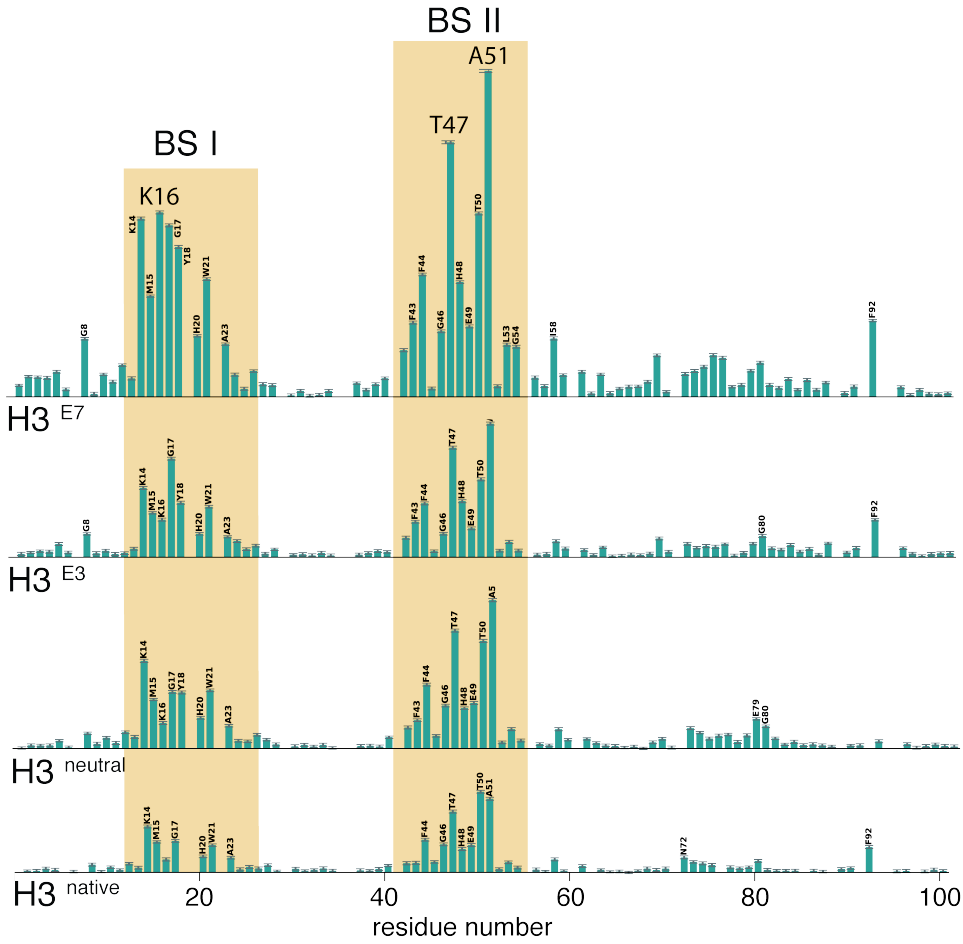


Figure S3-1. Chemical shift perturbations for H3 model peptides.

Chemical shift perturbations of the linear H3 peptide models. Two main binding sequences (BS I and BS II) are conserved in all peptides (yellow).

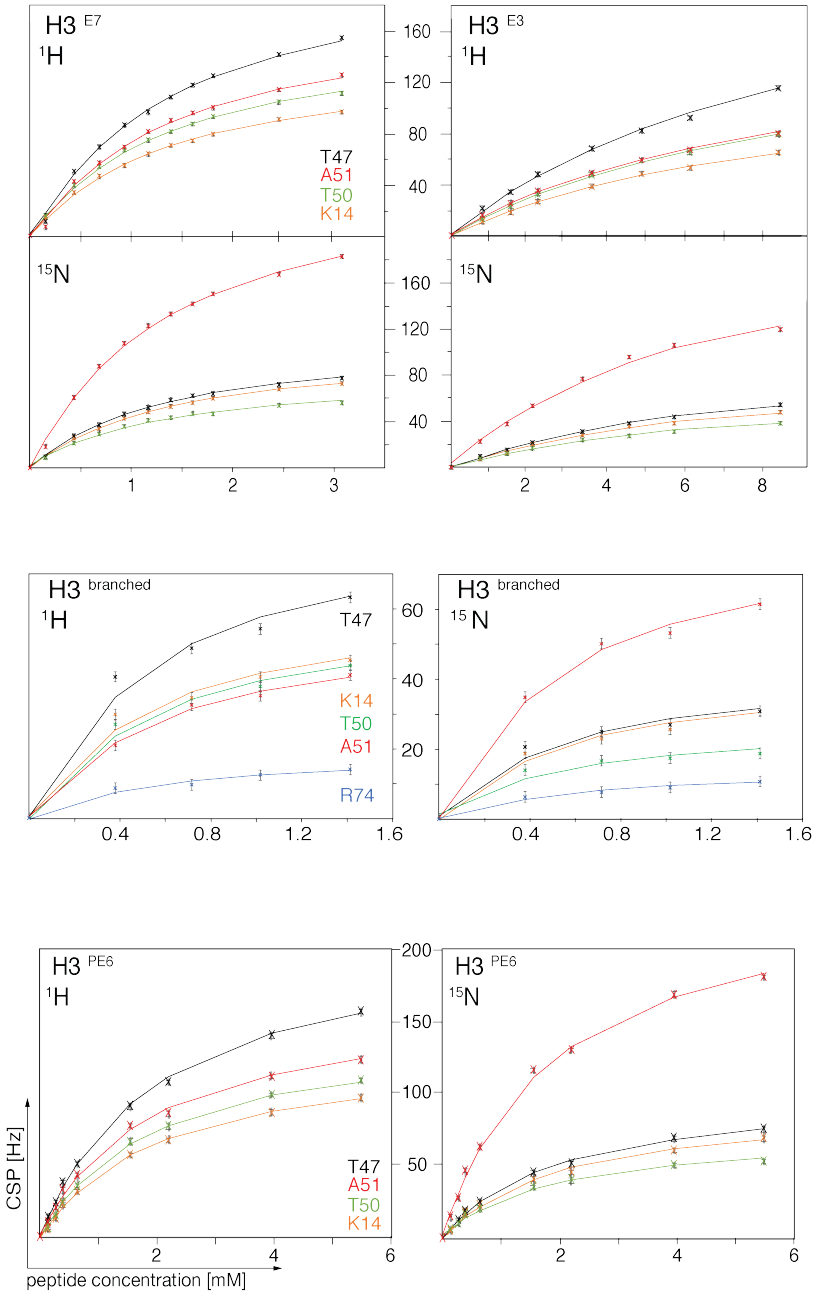


Figure S3-2. Fit data for reporter residues for H3E3, H3E7 and H3PE6.

Titration data for the reporter residues K14, T47, T50 and A51 for H3E7 (Top, left), H3E3 (Top, right) and H3PE6 (Bottom). Data points are shown as cross marks and the best fit is depicted as solid line. Color-coding annotated in figure.

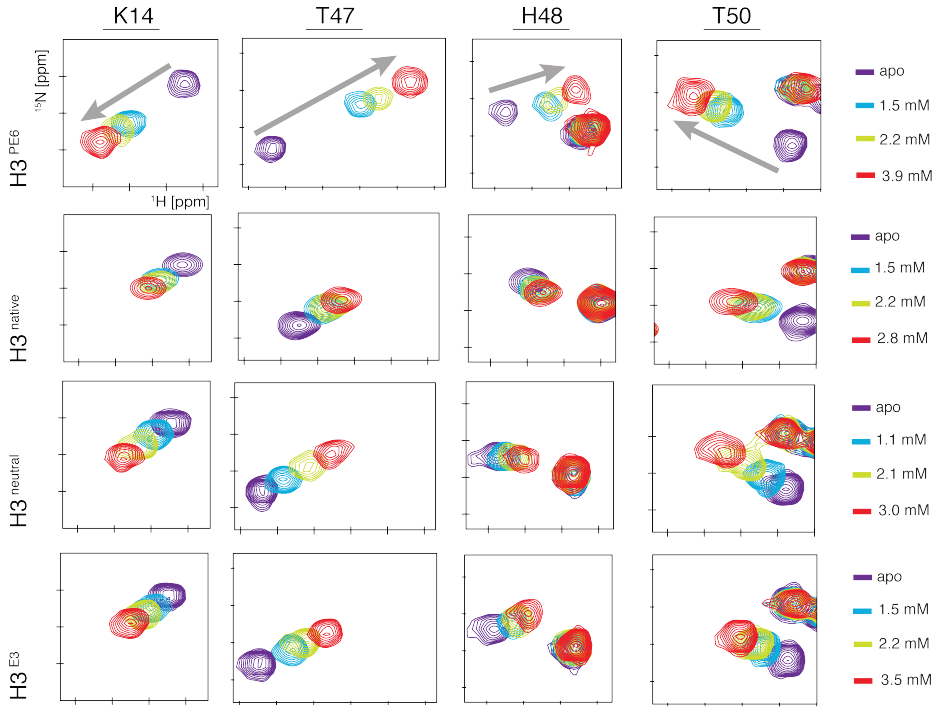


Figure S3-3. Shift trajectory for H3 tail peptide models.

CSP directionality for H3PE6 compared to the behavior of the same reporter residues in case of H3native, H3neutral and H3E3. All residues show a similar shift direction which indicates a similar bound-state complex structure. Color-coding and titration points indicated in figure.

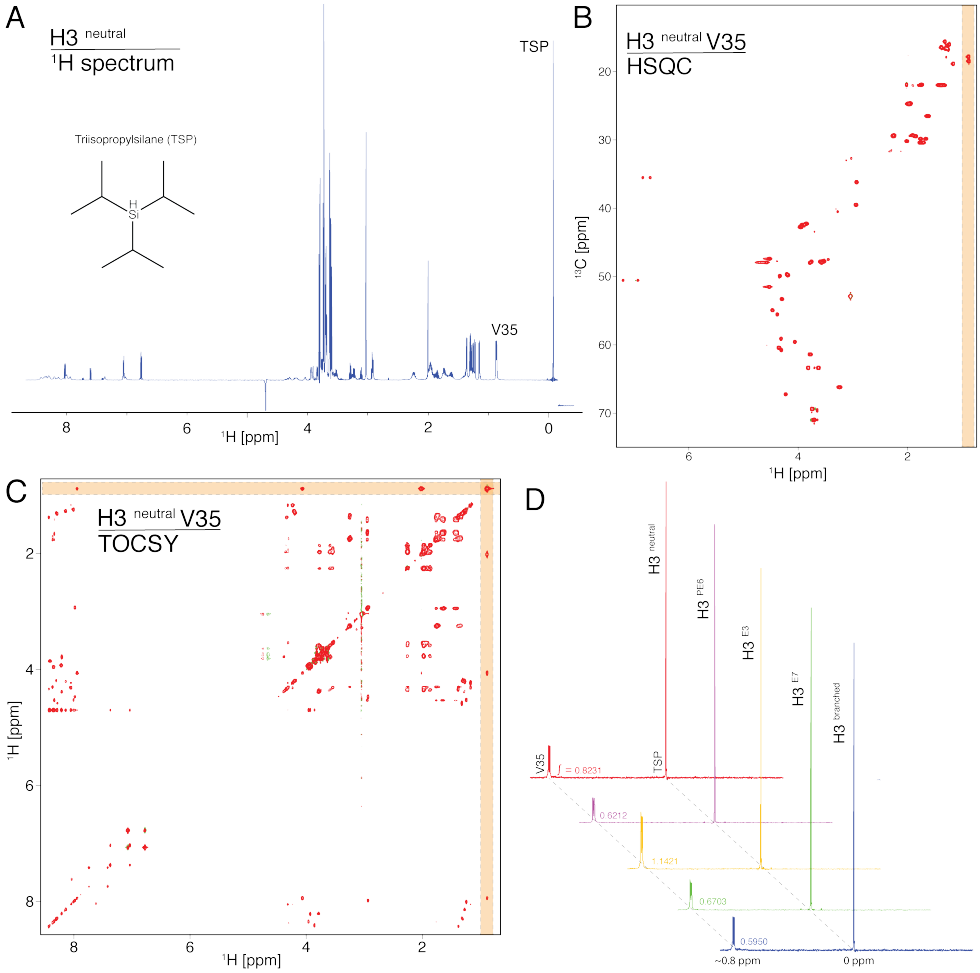


Figure S3-4. Concentration determination of peptide stocks.

(A) Full ^{1}H-1D spectrum of H3neutral with peaks for Val35 and the added reference compound TSP indicated. (B) ^{13}C-HSQC and (C) TOCSY of H3neutral with the chemical shift for Val35 indicated (orange). Both confirm peak as Val35. (D) Peaks for Val35 for all H3 peptide samples that were used for concentration determination. Integrals for TSP were set to 1 and the corresponding Val35 integrals are annotated in figure.

