

Probing Tandem Epigenetic Reader Domains



Thesis submitted for the degree of

Doctor of Philosophy in Systems Approaches to Biomedical Science

Angelina Measures, Lincoln College, University of Oxford

Supervisors:

Prof. Stuart Conway (University of Oxford)

Dr Rachel Grimley (Pfizer Neusentis)

Dr Ian Storer (Pfizer Neusentis)



Thesis submitted for the degree of:
Doctor of Philosophy in Systems Approaches to Biomedical Science

Probing Tandem Epigenetic Reader Domains

I Abstract

Epigenetic signals include chemical modifications on DNA and post-translational modifications to amino acids in histone proteins. One of the most common sites of modification are lysine residues: addition of acetyl or methyl groups are recognised by “reader” domains including bromodomains and plant homeodomains (PHDs). These domains mediate protein-protein interactions, affecting downstream events including gene transcription. PHDs and bromodomains are found in close proximity within some proteins, raising the possibility of combinatorial signalling eliciting a range of responses. Tripartite motif-containing proteins 24 and 33 (TRIM24 and TRIM33) contain an adjacent PHD and bromodomain which recognise methylation and acetylation states of lysine residues, respectively (Figure 1). These interact with histone-mimicking peptides that display a specific combination of unmodified, trimethylated and acetylated lysine residues. Work has been carried out towards inhibition of these epigenetic readers, through target analysis, assay development, and screening of putative binding partners. Refinement of expression systems for TRIM24 and TRIM33 enabled structural investigations, and development of biophysical assays to provide screening capacity. These assays were used to further characterise interactions through profiling varyingly modified histone-like peptides. Based on structural information and computational modelling, an *in silico* screen was performed to identify putative binding partners, and generate a chemical pharmacophore. *In vitro* screening of a bespoke collection of compounds based on these results, was supplemented with a broad screen targeting protein-protein interactions. Subsequent rounds of validation identified four compounds which bind to either TRIM24 or TRIM33 (Figure 2). Prompted by the reliance of competition based assays on knowledge of a high affinity interaction with a ligand or peptide, a dimethylisoxazole unit was used as an acetyl lysine mimicking residue. Incorporation in a histone-like peptide, *via* selective alkylation, resulted in increased affinity of the peptide to the BRD4 bromodomain. Application to epigenetic reader systems with unknown cognate binding partners would improve assay capacity for ligand identification. This approach will facilitate synthesis and screening of novel probe molecules, to elucidate the function of tandem recognition domains in interpreting the “epigenetic code”.

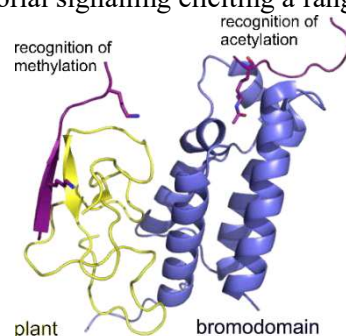


Figure 1 TRIM histone recognition by the tandem PHD bromodomain cassette (PDB: 3O34 & 3O37). Recognition of multiple modification sites on a H3 histone tail – lysine residues 4, 9, and 23.

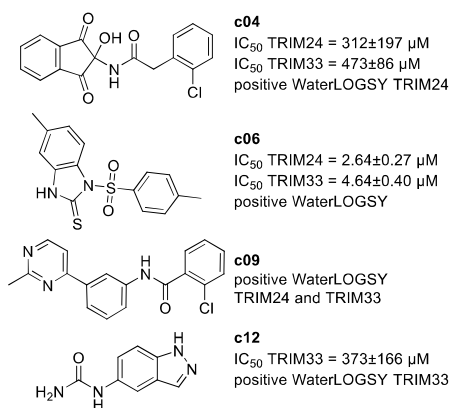


Figure 2 Compounds identified and validated as ligands for TRIM24 and TRIM33

II Acknowledgements

Many thanks to Stuart Conway for encouragement and supervision. For many interesting discussions about research, or philosophy about science in general; for letting me get overexcited about this project, but making sure I tidy up my toys before I run after the next game. Rachel Grimley and Ian Storer provided enthusiastic support, without which I would