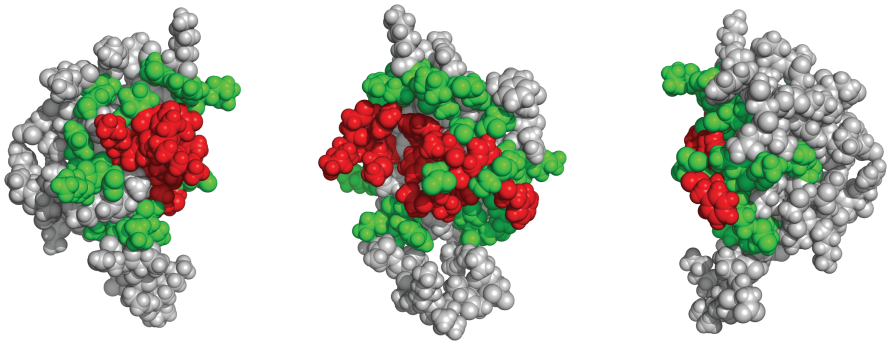
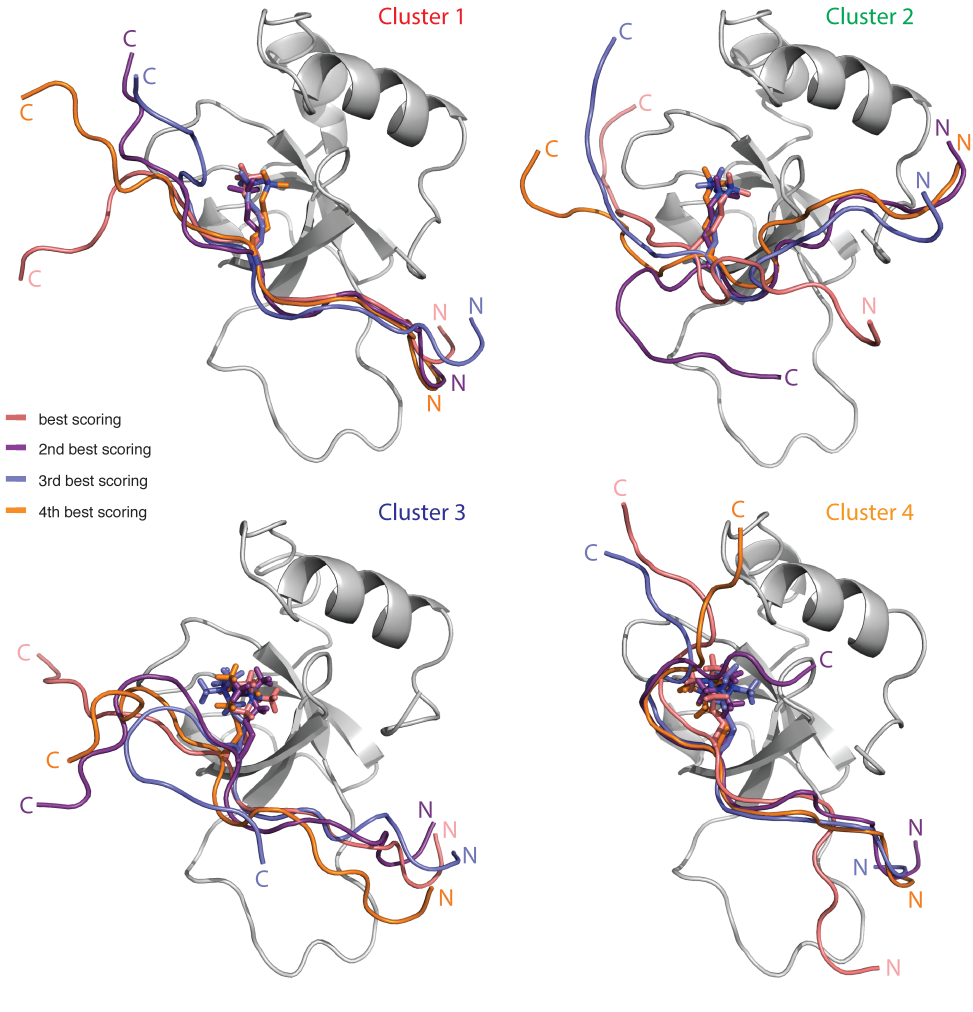


Figure S3-5. Input for HADDOCK modelling of $H3^{E7}$ complex with PSIP1-PWWP.

Representation of PSIP1-PWWP with active residues indicated in red and surrounding surface accessible passive residues colored green.



	Cluster 1	Cluster 4	Cluster 2	Cluster 3
HADDOCK score	-124.2 +/- 4.4	-116.5 +/- 10.2	-112.7 +/- 11.8	-80.5 +/- 14.6
Cluster size	172	5	14	7
RMSD from the overall lowest-energy structure	5.4 +/- 0.5	6.2 +/- 0.3	2.9 +/- 2.0	5.5 +/- 0.6
Van der Waals energy	-57.0 +/- 9.8	-64.6 +/- 3.9	-47.7 +/- 6.1	-40.3 +/- 6.8
Electrostatic energy	-373.0 +/- 60.2	-327.3 +/- 73.8	-413.1 +/- 115.1	-243.0 +/- 35.6
Desolvation energy	1.9 +/- 4.3	6.3 +/- 4.5	10.3 +/- 8.9	1.2 +/- 7.8
Restraints violation energy	55.3 +/- 32.62	72.4 +/- 52.17	72.8 +/- 35.88	72.3 +/- 30.06
Buried Surface Area	1568.6 +/- 97.8	1764.7 +/- 84.7	1507.2 +/- 67.6	1231.0 +/- 174.4
Z-Score	-0.9	-0.5	-0.3	1.7

Figure S3-6. Docking results for H3^{E7} and PSIP1-PWWP.

Representation of all clusters resulting from the docking of H3E7 and PSIP1-PWWP (Top). All four clusters are displayed with the corresponding first four structures. K36me3 is displayed as sticks. C-terminus (poly-Glu) and N-terminus (native H3 sequence) are indicated. Docking statistics are shown for all clusters in their scoring order (Bottom).

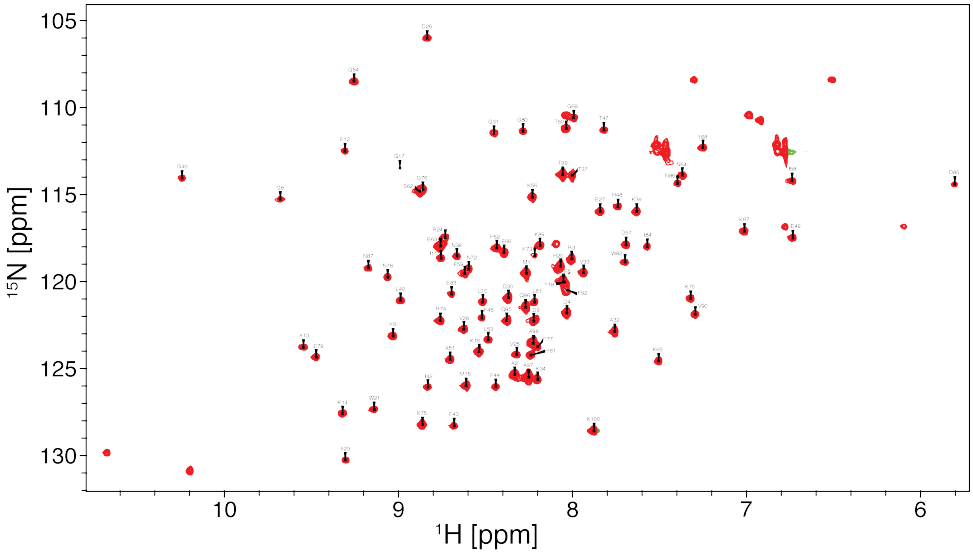


Figure S3-7. Assignment of PSIP1-PWWP.

Full 15N HSQC spectrum of PSIP1-PWWP (AA 2-100) showing the backbone amide signals used to follow chemical shift perturbations throughout the presented NMR titration experiments. ~87% of residues are assigned with the corresponding residues annotated in the spectrum (excluding side-chain hydrogen signals).

4

A Novel Screening Approach for Peptides Targeting the Trimethyllysine Recognition Domain of PSIP1

Based on research published as: I. Corbeski, **V. Horn**, R. A. van der Valk, U. B. Le Paige, R. T. Dame & H. van Ingen (2018). Microscale Thermophoresis Analysis of Chromatin Interactions.

Methods Mol Biol. 1837, 177 - 197

Synthesis and purification of the compounds $H3C^{E7}$ and $H3K_C^{E7}$ was performed in collaboration with N. J. Meeuwenoord at the Bio-organic synthesis group, Leiden University, Leiden (The Netherlands)

Abstract

An increasing number of therapeutic substrates address pathologies connected to post translational modifications of histone proteins, that together with DNA form nucleosomes which is the fundamental unit of chromatin. A promising approach herein is the interference of inhibitors with the readout of histone modifications by epigenetic reader proteins. These effector proteins often recognize their nucleosomal substrate by a synergetic engagement of multiple binding epitopes, prohibiting binding to isolated histone peptide models. We recently showed for one such protein, PSIP1, which binds nucleosomes with trimethylated lysine, that incorporation of both trimethyllysine and DNA mimicking in histone H3 derived peptide models significantly increases their binding affinity towards its PWWP reader domain. This observation may be used as a new rationale to potentially facilitate the development of a potent peptide-based inhibitor. Here, we set out to develop a workflow that allows for a more efficient screening of histone H3 tail-based peptide candidates. We report a protocol to efficiently substitute the native trimethyllysine residue for different side chain structures that combined with a microscale thermophoresis assay allows for fast identification of lead structures with low sample consumption. Through validation of the approach with NMR spectroscopy we find that fluorescence quenching upon binding reports specifically on aromatic cage binding and can be analyzed by simple fluorescence readout.

INTRODUCTION

DNA templated processes are for a large part regulated and orchestrated by the epigenetic landscape of chromatin, in the form of chemical modifications of its fundamental building block, the nucleosome. In contrast to the largely invariant DNA sequence, the histones proteins that make up the nucleosome can be decorated with a vast array of post translational modifications. These are maintained in a dynamic nature through epigenetic effector proteins that install (writers) or remove (erasers) them and are recognized by dedicated proteins (readers) [1-3]. Epigenetic effector proteins have been associated with a wide range of pathologies that make them a promising therapeutic target for small molecule and peptide-based inhibitors [4-7]. However, while there are significant advances reported for the inhibition of bromo domains that function as readers for acetylated side chains, the targeting of methylated lysine is still an intrinsically challenging task [8]. This is due to the binding motif of methyllysine, the aromatic cage. Especially the trimethylated state is specific to aromatic cage architectures with a shallow grove instead of a deep binding cavity as is the case for monomethyllysine [9]. This imposes a problem for the design of small molecule inhibitors that often rely on the insertion of structural moieties into a deep cavity. Therefore, we suggested that the inhibition of trimethyllysine might rather resemble a case of protein-protein interface (PPI) inhibition due to the extended interaction surface. The gap from small molecule cavity binding to extended interaction surfaces can be tackled by mid-molecular weight, peptide-based inhibitors [10]. Problematic herein is that several trimethyllysine reader proteins are reported to possess a very low binding affinity to peptide models of their nucleosomal substrate, the most severe being the PWWP domain of PSIP1. Due to its described role in many pathologies, PSIP1 is however a promising drug target [11-15]. To address this problem, we designed a peptide model of the H3 N-terminal tail that intrinsically has a very low binding affinity (~17 mM [16]) and added negative charges in order to use the synergetic binding of electrostatic charges and the aromatic cage for a two-epitope binding. Our study presented in chapter 3 resulted in a branched peptide structure where the native H3 tail sequence is combined with an orthogonal palindromic negatively charged backbone. We were able to show that we can increase the affinity of the PSIP1-PWWP peptide complex significantly. This is hypothesized to be due to additional electrostatic interactions formed between the peptide backbone and the positive protein surface, orienting the H3 tail for aromatic cage binding. However, with an obtained K_D of 350 μM this candidate peptide does not qualify as an inhibitor yet [17]. Nevertheless, it provides a lead structure for further screening efforts. A promising approach therein is

to increase the binding affinity of H3 tail peptides towards the aromatic cage. Exchange of the native trimethylammonium group for other functional groups holds the potential to boost binding affinity. Here, we introduce a synthesis protocol to incorporate bromoalkyl derived functional groups as thioester-based trimethyllysine substitutes and set up a microscale thermophoresis (MST) assay that allows for fast screening of compounds. In addition, we apply NMR spectroscopy to validate the assay and gain in-depth insights into binding modes and nature of the modifiable peptide platform.

Principles of MST Affinity Measurements

As a fast assay for the measurement of binding affinities with low sample consumption, MST relies on the thermophoretic effect. Analogous to electrophoresis, thermophoresis is the flow or directed movement of molecules along a temperature gradient [18]. Technological advances have made it possible to use small temperature gradients (typically 2–6 °C) and detect the resulting micrometer-scale movements, allowing the application in molecular biology as microscale thermophoresis (MST) [19]. MST typically requires fluorescent labeling of the protein of interest for high sensitivity of detection. Samples are loaded onto glass capillaries and a specific spot is heated by an infra-red laser (see Figure 4-1A). The resulting temperature gradient causes thermophoresis of the labeled molecule, typically away from the heated spot, which is observed through a decrease in fluorescence in the heated region (see Figure 4-1B). Since thermophoresis is sensitive to molecular size, charge, and hydration shell [20], changes in these properties due to binding will cause changes in thermophoresis. For analysis of interactions, MST curves are recorded as titration series with increasing amounts of ligand and normalized with respect to their initial equilibrium fluorescence (F_{norm}) (Figure 4-1B). A binding curve is extracted by plotting the F_{norm} values at the end of the laser on –period (phase IV, Figure 4-1B), which captures binding induced changes both in thermophoresis and in the intrinsic temperature dependence of the fluorophore (indicated by the arrow "Thermophoresis + T-Jump" in Figure 4-1B). Both effects can also be analyzed separately. Since the observed F_{norm} values are the population weighted averages of the unbound and bound molecules, standard methods can be used to fit the binding curve and extract the binding affinity. Furthermore, MST can be performed in virtually every buffer [21] and other characteristics such as thermodynamic properties, binding stoichiometry, and enzyme kinetics can be extracted with customized experimental designs [22]. Due to the intrinsic advantages of MST for the measurement of binding affinities (namely the low sample consumption and fast measurement and preparation times) we set out to implement this technique for the screening of potential PSIP1-PWWP binders.

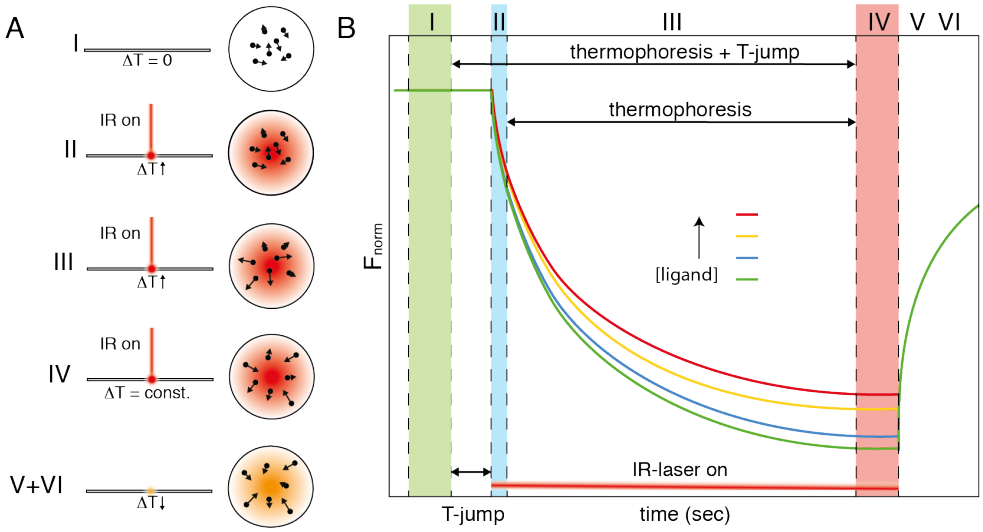


Figure 4-1. Biophysical basis of microscale thermophoresis in affinity assays. [23]

Schematic depiction of the MST experiment (A) and MST time trace (B). In (A), the capillaries (left) and particles in a cross-section thereof (right) during different stages of the experiment (I–VI) are shown. Starting from equilibrium (phase I), laser irradiation is started (phase II), causing particles to move out or into the heated volume (thermophoresis, phase III). When thermophoresis is counter balanced by mass diffusion a steady state is reached (phase IV). When the laser is switched off, the particle concentration re-equilibrates (phase VI). In (B), a four-sample titration series is shown, ligand concentration indicated. Each sample is characterized by a distinct thermophoresis curve. The rapid change in normalized fluorescence in phases II and V is caused by the temperature dependence in fluorescence (T-jump). Binding curves can be extracted by plotting F_{norm} values for regions "Thermophoresis + T-jump" (I vs. IV), Thermophoresis (start vs. end of III), or T-jump (I vs. II).

RESULTS

FLUORESCENT DYES INTERACT WITH THE AROMATIC CAGE OF PSIP1-PWWP

The setup of an MST assay for the investigation of the binding affinity between H3 tail peptide models and PSIP1-PWWP requires one binding partner to be labelled with a fluorescent tag. An attractive approach is to use peptide chemistry methods to attach the fluorescence dye 5,6-carboxyfluoresceine (6-FAM) at the end of the sequence (Figure 4-2). This has the advantage that the approach is cheap, can readily be incorporated in the synthesis protocol and an on-resin labelling allows for the

efficient removal of unreacted dye. Because of the low affinity of the studied interactions we had to verify that the fluorescent tag does not engage in binding to the protein since this could compete with peptide binding and potentially interfere with the measurement.

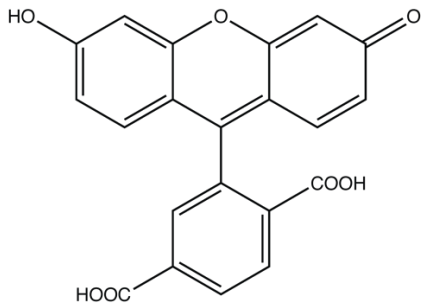


Figure 4-2. Structural formula of 6-FAM.

Depiction of the structural formula of the fluorescence label 5,6-carboxyfluoresceine (6-FAM) used for labelling of the peptide substrates prior to MST.

This is of particular importance since we showed that the binding of H3 tail peptide models is largely driven by electrostatic interactions. While a test like this can be carried out by conducting an MST study with the fluorescent dye as binding partner for PSIP1-PWWP, we wanted to gain a more complete picture of a possible interaction. Therefore, we used NMR spectroscopy which is especially well suited to investigate binding processes on a structural level and map binding epitopes. We carried out HSQC titration experiments of ^{15}N labelled PSIP1-PWWP with increasing concentrations of 6-FAM added. Strikingly, we observe chemical shift perturbations (CSP) for specific residues in PSIP1-PWWP which indicates a binding of 6-FAM (Figure 4-3 A+B). The magnitude of CSPs is lower than for our lead peptide $H3^{branched}$ (see Chapter 3) but the three binding epitopes previously identified for H3 tail peptides (BS I-III) are clearly affected by addition of the fluorophore (Figure 4-3 C). We conclude from this observation that the labelling of H3 tail peptides is not orthogonal to PSIP1-PWWP complex formation and a clear separation of processes might not be possible. To exclude interference from the dye, the protein can be covalently labeled with the fluorophore instead. The manufacturer Nano Temper provides two labelling schemes optimized for MST use. Firstly, a dye with an amine-reactive NHS-ester group that selectively labels lysine residues, and secondly, a maleimide-based dye coupling to cysteine residues. Since we showed previously that lysine residues play an important role in the complexation of PSIP1-PWWP (see Chapter 3) we designed a cysteine mutant that allows for selective and specific PSIP1-PWWP labelling.

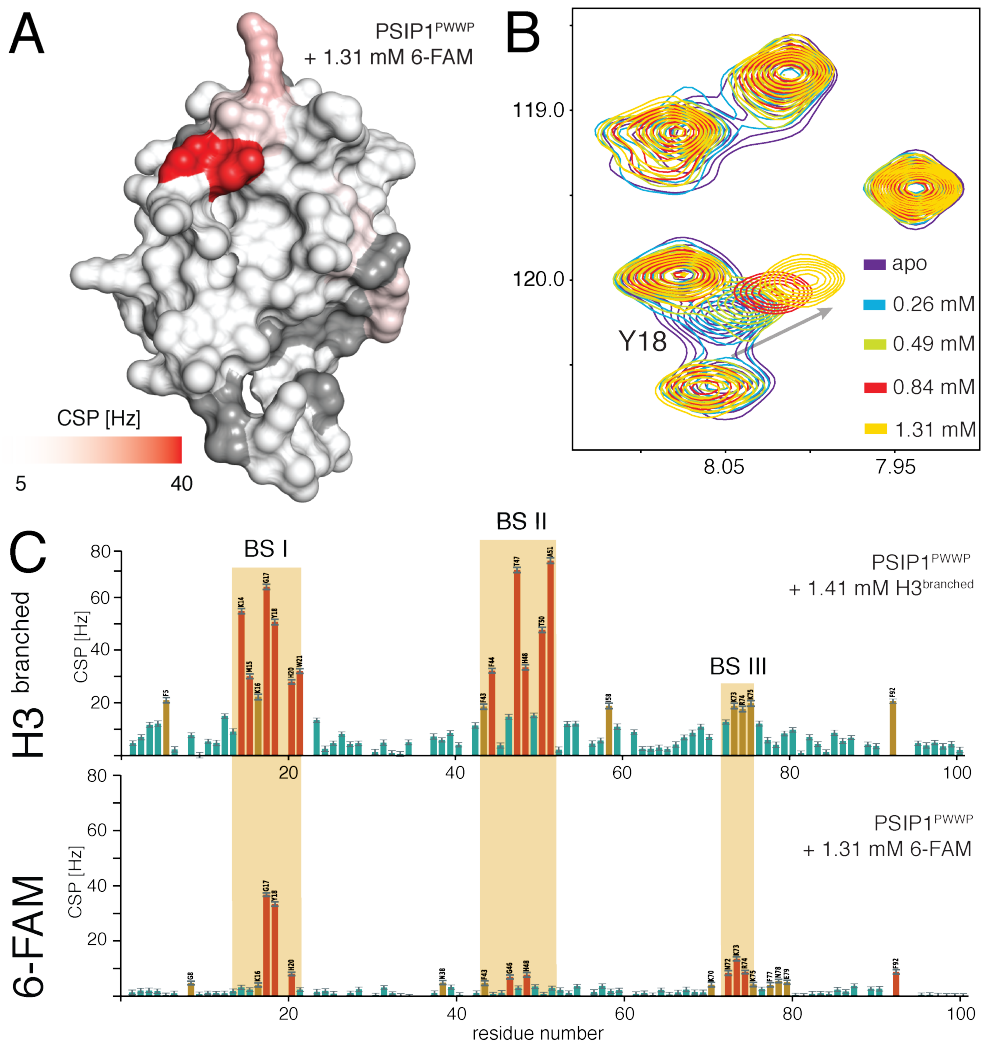


Figure 4-3. NMR titration experiment to map binding of 6-FAM to the PSIP1-PWWP aromatic cage.

(A) A depiction of observed chemical shift perturbations on the PSIP1-PWWP surface shows interactions in the region of the aromatic cage. Especially, residue Y18 shows interactions with 6-FAM. (B) Zoom on the NMR signal for Y18 showing a peak displacement in dependency of addition of increasing concentrations of 6-FAM. Color-coding and concentrations indicated in figure. (C) Comparison of binding sequences for 6-FAM and $H3^{branched}$. The two binding sequences BS I and BS III occur for 6-FAM as well while only minor peak displacement is observed for BS II.

Introduction of a Cysteine Reaction Site Allows for Selective Labelling of PSIP1-PWWP

As a mutation site to introduce a cysteine residue, allowing for selective coupling of a fluorescent dye, we selected residue N86 due to its location on the opposite side of the aromatic cage (Figure 4-4 A). Though we cannot exclude the possibility of intermolecular binding of the label to the aromatic cage of another PSIP1-PWWP molecule, the very low concentrations of labelled binding partner in MST experiments makes productive encounters insignificant, especially due to the low affinity observed for dye/cage interactions (Figure 4-3). Another advantage of this setup arises when compared to a labelled peptide strategy. There the peptide concentration would be very low while the PSIP1-PWWP concentration would have to be approximately 20-times the K_D value to yield an ideal binding curve. However, this limits the applicability to medium and high affinity peptide compounds and excludes low affinity binders due to limitations in solubility of the protein or effects on the viscosity that in return influence the diffusion behavior (and subsequently the MST measurement) unrelated to binding.

After expression and purification of the N86C mutant construct, we assured the integrity of the mutation and proper native-like folding of PSIP1-PWWP N86C by NMR spectroscopy. From the ^{15}N-HSQC spectra of *wt*PSIP1-PWWP and PSIP1-PWWP N86C it can be seen, that the majority of residues is unaffected by the mutation (Figure 4-4 B). When focusing on the mutated residue N86 we see that the corresponding peak disappeared which is in line with the mutation of the residue. Furthermore, we see two peaks disappear that are attributed to Asn side chain amino group protons (Figure 4-4 C). Additionally, several peaks experience chemical shift perturbations in the mutant construct (Figure 4-4 D). While in some cases there is a certain ambiguity of which effected signal belongs to which native signal, it can be concluded that all residues in question are in the direct surrounding of the expected mutation site N86C. We next labelled PSIP1-PWWP N86C with maleimide reactive dye (Fluorophore NT-647) and tested for fluorescence readout (Figure 4-4 E). A scan over the diameter of the capillary shows sufficient fluorescence signal and a smooth signal which excludes sticking of the substrate to the glass capillary.

We therefore conclude that the mutant protein is folded in a native conformation, can efficiently be labeled and behaves well in MST capillaries and is thus suitable for the use in MST assays.

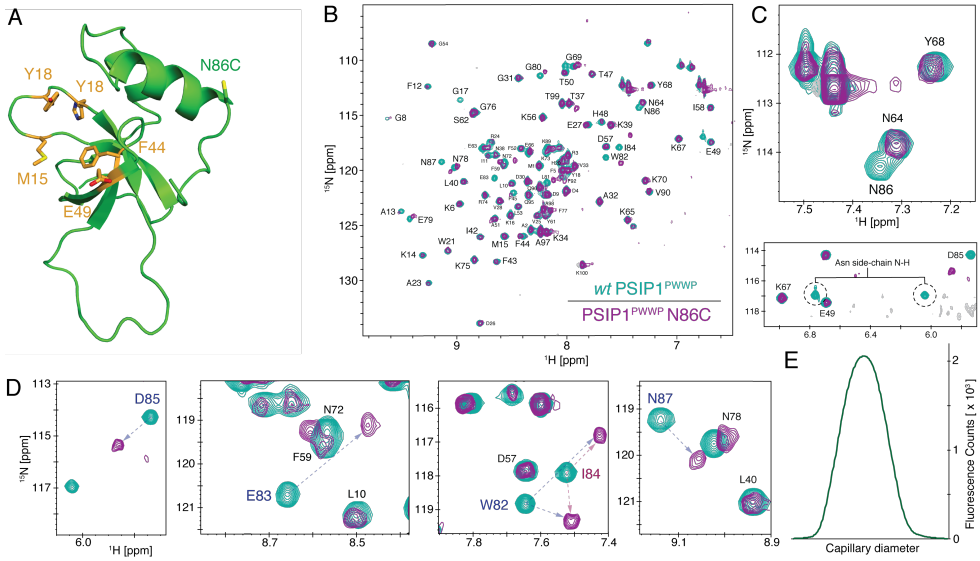


Figure 4-4. Structural analysis of PSIP1-PWWP N86C by NMR spectroscopy.

(A) Structure of N86C PSIP1-PWWP derived from *wt*PSIP1-PWWP [16]. Residues that form the aromatic cage moiety (orange) and the introduced N86C mutation are depicted as sticks. (B) Overlay of the HSQC spectra of PSIP1-PWWP N86C (purple) and *wt*PSIP1-PWWP (cyan). Large similarities in peak positions ensure a proper fold of the mutant construct. (C) Missing signals for the N86 backbone amide and two side chain protons confirm the mutation. (D) Specific displacement of peaks for residues that are situated in the direct sequence environment of N86C additionally confirms the integrity of the mutation. (E) A capillary scan for PSIP1-PWWP N86C shows a symmetrical peak with sufficient fluorescence intensity for application in MST experiments.

A SYNTHESIS PROTOCOL FOR FAST ADAPTION OF H3 TAIL LEAD COMPOUNDS

After the setup of a labelling scheme for an MST assay we sought to further streamline the screening for H3 tail-based PSIP1-PWWP binders. It was shown, that binding to the aromatic cage can be addressed with other functional groups than just the trimethylammonium of trimethyllysine [9]. We therefore resorted to H3 tail peptides with a Cys residue at position 36 as a platform to explore diverse functional side chain groups as substitutes for the native trimethyllysine. This setup allows the usage of a monomethoxytrityl (Mmt) side chain protection group that can be selectively cleaved and allows for an isolated on-resin reaction site during solid phase peptide synthesis (SPPS). The free thiol group can then be used for alkylation reactions with bromo-alkyl compounds. This modification was previously shown to be applicable to generate site specific trimethyllysine analogue carrying histone proteins [24]. However, the described protocol is limited to aqueous buffers with high excess of (2-bromoethyl)

trimethylammonium bromide. Furthermore, the aqueous conditions prohibit the use of water insoluble halo-alkyls. We therefore present a modified alkylation protocol that is compatible with on-resin modifications (Figure 4-5 and detailed in M&M). Importantly, the modified DMF-based protocol relies on Cs_2CO_3 as base to stabilize the deprotonated state of the cysteine thiol group instead of basic alkylation buffer (pH 7.8) for the original water-based protocol [24]. A further key difference here is that the use of 16 Eq. instead of 40 Eq. (2-bromoethyl) trimethylammonium bromide proved sufficient for quantitative peptide alkylation.

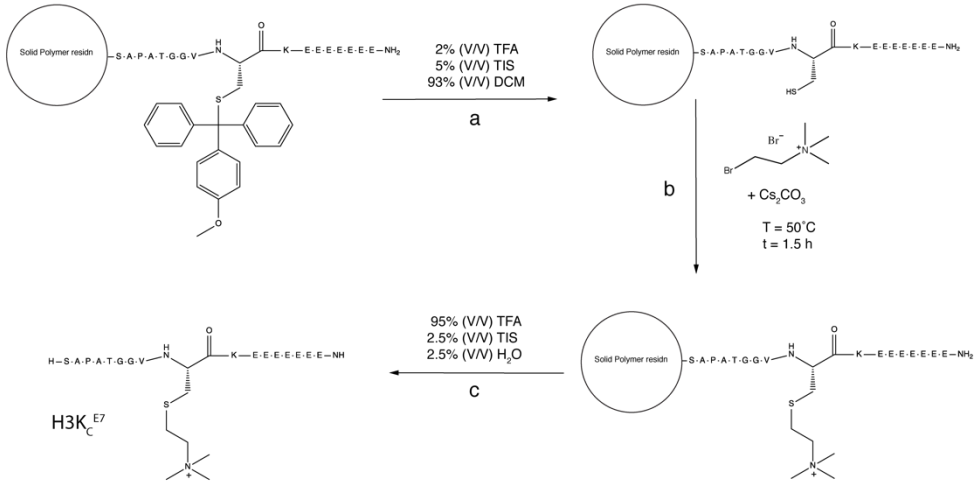


Figure 4-5. Reaction scheme for the selective on-resin alkylation of cysteine in a H3C^{E7} peptide.

Route for the selective removal of the Mmt side chain protection group from cysteine in resin-bound peptide (a). Shaking of resin-bound peptide at 50°C for 1.5 h together with (2-bromoethyl) trimethylammonium bromide (16 Eq.) and Cs_2CO_3 (16 Eq.) in DMF results in fully alkylated product (b). Final cleavage is performed under acidic conditions with a trifluoracetic acid (TFA)/ triisopropylsilane (TIS)/ dichlormethane (DCM) solution (c).

To test the applicability of the protocol we synthesized the H3 tail peptide H3^{E7}, that we introduced in chapter 3, in an alkylated version (H3K$_C^{E7}$) (Figure 4-5). Indeed, we observe an efficient alkylation of a cysteine carrying H3^{E7} analogue. Though a suitability of the protocol for a wider range of halo-alkyl derivatives is still to be shown, we suggest that the described reaction can be used to choose for different functional side chain groups to optimize for aromatic cage binding of H3 tail peptides.

Binding Affinity of $H3K_C^{E7}$ Derived from an MST-based Assay

We next carried out MST experiments using a serial dilution of $H3K_C^{E7}$ and a constant concentration of PSIP1-PWWP. Interestingly, we observe a concentration dependent increase of fluorescence throughout the dilution series. This fluorescence quenching can be caused by complex formation. To test this hypothesis, the MST assay was repeated using $H3K_C^{E7}$ without prior alkylation ($H3C^{E7}$). There no changes in fluorescence were observed (Figure 4-6 A). In addition, we heated the samples with labelled PSIP1-PWWP and alkylated $H3K_C^{E7}$ in SDS-buffer. Protein denaturing caused the initial fluorescence for all peptide concentrations to be equal. We thus conclude that the fluorescence quenching is caused by binding of the peptide and that the quenching is dependent on the presence of the trimethyllysine analogue. We fitted the data for initial fluorescence, as well as T-jump and Thermophoresis + T-Jump to extract the binding affinity. Interestingly, all three fits arrive at a comparable value for the affinities for $H3K_C^{E7}$, ranging from 1.4 mM to 1.9 mM (Figure 4-6 B-D). These values are slightly decreased when compared to $H3^{E7}$ (1.2 mM) as determined by NMR (See chapter 3). However, the good agreement between data observed for the native $H3^{E7}$ peptide substrate and the introduced screening assay show that it can be used for the screening of peptide-based leads for PSIP1-PWWP inhibition.

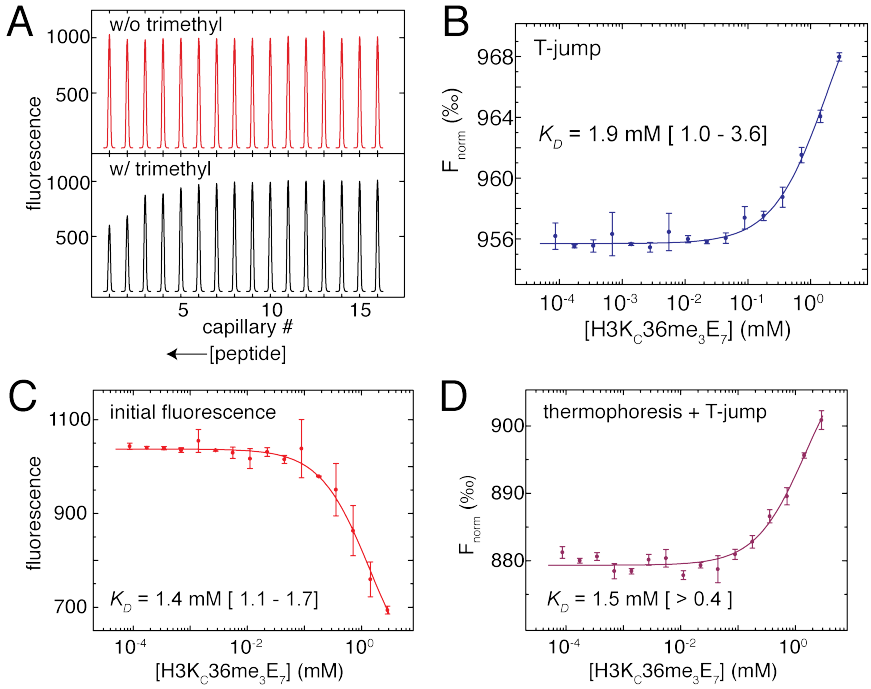


Figure 4-6. MST of nucleosome-mimicking H3K$_C^{E7}$ binding to the PSIP1-PWWP domain [23].

(A) Capillary scans for H3K$_C^{E7}$ (SAPATGGVCKEEEEEEE) with (black) and without (red) cysteine-based trimethyllysine analogue. (B, C, D) Analysis of the MST measurements was done by fitting the data for T-jump (B), initial fluorescence (C) and Thermophoresis + T-jump (D). Best-fit values for the K_D and 95% confidence limits are indicated in the plots.

CYSTEINE-BASED KME3 ANALOGUES IMPAIR AROMATIC CAGE BINDING

To validate the finding from the MST assay and further investigate the structural impact of using a cysteine-based Kme3 mimic, we conducted NMR titration experiments. First, we conducted an NMR titration experiment of H3C^{E7} together with *wt*PSIP1-PWWP to see whether the absence of fluorescence quenching in the MST assay is due to a lack of interaction. In stark contrast to H3^{E7} (see Figure 3-3 in Chapter 3) we observe only minor chemical shift perturbations, even up to a final peptide concentration of 5.97 mM (Figure 4-7 A-C). Interestingly, a global fit for the observed CSP data results in a K_D of 1.5 mM (95% prob. 1.27 – 2.10) (Figure 4-7 D). This shows that there is a complex formation also for H3C^{E7}, despite the absence of a binding related change in fluorescence in the MST experiment. This indicates that the fluorescence decrease is specifically reporting on aromatic cage binding and might be exploited for the further optimization of side chain functional groups.

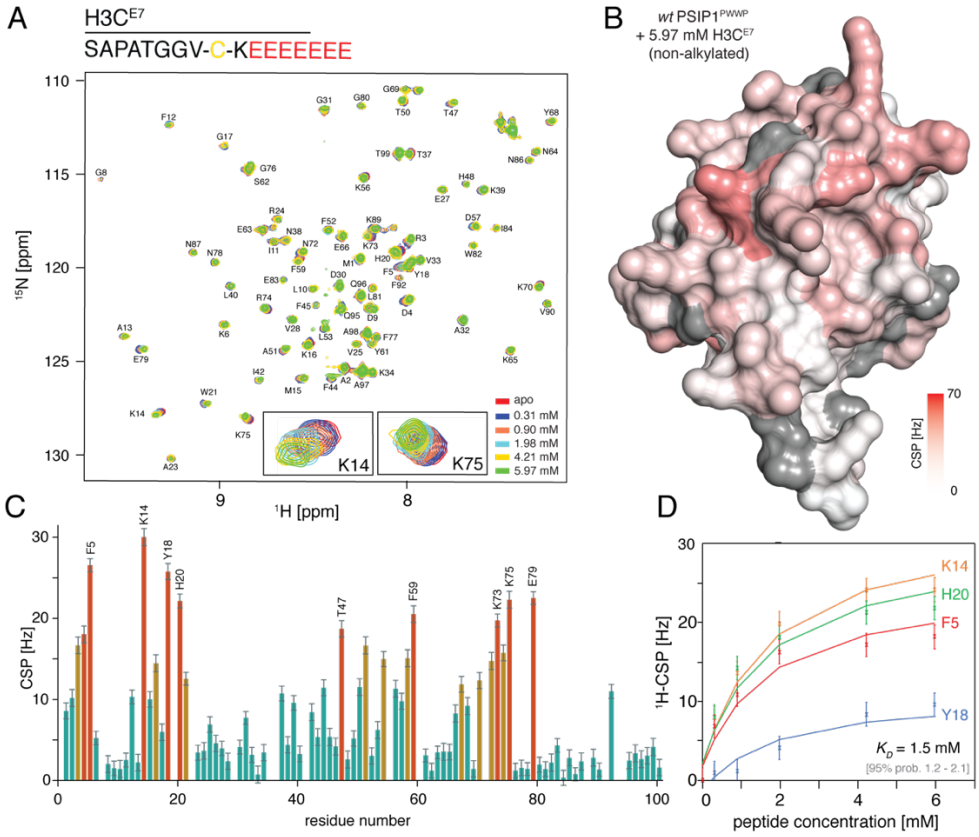


Figure 4-7. NMR titration study of H3C^{E7} binding to PSIP1-PWWP.

(A)Overlay of ^{15}N-HSQC spectra of PSIP1-PWWP with increasing concentrations of H3C^{E7} added up to a final concentration of ~6 mM. (B) Observed chemical shift perturbations depicted on the surface representation of PSIP1-PWWP are distributed over the positively charged DNA binding interface. (C) The unspecific binding of H3C^{E7} is further highlighted by the lack of prominent CSPs for aromatic cage binding reporter residues in BS I and BS II. Furthermore, the total extent of observed CSPs at ~6 mM is limited to ~30 Hz. (D) Selected examples of the global fits for the major shifting residues. Color-coding annotated in figure.

We next titrated H3K$_C$C^{E7} to a final concentration of 3.02 mM. We observed specific CSPs for several residues that we established previously as reporter residues for H3 tail peptide binding (Figure 4-8 A). The shifting residues are arranged in and around the aromatic cage which indicates a conserved binding mode compared to H3^{E7} (Figure 4-8 B). However, we also observed that besides the binding sequences (BS) BS I and BS

II $H3K_C^{E7}$ also shows engagement of BS III (Figure 4-8 C). Fitting of the CSP derived binding curves results in a K_D for $H3K_C^{E7}$ of 0.8 mM (0.7-1.2 mM 95% probability limits) (Figure 4-8 D) compared to 1.2 mM for $H3^{E7}$ (see Chapter 3). The fits further allow to compare the binding modes of $H3K_C^{E7}$ and $H3^{E7}$ based on the fitted bound-state chemical shifts in the respective titrations. Here, we focused on residues in and directly around the aromatic cage that report on binding of the trimethyllysine or its analogue and the formation of the intermolecular beta-sheet. Strikingly, the two datasets show significantly different chemical shifts for the bound-state of the protein for these residues. This includes the bound-state peak position of residues A51 and K16 that are significantly and uniformly decreased to ~ 16 % with respect to the native peptide. A very similar reduction was observed for $H3^{branched}$ (Figure 4-8 F). This points towards a binding mode that is comparable to $H3^{branched}$ even though $H3K_C^{E7}$ is structurally most similar to $H3^{E7}$. This is also indicated by the pronounced CSPs of positive residues K73, R74 and K75 in the $H3K_C^{E7}$ titration. These residues in BSIII, were not affected by $H3^{E7}$ binding, but were significantly affected in titration with $H3^{branched}$. Thus, the data suggest that both in term of the interaction to the aromatic cage and the positively charged PWWP surface, $H3K_C^{E7}$ resembles the binding mode $H3^{branched}$ more than $H3^{E7}$.

We previously hypothesized that the extensive interaction with the DNA binding surface can cause reduced interaction with aromatic cage, thus decreasing in bound-state peak position for reporter residues of aromatic cage binding. In this model only a distinct fraction of the aromatic cage is occupied even under saturating conditions. In case of $H3^{branched}$ we hypothesized that the decreased aromatic cage binding is due to a lack in structural optimization to simultaneously engage both epitopes. In case of $H3K_C^{E7}$, aromatic cage binding could be similarly disfavored due to substitution of the native carbon by a sulfur that differs in size and electronegativity.

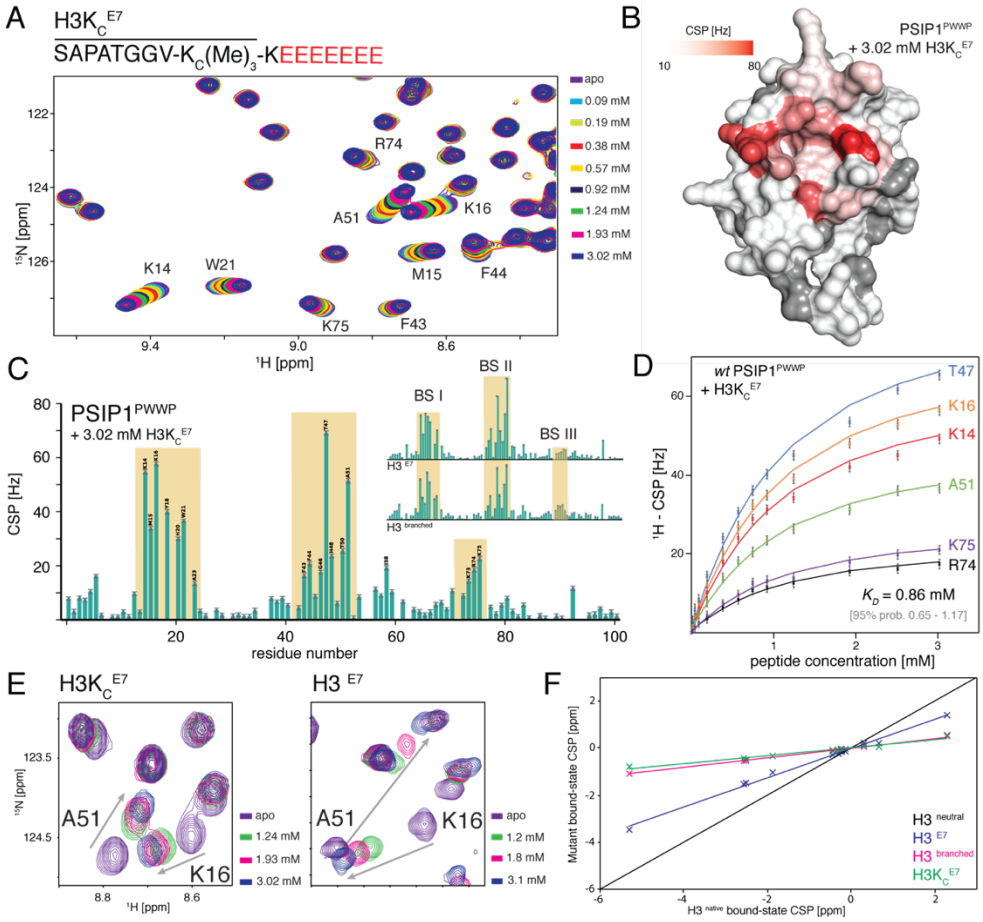


Figure 4-8. NMR Titration of H3K$_C^{E7}$ shows varied binding mode of an unacetylated N-terminus.

(A) Overlay of ^{15}N-HSQC spectra of PSIP1-PWWP with increasing concentrations of H3K$_C^{E7}$ added. CSPs can be observed for key residues reporting on aromatic cage binding. Color-coding and concentrations indicated in figure. (B) Depiction of CSPs on the structure of PSIP1-PWWP maps a binding site in and around the aromatic cage composed of BS I and BS II. In addition, BS III as previously observed in H3^{branched} is present. (C) While being an altered version of H3^{E7}, H3K$_C^{E7}$ shows all three binding sequences as observed for H3^{branched}. (D) Best-Fits for the respective two highest shifting residues for BS I – III upon H3C$_K^{E7}$ titration (solid lines, color-coding indicated in figure). K_D and 95% confidence interval stated in figure. (E) Close up on the displacement of residue A51 reporting on acidic patch binding. The total shift margin at comparable concentrations is significantly decreased between H3^{E7} (native Kme3) and H3C$_K^{E7}$ (cysteine-based analogue). This indicates a difference in bound-state structure of trimethyllysine. (F) Comparison of the free-bound chemical shift difference for binding H3^{native} and H3^{E7} (blue), H3^{branched}(pink) or H3K$_C^{E7}$ (green), based on best-fits of the titration data. Black line indicates perfect correlation.

Discussion

Here, we introduce a workflow for an efficient fluorescence-based screening assay for H3 tail-based peptide compounds as potential inhibitors of PSIP1-PWWP. Fluorophore 6-FAM was found to specifically interact with the PSIP1-PWWP domain, in or in the vicinity of aromatic cage. Similar direct involvement of fluorescent dyes has been reported and exploited for the fluorescence detection of protein binding [25, 26]. To avoid this interference, we used the N86C mutant protein for selective fluorescent labelling and application in MST-based assays. Due to the low sample consumption the assay is especially well suited for the study of low affinity interactions. This is of high relevance since we previously showed that insights into binding mechanisms for low affinity interactions can be used to derive the structural basis of potential inhibitors (Chapter 3). Furthermore, we transferred the idea of halo-ethylammonium derived trimethyllysine analogue incorporation to on-resin peptides during solid phase peptide synthesis. We demonstrate the methodology using a methyllysine analogue of the linear $H3^{E7}$ peptide that we studied previously by NMR spectroscopy. A slightly lower binding affinity as found for $H3K_C^{E7}$ in the MST assay (1.4/1.5/1.9 mM, vs 1.2 mM for NMR). While these differences are too small to be significant given the accuracies of the different methods, NMR experiments with $H3K_C^{E7}$ show a different structural binding mode. The bound-state shifts, resembling those of $H3^{branched}$ more so than $H3^{E7}$, point to reduced aromatic cage binding and increased specific interaction with the positive DNA binding surface around BSIII. Since the main difference between $H3^{E7}$ and $H3K_C^{E7}$ is the use of a thioether Kme_3 analogue, we suggest that the increased electronegativity of the sulfur together with the slightly altered side chain structure in the thioether mimic decrease the suitability of aromatic cage binding which in return allows for an efficient engagement of the positive residues in BS III. The resulting complex thus resembles that observed for $H3^{branched}$ where the increased electrostatic charges 'pull' the trimethyllysine out of the cage. Alternatively, the different acetylation state of the N-terminus could play a role in this change in binding mode. The substrate $H3K_C^{E7}$ possesses a free and positively charged N-terminus while $H3^{E7}$ is acetylated and neutral. While this charge is remote from the binding site, the influence of the N-terminal charge needs to be investigated more thoroughly before conclusive remarks about the nature of change in binding mode can be made.

Together, our data underscore the potential of optimizing the functional side chain group at position 36, as the thioether based peptide results in slightly increased overall binding affinity compared to the native trimethyllysine $H3^{E7}$ (0.8 mM vs. 1.2 mM). The MST assay is here of special use since we see that the fluorescence decrease that is

observed is dependent of the presence of trimethyllysine. We showed by NMR that H3C^{E7} is still able to engage in PSIP1-PWWP binding, which indicates that only binding of negative H3 tail peptides with aromatic cage engagement is reported. However, this also means that it is to be expected to also observe MST effects for substrates that do not engage in aromatic cage binding. Though this still has to be shown, it would make the additional use of microscale thermophoresis obsolete and we thus suggest focusing on the isolated readout of initial fluorescence which would allow for an even faster screening. A potential additional advantage is to allow for selection of compounds that have both high affinity and tight interaction with aromatic cage as evidenced by fluorescence quenching.

CONCLUSION

We here presented a screening setup that allows for fast screening of H3 tail peptide-based candidates for PSIP1-PWWP inhibition. The screening is based on a fluorescence assay, using labelled protein. The presented protocol for derivation of the H3 tail peptides based on alkylation of an introduced cysteine allows to substitute native trimethylammonium group for entities with different, more suitable characteristics to enforce binding to the trimethyllysine pocket. Our results so far highlight that for simultaneous engagement of both methyllysine pocket and the DNA binding surface is crucial to obtain high affinity engagement to PSIP1-PWWP. Overall, the setup is ideal for initial screenings that can be further verified by more accurate but time-consuming methodologies.

Materials and Methods

Protein Expression

For NMR studies, PSIP1-PWWP (AA 2-100, Uniprot-ID: O75475) and PSIP1-PWWP mutant N86C were expressed as a GST-fusion protein in BL21 Rosetta2 (DE3) cells (Novagen). PSIP1-PWWP constructs were produced in M9 minimal medium containing $^{15}NH_4Cl$. Cells were grown to an OD of 0.6 and induced with 500 µM IPTG. The protein was expressed for 3 h at 37 °C. The cells were harvested in lysis buffer (50 mM Tris pH 8.0, 300 mM KCl, 0.5 mM PMSF, 1 mM DTT, 3.3 mM TritonX-100, complete EDTA-free protease inhibitor (Roche) and lysed by French-press. The soluble fraction was collected after ultracentrifugation at 25000 rpm and 4 °C for 40 min, and subsequently loaded on a 5 mL GSTrap HP column (GE Healthcare). The column was washed with wash buffer I (50 mM Tris pH 8.0, 300 mM KCl, 1 mM DTT, complete EDTA-free protease inhibitor (Roche)) and wash buffer II (50 mM KPi pH 7.0, 100 mM KCl, 1 mM DTT, complete EDTA-free protease inhibitor (Roche)). The GST-fusion protein was eluted using elution buffer (50 mM KPi pH 7.0, 100 mM KCl, 50 mM reduced glutathione, 1 mM DTT, complete EDTA-free protease inhibitor (Roche)). Thrombin cleavage of the GST-tag was carried out for 3h at 37 °C after adding 10U of thrombin to the combined eluted fractions, then quenched by addition of 1 mM PMSF and incubation for 15 min at 37 °C. The PSIP1-PWWP domain was then purified using ion-exchange chromatography over a HiTrap SP HP column (GE Healthcare). A gradient from IEX buffer (20 mM KPi pH 7.0, 0.2 mM PMSF, 0.5 mM DTT) w/o KCl to IEX buffer with 1 M KCl was used to elute PSIP1-PWWP from the column. The PWWP-containing fractions were then combined and dialyzed to NMR buffer (20 mM NaPi (pH = 7.0), 150 mM NaCl, cOmplete protease inhibitor, 10 %(V/V) D_2O)

Mutagenesis

For the construct PSIP1-PWWP N86C, the point mutation was introduced into the PSIP1-PWWP plasmid using the following primers:

Fwd: GGTTTATGGGAGATAGATTGCAATCCAAAAGTGAAATTTTC

Rev: GAAAATTTCACTTTTGGATTGCAATCTATCTCCCATAAACC

NMR SPECTROSCOPY

All NMR experiments were carried out at 298 K on a Bruker Avance spectrometer operating at 600 MHz (with cryo probe) ($H3K_C^{E7}$, 6-FAM, quality control) or 750 MHz ($H3C^{E7}$) ^1H Larmor frequency. Processing was done using the Bruker's TopSpin. Spectra were analyzed using Sparky (Goddard and Kneller, UCSF). Samples for the titration experiment contained ~140 μM ^{15}N-*wt*PSIP1-PWWP in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)).

Samples for the quality control experiments contained ~350 μM ^{15}N-*wt*PSIP1-PWWP in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)) and ^{15}N-PSIP1-PWWP N86C in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)).

Sample for the NMR titration experiments with $H3K_C^{E7}$ contained 400 μM ^{15}N-*wt*PSIP1-PWWP in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)).

Sample for the NMR titration experiments with 6-FAM contained 380 μM ^{15}N-*wt*PSIP1-PWWP in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche))

Sample for the NMR titration experiments with unalkylated $H3C^{E7}$ contained 300 μM ^{15}N-*wt*PSIP1-PWWP in NMR buffer (20 mM NaPi, pH 7.0, 150 mM NaCl, 10% D2O, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche))

MICROSCALE THERMOPHORESIS

All measurements were performed on a NanoTemper NT.115 instrument. Experiments on $H3K_C^{E7}$ and unalkylated $H3C^{E7}$ were performed with 0.35 μM NT-467 labeled PSIP1-PWWP N86C titrated with a serial dilution of the respective peptide (starting from 2.85 mM) in 10 mM Tris pH 7.5, 100 mM KCl, supplemented with 0.05% Tween-20, 25 C, 20%/20% LED/MST power, with 30/5 s laser-on/off time. All measurements were done using NanoTemper Monolith NT.115 Premium Capillaries. SD-test was performed by centrifuging tubes 1-3 (highest concentrations)

and tubes 14-6 (lowest concentrations) for 10 min at 15,000 g. 10 µL of the samples were added to 10 µL of 2x SD mix (4% SDS / 40 mM DTT) each and incubated for 5 min at 95 °C. The samples were then centrifuged and transferred to capillaries for subsequent fluorescence measurements. Data analysis was done using the supplied software NTAnalysis_v1.5.41. Fitting of the recorded fluorescence data was done using custom MatLAB scripts.

Peptide Synthesis

Unalkylated $H3C^{E7}$ was synthesized automatically, using standard Fmoc-based solid phase peptide synthesis. The backbone sequence (SAPATGGVCKEEEEEEE) was synthesized on a SRAM resin and the N-terminus acetylated after the final coupling step. The crude product was cleaved off the solid support using a cleavage solution of 95%(V/V) TFA, 2.5%(V/V) TIPS and 2.5%(V/V) H2O. Purification of the peptide was done using LC-MS (Column machine etc.) isolating a peak with molecular mass of 1834.8 g/mol (m/z = 1) (theoretical: 1833.76 g/mol). The product was dissolved in NMR buffer. The stock concentration was 20.34 mM at pH 6.97.

$H3K_C^{E7}$ was synthesized automatically, using standard Fmoc-based solid phase peptide synthesis. The backbone sequence (SAPATGGVCKEEEEEEE) was synthesized on a Wang resin without N-terminal acetylation after the final coupling step. The Cysteine side chain thiol group was protected with an Mmt group. For alkylation, Cys(Mmt) was selectively deprotected on-resin with a cleavage solution of 2%(V/V) TFA, 98%(V/V) DCM. Completion was ensured by monitoring the yellow color of the trityl-cation. The resin was washed with base before further reactions. Alkylation was performed by adding 16 Eq. (2-bromoethyl)-trimethylammonium bromide and 16 Eq. Cs_2CO_3 to the reactor in DMF. Both compounds show limited solubility and partially remain as solid. The reactor is rotated at 50 °C for 1.5 h under exclusion of UV light. After the reaction, the reactor is washed with H2O to remove remaining catalyst/reagent. The resin is then washed with DCM and dried for subsequent cleavage. The crude product was cleaved off the solid support using a cleavage solution of 95%(V/V) TFA, 2.5%(V/V) TIPS and 2.5%(V/V) H2O. Purification of the peptide was done using LC-MS (Column machine etc.) isolating a peak with molecular mass of 1877.9 g/mol (m/z = 1) (theoretical: 1877.84 g/mol). The product was dissolved in NMR buffer. The stock concentration was 10.19 mM at pH 7.0.

PEPTIDE CONCENTRATIONS

Determination of peptide stock concentrations was done using 1D NMR spectra with TSP as reference compound (Figure S4-1). The integral of the isolated singlet arising from TSP methyl groups (0.0 ppm) was put in relation to the peak for Val35 methyl groups in the peptides (~0.8 ppm) to back-calculate the effective peptide concentrations. The Val signal was verified in Chapter 3 of this thesis by 2D NMR spectra (TOCSY, ^{13}C-HSQC, see Figure S3-4).

REFERENCES

1. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
2. Speranzini, V., et al., *Touch, act and go: landing and operating on nucleosomes.* EMBO J, 2016. **35**(4): p. 376-88.
3. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nature Reviews Genetics, 2016. **17**: p. 487.
4. Dawson, M.A., T. Kouzarides, and B.J.P. Huntly, *Targeting Epigenetic Readers in Cancer.* New England Journal of Medicine, 2012. **367**(7): p. 647-657.
5. Ren, C., et al., *Small-Molecule Modulators of Methyl-Lysine Binding for the CBX7 Chromodomain.* Chemistry & Biology, 2015. **22**(2): p. 161-168.
6. Jain, A.K. and M.C. Barton, *Bromodomain Histone Readers and Cancer.* Journal of Molecular Biology, 2017. **429**(13): p. 2003-2010.
7. Zaware, N. and M.-M. Zhou, *Chemical modulators for epigenome reader domains as emerging epigenetic therapies for cancer and inflammation.* Current Opinion in Chemical Biology, 2017. **39**: p. 116-125.
8. Wagner, T., et al., *Mind the Methyl: Methyllysine Binding Proteins in Epigenetic Regulation.* ChemMedChem, 2014. **9**(3): p. 466-483.
9. Milosevich, N. and F. Hof, *Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins.* Biochemistry, 2016. **55**(11): p. 1570-1583.
10. Nevola, L. and E. Giralt, *Modulating protein–protein interactions: the potential of peptides.* Chemical Communications, 2015. **51**(16): p. 3302-3315.
11. Yokoyama, A. and M.L. Cleary, *Menin critically links MLL proteins with LEDGF on cancer-associated target genes.* Cancer Cell, 2008. **14**(1): p. 36-46.
12. Ciuffi, A. and F.D. Bushman, *Retroviral DNA integration: HIV and the role of LEDGF/p75.* Trends in Genetics, 2006. **22**(7): p. 388-395.
13. Ganapathy, V. and C.A. Casiano, *Autoimmunity to the nuclear autoantigen DFS70 (LEDGF): What exactly are the autoantibodies trying to tell us?* Arthritis & Rheumatism, 2004. **50**(3): p. 684-688.
14. Daugaard, M., et al., *Lens Epithelium-Derived Growth Factor Is an Hsp70-2 Regulated Guardian of Lysosomal Stability in Human Cancer.* Cancer Research, 2007. **67**(6): p. 2559.
15. Huang, T.-s., et al., *LEDGF/p75 has increased expression in blasts from chemotherapy-resistant human acute myelogenic leukemia patients and protects leukemia cells from apoptosis in vitro.* Molecular cancer, 2007. **6**: p. 31-31.
16. van Nuland, R., et al., *Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain.* Epigenetics & Chromatin, 2013. **6**(1): p. 12.
17. Greschik, H., R. Schüle, and T. Günther, *Selective targeting of epigenetic reader domains.* Expert Opinion on Drug Discovery, 2017. **12**(5): p. 449-463.
18. Ludwig, C., *Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösungen.* Sitzungsber Akad Wiss Wien Math-Naturwiss 1856. **20:539**.

19. Wienken, C.J., et al., *Protein-binding assays in biological liquids using microscale thermophoresis.* Nat Commun, 2010. **1**: p. 100.
20. Duhr, S. and D. Braun, *Why molecules move along a temperature gradient.* Proceedings of the National Academy of Sciences, 2006. **103**(52): p. 19678.
21. Seidel, S.A.I., et al., *Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions.* Methods, 2013. **59**(3): p. 301-315.
22. Jerabek-Willemsen, M., et al., *MicroScale Thermophoresis: Interaction analysis and beyond.* Journal of Molecular Structure, 2014. **1077**: p. 101-113.
23. Corbeski, I., et al., *Microscale Thermophoresis Analysis of Chromatin Interactions*, in *Bacterial Chromatin: Methods and Protocols*, R.T. Dame, Editor. 2018, Springer New York: New York, NY. p. 177-197.
24. Simon, M.D., *Installation of site-specific methylation into histones using methyl lysine analogs.* Curr Protoc Mol Biol, 2010. **Chapter 21**: p. Unit 21 18 1-10.
25. You, F., et al., *Cell-Free Bioassay for Measurement of Dioxins Based on Fluorescence Enhancement of Fluorescein Isothiocyanate-Labeled DNA Probe.* Analytical Chemistry, 2006. **78**(20): p. 7138-7144.
26. Ogawa, T., et al., *Fluorescence enhancement of fluorescein isothiocyanate-labeled protein a caused by affinity binding with immunoglobulin g in bovine plasma.* Sensors (Basel, Switzerland), 2009. **9**(10): p. 8271-8277.

SUPPLEMENTARY INFORMATION

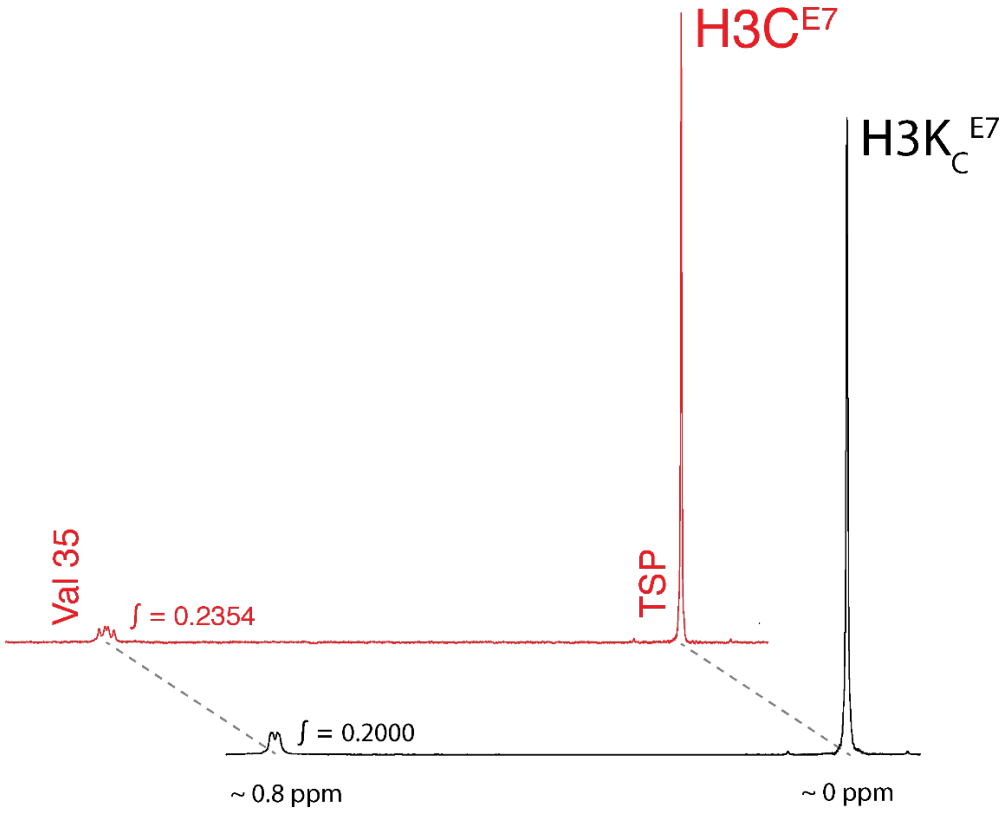


Figure S4-1. Determination of peptide concentrations of H3K$_C^{E7}$ and unalkylated H3C^{E7}.

Peaks for Val35 for all H3 peptide samples that were used for concentration determination. Integrals for TSP were set to 1 and the corresponding Val35 integrals are annotated in figure. Respective TSP concentrations were 3.39 mM (H3K$_C^{E7}$) and 0.5647 mM (unalkylated H3C^{E7}). Stock dilutions were 1:10 (H3K$_C^{E7}$) and 1:100 (unalkylated H3C^{E7}).

5

General Conclusions & Perspectives

JOINING FORCES IN STRUCTURAL BIOLOGY

Understanding the organization and activity of DNA has come a long way since Walther Flemming first introduced the term chromatin in the 1880s. A milestone herein is the discovery and subsequent determination of an atomic resolution structure of the fundamental basis of chromatin, the nucleosome. This facilitated the understanding of how effector proteins operate on the nucleosome and drive chromatin biology. Since a growing number of these proteins is attributed important roles in pathologies, an increasing interest in inhibitors for protein-nucleosome interactions is emerging. To direct the efforts to find modulators for epigenetic interactions, an understanding of the underlaying mechanisms of binding and recognition is vital. This raises the need to resolve the 3D structures of nucleosome-protein complexes. The main techniques used in the structural biology study of these complexes are X-ray crystallography, NMR spectroscopy and cryo-electron microscopy, all of them with their own advantages and limitations. While X-ray crystallography is able to yield high-resolution structures, its general limitation is the crystallization process and possible molecular rearrangements upon crystallization. Even though cryo-electron microscopy leads to increasingly well resolved structures and freezes the molecule in its native fold, it has the same limitation as X-ray crystallography: the resulting structures are isolated snapshots that lack dynamics and pathways between conformational states. However, dynamics are of high importance in the interaction of biomolecules. Here, NMR spectroscopy offers insights into the dynamic nature of protein complexes and is especially well suited for flexible proteins. Additionally, NMR offers a clean robust interrogation of protein interactions, at atomic resolution. In return however, NMR is limited by the size of the molecular assembly and studying large complexes requires sophisticated isotope labelling schemes. Even though the first crystal structure of the nucleosome was published in 1997, the limitations of the above techniques caused the number of atomistic resolution structures of nucleosome-bound effector proteins or protein complexes to be limited compared to the entirety of protein-nucleosome interactions. This sparked ever more successful applications of the integrative use of biophysical data on different aspects of protein binding in data-driven modelling to resolve otherwise inaccessible structures. The results of these remarkable studies are reviewed in **Chapter 1**.

In this thesis, I aimed to show the immense potential of integrative modelling to direct the development of inhibitors of epigenetic effector protein activity.

The Power and Challenges of Structural Modelling as Seen by RNF168

In **Chapter 2**, we investigated the molecular basis of how the ubiquitin E3 ligase RNF168 recognizes its nucleosomal substrate and directs its corresponding E2 subunit towards its target residue for specific ubiquitination. Besides valuable previously reported insights into the activity and identification of crucial residues of this epigenetic writer protein, a structure of a nucleosome-bound state was yet to be shown. We used NMR spectroscopy to show that RNF168 binds the acidic patch on the interface between histone H2A and H2B. On the RNF168 side, we showed with mutagenesis that besides the previously reported residue R57, various arginine residues distributed over an α-helix compose the interaction interface with the nucleosome. With the use of this data in structural modelling alone, two energetically similar conformations of the RNF168-nucleosome complex were obtained that differed by 180° in their RNF168 orientation on the nucleosome surface. This is of special interest in the light of substrate specificity. While the addition of the corresponding E2 subunit UbcH5c to one conformation positioned the catalytic center in good agreement with ubiquitination of lysine on the N-terminal tail of H2A, the second directed the ubiquitination complex towards the C-terminal tail. Interestingly, another E3 ligase, Ring1B, is reported to ubiquitinate lysine residues in the latter region. This is however not the case for RNF168 and posed an important challenge to address. The information accessible by NMR and mutagenesis allowed in this case for the mapping of interaction surfaces on both nucleosome and RNF168 but did not provide information on their relative orientation or on intermolecular distances in this case. This can potentially lead to additional ambiguity in the resulting structures. Additionally, potential incomplete mapping of the interface may bias the outcome. Another possible issue is that conformational changes of subunits upon binding cannot be reliably modelled. It thus could be conceived that structural rearrangements within the H2A/H2B dimer upon binding would favor the target specific positioning of RNF168. To discriminate between the two possible orientations of RNF168 on the nucleosome, cross-linking mass spectrometry was employed to get additional distance restraints. Indeed, a cross-link specific for the N-terminal orientation allowed for a proper discrimination. Furthermore, we validated our model by engineering a H2B E110A mutant. This acidic patch residue is in the interaction interface of only the conformation that directs for N-terminal ubiquitination. Indeed, the mutant construct was observed to silence RNF168 activity. In contrast, Ring1B activity was not affected. This serves as a prime example of how different biophysical techniques can be employed to derive a structural model and highlights an important additional point.

Structural modeling will always result in a structure and its quality crucially depends on the input data, thus requiring careful biochemical validation.

The Shape of the Basic Patch Potentially Dictates Specific Acidic Patch Binding

While we were able to verify the orientation of nucleosome-bound RNF168-UbcH5c, the fundamental molecular interactions that are causing the drastic change in orientation on the nucleosome between RNF168 and Ring1B still have to be proven. One hypothesis is that the RNF168 basic helix and a surface groove on the acidic patch align especially well for specific insertion. This is supported by the importance of all arginine residues in the basic anchor helix. Interestingly, acidic patch binding domains often show these additional basic residues that interact with the acidic patch (Figure 5-1). While the "arginine anchor" is generally attributed with the binding to the acidic patch, the relative position of secondary positively charged residues may favor a specific orientation by a well-defined network of electrostatic interactions with the acidic patch. This is possibly true for RNF168 as well. It cannot be excluded that mutants that lack Arg residues, other than the arginine anchor, in the anchor helix still bind the acidic patch, but are not positioned properly for effective ubiquitination. While this is also suggested for Arg mutants in our study, additional studies investigating the specific binding modes of these mutants will yield significant information on the precise impact of the overall shape and size of acidic patch binding epitopes. This is of particular interest, since we showed that structure-guided interference with the electrostatic interaction surface allows for selective *in vitro* silencing of RNF168, as demonstrated by the H2B E110A nucleosome mutant. Following this rational, it might be possible to select specific histone mutations for a large variety of acidic patch binders essentially giving control over selective activity in vitro. Furthermore, identification of specific basic patch interaction sites in combination with structural models can help to identify binding epitopes to be targeted by small molecule inhibitors.

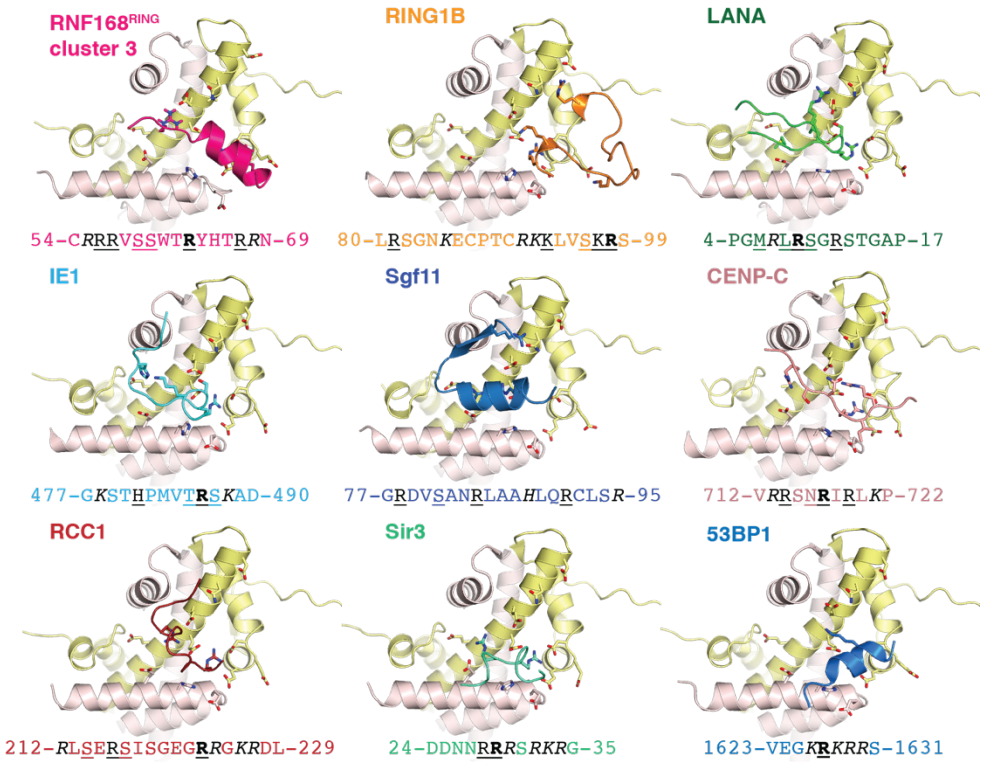


Figure 5-1. Network of electrostatic interactions for selected acidic patch binding proteins.

The acidic patch binding elements of RNF168, RING1B, LANA, IE1, Sgf11, CENP-C, RCC1, Sir3 and 53BP1 are show as cartoon with selected side chains involved in the interaction shown as sticks. The corresponding amino acid sequence is shown below each figure. Underlined residues are involved in hydrogen-bonding or hydrophobic interaction with the acidic patch. Positively charged residues shown in black, arginine anchor in bold, flanking residues in italic.

The Use of Structural Models in Rational Drug Design

In our work on RNF168, we made use of our structural model to rationally design a nucleosome mutant that selectively interferes with the activity of epigenetic effector proteins. This is of special importance since the regulation of histone modifications is a promising target for epigenetic therapeutics. An example hereof is the inhibition of acetyllysine reader domains and histone deacetylases by small molecules. While there is a growing number of drug candidates advancing to clinical trials, the number of potential inhibitors for the readout of trimethyllysine (Kme3) modifications is significantly lower. This is due to the shallow nature of the epitope that recognizes it, the aromatic cage. One of these aromatic cage-containing proteins is PSIP1 that recognizes a trimethylated lysine on the histone H3 tail via its PWWP reader

domain. It was previously shown that PSIP1-PWWP shows the lowest binding affinity for this specific modification in context of a H3 tail peptide model. Furthermore, it was established that the underlaying reason is a synergetic binding of PSIP1 to Kme3 through the aromatic cage and electrostatic interactions between positive residues on the protein surface and nucleosomal DNA. These insights were derived from a structural model based on NMR data and mutagenesis. In **Chapter 3**, we set out to use insights gained from a structural model to design H3 tail-derived peptides that incorporate both Kme3 and glutamate residues to mimic the electrostatic environment of nucleosomal DNA. We were able to show that this increases the affinity of these substrates for PSIP1-PWWP binding significantly. This is especially true for a branched peptide where an orthogonal glutamate backbone gives a more nucleosome-resembling structure to the negative charges.

NOTHING IN LIFE'S FOR FREE - TRADING SPECIFICITY FOR AFFINITY

While we observed an up to 50-times increase in affinity, investigating the protein-peptide interactions by NMR spectroscopy led to an interesting observation. We showed that the increase in affinity is mostly based on the electrostatic binding of the positive DNA interaction surface of PSIP1-PWWP while a decreasing fraction of the native H3 tail sequence is bound to the aromatic cage. This is of high interest since it shows that the electrostatic interactions lower specificity and are in competition with aromatic cage binding. In context of the native nucleosome-reader complex the question is raised how complex formation is achieved with such specificity in vivo. A possible explanation might be the static nature of nucleosomal DNA compared to highly flexible peptide tails. It might therefore be of interest to restrict the conformational space of the flexible peptide backbone and study the subsequent effect on the complex structure. We also hypothesized that a sterically non-optimized linker length between Kme3 and the negative charged backbone can cause a competition in binding their respective epitope. Therefore, a variation in linker length might deliver important insights into how the selectivity of H3 tail-derived peptides can be increased. This is of utter importance since, besides a high binding affinity, selectivity is a main criterium for a potential inhibitor.

IMPROVING THE SELECTIVITY OF PSIP1 BINDING H3 TAIL PEPTIDES

While we showed how the affinity of PSIP1-PWWP binding peptides can be increased, we paid for that with a loss in specificity. Even though variations in peptide

structure might result in higher affinity while retaining specificity, we additionally focused on the improvement of aromatic cage binding. In **Chapter 4**, we introduced a synthesis protocol that allows for the introduction of functional groups based on a cysteine residue in place of Kme3. This enables the screening for side chains that increase aromatic cage binding and subsequently target specificity. We showed the applicability of our protocol by synthesizing a cysteine-based Kme3 mimic in a proof-of-concept study. The suitability for other side chain structures however is still to be shown. We compared the resulting complex structures by NMR spectroscopy and found that the sulfur in place of carbon seems to have a negative effect on aromatic cage binding. We see that the negatively charged N-terminal peptide stretch encounters positive PSIP1-PWWP residues not observed for the native Kme3 equivalent of the peptide while the fraction of bound H3 tail sequence is decreased. While this seems counter intuitive, we must not forget that the mimicked Kme3 shows an initially already poor affinity for the aromatic cage. Therefore, the effect of the sulfur might be especially pronounced in this case. A more optimized engagement of the aromatic cage by a functional group other than trimethylammonium might likely render the sulfur-induced effect insignificant.

Fluorescence Screening of Potential Peptide Inhibitors

With a fast synthesis protocol for peptide substrates with varying aromatic cage-interacting side chain functional groups at hand we intended to improve the screening of these potential lead compounds. While NMR spectroscopy delivers vital insights into binding epitopes and structure of a complex, it is demanding on both time and sample consumption. We therefore set up an assay based on microscale thermophoresis (MST). For this purpose, we designed a mutant construct of PSIP1-PWWP for fluorescent labelling. With this at hand, we were able to reproduce the results obtained from NMR by fitting of specific fluorescence quenching, induced by peptide binding. The quenching showed to be specific for peptides with intact aromatic cage interaction and can therefore serve as a specific reporter for aromatic cage binding throughout screening for alternative functional side chain groups. Furthermore, we showed by NMR that peptides without aromatic cage engagement still bind to PSIP1-PWWP. This means that unalkylated peptides are likely to still show an MST effect. Though this has yet to be proven, it would make the use of an MST assay unnecessary and favor readout of initial fluorescence and fitting the data for binding dependent fluorescence quenching. Together, this provides an even more applicable approach to screen for aromatic cage binding moieties to increase the specificity and affinity of H3 tail-derived PSIP1-PWWP binders.

Closing Remarks

The studies presented in this work show the potential of the integrative use of biophysical data in defining the structural basis of protein interactions. Even if the results obtained hold a degree of ambiguity, this approach allows to iteratively refine and validate the model and interpret its meaning for the molecular basis of protein function. Often all three points at the same time. This dynamic nature makes the use of structural models in the design of therapeutic compounds especially useful since the inhibition of a certain protein function might not require a structure to be accurate down to the last atom but rather highlight key interactions or structural features that can be addressed in context of small molecule or peptide inhibitors. I therefore strongly advocate for using and sharing biophysical data, even seemingly not publishable negative results. This has the potential to allow for a synergetic rather than competitive use of techniques in structural biology to overcome their intrinsic limitations.

English Summary

Proteins and protein complexes that install, recognize and erase modifications on histone proteins play a fundamental role in orchestrating all DNA-templated processes. The pronounced involvement of these effector proteins in health and disease sparked a high interest in insights on how these effector proteins operate on their nucleosomal substrate on a molecular level. Vital insights into the molecular basis of chromatin biology are gained from how proteins form complexes with their nucleosomal substrate. Key herein are a growing number of structures of protein-nucleosome complexes obtained by X-ray crystallography and cryo-electron microscopy (cryo-EM). However, all biophysical techniques are subject to intrinsic limitations. One of the biggest challenges in structural biology is to overcome these limitations in order to get a deeper knowledge on the interactions between effector proteins and the nucleosome.

Chapter 1 presents an overview of nucleosome-bound protein structures. We first introduce the epitopes which form the docking platforms for proteins on the nucleosome and summarize the high-resolution structures of complexes that were obtained by X-ray crystallography and cryo-EM. We then focus on approaches that utilized data from various biophysical sources in structural modelling studies and introduce the resulting complexes including important insights into their molecular basis of recognition.

We applied this methodology in **Chapter 2** to investigate the molecular basis of nucleosome binding of RNF168 which plays a key role in DNA damage response. Its activity depends on ubiquitination of H2A K13/K15 by the RNF168 ^{RING} domain. We used NMR spectroscopy, cross-linking mass spectrometry and mutagenesis data, to show that RNF168^{RING} binds the acidic patch and to derive a data-driven structural model of the nucleosome-bound E3-E2 complex. This yielded atomistic insights into the interactions between several Arg residues on the basic binding helix of RNF168^{RING} and the acidic patch. The orientation of the E3 subunit on the nucleosome positions the catalytic center for ubiquitin transfer on the corresponding E2 in close proximity to the target lysine on the H2A N-terminus. Additionally, we tested our model by designing an H2A mutant that interferes with the electrostatics of the acidic patch in a region that is interacting with RNF168 while not being involved in binding of Ring1B, another acidic patch binding E3 subunit. Indeed, we show that this mutation causes selective silencing of RNF168 activity while Ring1B remains untouched. We then compared our resulting structural model with recent E3-E2~Ub structures that

highlighted the importance of an active 'closed' and inactive 'open' state of the complex. Comparison of these complexes with our structure showed that it is in agreement with a 'closed' state while suggesting clashes between ubiquitin and the nucleosome for 'open' states. We thus suggest that the nucleosome itself, by steric occlusion, constrains the conformational space of Ub in the E3-E2~Ub complex and thus promotes the formation of closed complexes.

In **Chapter 3**, we set out to utilize insights derived from a structural model of the epigenetic reader domain PSIP1-PWWP bound to its nucleosomal substrate. Highlighted are the importance of the reader binding to trimethylated lysine on the histone H3 tail and nucleosomal DNA alike. We aimed to use this information to increase the intrinsically poor binding affinity of a trimethyllysine-containing H3 tail peptide to PSIP1-PWWP by incorporating negatively charged residues as DNA mimics. By designing several peptide structures with increasing negative charge, we were able to show that the binding affinity continuously increases. This is especially the case for a H3 tail peptide with an orthogonal negatively charged backbone, resembling the charge distribution of a nucleosome. With analysis of bound-state peak positions derived from NMR titration studies, we found that the boost in affinity is caused by the electrostatic interactions while lowering the fraction of H3 tail sequence bound to its native binding epitope in and around the aromatic cage.

Due to its involvement in various pathologies, PSIP1-PWWP is an interesting drug target. However, this poses a challenge for the use of potential H3 tail-based inhibitors for PSIP1-PWWP activity. While we show that the incorporation of negative charges increases the binding affinity, it also causes a loss in specificity. We addressed this problem in **Chapter 4**, where we focused on a methodology for the optimization of aromatic cage binding to increase both affinity and specificity. We introduced a way to incorporate desired non-natural side chain functional groups into the peptide sequence based on alkylation of a cysteine residue. We then showed the applicability by comparing the PSIP1-PWWP-bound complexes of a negative N-terminal glutamate stretch containing H3 tail peptide with either native trimethyllysine (Kme3) or a cysteine-based mimic thereof. We found that while both displayed comparable binding affinities, the bound-state conformation of cysteine-based Kme3 showed a lower fraction of tail bound. This indicates that the sulfur has an effect that counters efficient aromatic cage complexation. However, due to the low affinity of Kme3, it has yet to be seen if the observed effect proves significant also for higher affinity functional groups for aromatic cage interaction other than trimethylammonium.

To allow for a fast screening of compounds optimized for aromatic cage binding, we decided to set up a microscale thermophoresis (MST) assay. However, we showed that the aromatic cage of PSIP1 interacts with the fluorescence label. Due to the low affinity of the peptide protein interaction the approach to label the peptide substrate was therefore not applicable. We instead engineered a PSIP1 mutant that allowed for selective protein labelling. In the subsequent measurements a quenching of initial fluorescence was observed in dependence of peptide concentration added to the protein. By denaturing the samples and observing a recovery of fluorescence, we were able to correlate this effect to peptide binding. Interestingly, we see that this effect is absent for cysteine-containing H3 tail peptides that were not modified to Kme3 analogues. However, NMR studies showed that these peptides still engage in PSIP1-PWWP binding purely based on electrostatic interactions. We concluded that the fluorescence quenching reports specifically on aromatic cage binding and is therefore of high value for the screening of aromatic cage interacting functional groups. Furthermore, we saw the initial fluorescence to be sufficient to derive binding affinities and proposed a synthesis protocol and assay for the screening of potential peptide-based PSIP1-PWWP inhibitors.

In **Chapter 5**, I provide a critical reflection on the studies presented in this thesis and discuss suggestions for future studies based on the findings presented in this thesis.

Nederlandse Samenvatting

Eiwitten en eiwitcomplexen die modificaties op de histon eiwitten in nucleosomen aanbrengen, herkennen en verwijderen spelen een fundamentele rol in de regulatie van DNA-gerelateerde processen. De belangrijke rol van deze klasse van eiwitten bij gezondheid en ziekte heeft geleid tot een intense interesse in hoe deze eiwitten op moleculair niveau op hun substraat opereren. Essentiële inzichten in de moleculaire basis van de biologie van chromatine zijn verkregen door te bestuderen hoe deze eiwitten complexen vormen met nucleosomen. Een groeiend aantal structuren van eiwit-nucleosoom complexen bepaald met Röntgenkristallografie en cryo-electronenmicroscopie (cryo-EM) spelen hierin een sleutelrol. Alle biofysische technieken kennen echter hun intrinsieke beperkingen. Het overwinnen van deze beperkingen om diepgaande kennis te verkrijgen van de interacties tussen effectoreiwitten en nucleosomen is eén van de grootste uitdagingen in structuurbiologie.

Hoofdstuk 1 geeft een overzicht van de structuren van eiwit-nucleosoom complexen. We introduceren de verschillende type bindingsplekken voor eiwitten op het nucleosoom en beschrijven structuren verkregen met Röntgenkristallografie en cryo-EM. Daarna bespreken we methodes die data uit verschillende biofysische bronnen gebruiken voor het genereren van structurele modellen samen met de resulterende modellen en de belangrijkste inzichten in de moleculaire basis van herkenning.

In **Hoofdstuk 2** hebben we deze methodologie toegepast om de moleculaire basis van binding door RNF168 aan het nucleosoom te onderzoeken. Dit eiwit speelt een cruciale rol in het herstellen van DNA-schade. De activiteit van RNF168 hangt af van de ubiquitinering van H2A K13/K15 door het RNF168^{RING} domein. We hebben NMR-spectroscopie, cross-linking massaspectrometrie en data uit mutagenese gebruikt om aan te tonen dat RNF168^{RING} aan de 'acidic patch' bindt en om een structuurmodel van het nucleosoom-gebonden E3-E2 complex te ontwikkelen. Dit heeft een atomair begrip opgeleverd van de interactie tussen de acidic patch en verschillende Arg residuen van de basische bindingshelix van RNF168^{RING}. Door de specifieke oriëntatie van de E3 subunit op het nucleosoom wordt het katalytische centrum van het E2 enzym dicht bij het te modificeren lysine residu op de H2A N-terminus gebracht. We hebben ons model getest door een H2A-mutant te ontwerpen waarin de electrostatische eigenschappen van de acidic patch verstoord zijn in een regio die wel interactie heeft met RNF168 maar niet betrokken is bij de binding van Ring1B, een andere E3 subunit die ook aan de acidic patch bindt. We laten zien dat deze mutatie inderdaad selectief de

activiteit van RNF168 verstoort en de activiteit van Ring1B ongemoeid laat. Vervolgens hebben we ons structuurmodel vergeleken met recente E3-E2~Ub structuren die het bestaan van een actieve 'gesloten' en een inactieve 'open' conformatie van het complex aantonen. Uit deze vergelijking blijkt dat onze structuur overeenkomt met een 'gesloten' conformatie, terwijl er voor 'open' conformaties botsingen zouden opteden tussen de ubiquitine en het nucleosoom. Daarom stellen wij voor dat het nucleosoom zelf, door sterische hinder, de conformationele ruimte van ubiquitine in het E3-E2~Ub complex beperkt en zo de vorming van gesloten complexen bevordert.

In **Hoofdstuk 3** hebben we een manier gezocht om de inzichten te benutten die zijn voortgekomen uit een structuurmodel van het epigenetische 'reader' domein PSIP1-PWWP gebonden aan een nucleosoomsubstraat. Dit model laat zien dat dit domein bindt aan zowel de getrimethyleerde lysine op de histon H3-staart als het nucleosomaal DNA. We wilden deze informatie gebruiken om de intrinsiek lage bindingsaffiniteit van een H3-peptide met trimethyllisine voor PSIP1-PWWP te verhogen, door negatief geladen residuen in te bouwen om zo DNA na te bootsen. Door het gebruik van verschillende peptidestructuren hebben we aangetoond dat de bindingsaffiniteit positief gecorreleerd is aan de negatieve lading. Dit is in het bijzonder het geval voor een H3-peptide met een vertakte, negatief geladen hoofdketen, die de verdeling van de lading in een nucleosoom nabootst. Met behulp van NMR-titratiestudies hebben we gevonden dat de verhoogde affiniteit veroorzaakt wordt door de electrostatische interacties, maar ook dat de trimethyllysine in de H3-sequentie minder goed bindt aan zijn natuurlijke bindingsepitoop in en rondom de 'aromatic cage'.

PSIP1-PWWP speelt een rol in verschillende ziekten en zou daarom een interessant doelwit kunnen zijn voor medicijnen. Voor het gebruik van potentiële remmers van PSIP1-PWWP-activiteit op basis van de H3-staart komt echter een uitdaging om de hoek kijken. Hoewel we laten zien dat het inbouwen van negatieve ladingen de bindingsaffiniteit vergroot, veroorzaakt het ook een verlies van specificiteit. Dit probleem bespreken we in **Hoofdstuk 4**, waarin we ons hebben gericht op een methode voor het optimaliseren van aromatic-cage-binding om niet alleen de affiniteit maar ook de specificiteit de vergroten. We hebben een manier geïntroduceerd om niet-natuurlijke functionele zijketens in te bouwen in de peptidesequentie door alkylering van een cysteïneresidu. Vervolgens hebben we de toepasbaarheid hiervan laten zien door de PSIP1-PWWP-gebonden complexen van een H3-peptide met een serie negatief geladen glutamaten aan de N-terminus te vergelijken met ofwel een natuurlijke trimethyllysine (Kme3) dan wel een nabootsing daarvan op basis van cysteïne-alkylatie. Hieruit bleek dat in de gebonden conformatie van de Kme3 op basis van cysteïne een kleiner deel van het H3-peptide gebonden is, terwijl beide complexen een vergelijkbare

bindingsaffiniteit vertonen. Dit betekent dat het zwavelatoom een negatief effect heeft op de vorming van een aromatic-cage-complex. In verband met de lage affiniteit van Kme3 is echter nog niet duidelijk of dit effect ook significant is voor functionele groepen met een hogere affiniteit voor de aromatic cage dan trimethyl-ammonium.

Om verbindingen die geoptimaliseerd zijn voor binding aan de aromatic cage sneller te kunnen screenen, hebben we besloten om een microscale thermophoresis (MST) assay op te zetten. Het bleek echter dat de aromatic cage van PSIP1 een interactie aangaat met het fluorescente label. Door de lage affiniteit van de peptide-eiwitinteractie was het labelen van het peptidesubstraat dan ook niet toepasbaar. In plaats daarvan hebben we een PSIP1-mutant gemaakt die selectieve eiwitlabeling mogelijk maakt. In de daaropvolgende metingen werd een uitdoving van de initiële fluorescentie geobserveerd die afhankelijk bleek van de toegevoegde concentratie van het peptide. Door de monsters te denatureren en vervolgens het herstel van de fluorescentie te observeren, waren we in staat om dit effect te correleren aan peptidebinding. Interessant is dat dit effect afwezig is voor cysteïne-bevattende H3-peptiden die niet gemodificeerd zijn tot Kme3-analogen. NMR-studies hebben echter aangetoond dat deze peptiden nog steeds binden aan PSIP1-PWWP, uitsluitend gedreven door elektrostatische interacties. We concluderen dat het uitdoven van de fluorescentie specifiek informatie geeft over binding met de aromatic cage en daarom van grote waarde is voor het screenen van functionele groepen die daar een interactie mee aangaan. Verder hebben we gezien dat de initiële fluorescentie voldoende is om bindingsaffiniteiten af te leiden en hebben we een syntheseprotocol en een assay voor de screening van potentiële peptide-gebaseerde PSIP1-PWWP-remmers voorgesteld.

Hoofdstuk 5 bevat een kritische reflectie op het onderzoek dat beschreven is in dit proefschrift, alsmede suggesties voor toekomstig onderzoek naar aanleiding van de hierin gepresenteerde bevindingen.

Curriculum Vitae

Velten Horn was born on March 28th 1988 in Gross-Umstadt, Germany. He completed his Bachelor and Master studies in chemistry at Technische Universität Darmstadt. The research resulting in his master thesis was conducted under the supervision of Dr. Daniel Tietze in the group of Prof. Dr. Gerd Buntkowsky and was titled: "Subtype specific inhibitiors of Gq11 G-proteins – NMR structure determination and theoretical binding studies". He then moved to Leiden, Netherlands to pursue a PhD in the field of structural biology at Leiden University in the group of Dr. Hugo van Ingen with Prof. Dr. Marcellus Ubbink as his promotor. For the last year of his PhD research, he followed Hugo van Ingen to Utrecht University where he concluded the studies presented in this work.

List of Publications

PhD-related publications:

V. Horn, S. Jongkees & H. van Ingen (2019). Mimicking a nucleosomal context in a peptide: Increasing the binding affinity towards the epigenetic reader PSIP1.

Manuscript in preparation

V. Horn, M. Uckelmann, H. Zhang, J. Eerland, I. Aarsman, U. B. Le Paige, C. Davidovich, T. Sixma & H. van Ingen (2019). Structural basis of specific H2A K13/15 ubiquitination by RNF168.

Nat Commun. 10, 1751

V.Horn & H. van Ingen (2019). Recognition of nucleosomes by chromatin factors: Lessons from data-driven docking-based structures of nucleosome-protein complexes.

Accepted for publication (DOI: 10.5772/intechopen.81016)

I. Corbeski, **V. Horn**, R. A. van der Valk, U. B. Le Paige, R. T. Dame & H. van Ingen (2018). Microscale Thermophoresis Analysis of Chromatin Interactions.

Methods Mol Biol. 1837, 177 - 197

S. Xiang, U. B. Le Paige, **V. Horn**, K. Houben, M. Baldus & H. van Ingen (2018). Site-specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy.

Angew Chem Int Ed Engl. 57(17), 4571 - 4575

Publications from previous research work:

R. Schrage, A.-L. Schmitz, E. Gaffal, S. Annala, S. Kehraus, D. Wenzel, K. M. Büllesbach, T. Bald, A. Inoue, Y. Shinjo, S. Galandrin, N. Shridhar, M. Hesse, M. Grundmann, N. Merten, T. H. Charpentier, M. Martz, A. J. Butcher, T. Slodczyk, S. Armando, M. Effern, Y. Namkung, L. Jenkins, **V. Horn**, A. Stößel, H. Dargatz, D. Tietze, D. Imhof, C. Galés, C. Drewke, C. E. Müller, M. Hölzel, G. Milligan, A. B. Tobin, J. Gomeza, H. G. Dohlman, J. Sondek, T. K. Harden, M. Bouvier, S. A. Laporte, J. Aoki, B. K. Fleischmann, K. Mohr, G. M. König, T. Tüting & E. Kostenis

The experimental power of FR900359 to study Gq-regulated biological processes. (2015)

Nat Commun. 6, 10156

A. – L. Schmitz, R. Schrage, E. Gaffal, T. H. Charpentier, J. Wiest, G. Hiltensperger, J. Morschel, S. Hennen, D. Häußler, **V. Horn**, D. Wenzel, M. Grungmann, K. M. Büllesbach, R. Schröder, H. H. Brewitz, J. Schmidt, J. Gomeza, C. Galés, B. K. Fleischmann, T. Tüting, D. Imhof, D. Tietze, M. Gütschow, U. Holzgrabe, J. Sondek, T. K. Harden, K. Mohr & E. Kostenis

A Cell-Permeable Inhibitor to Trap Gαq Proteins in the Empty Pocket Conformation. (2014)

Chem Biol. 21(7), 890 – 9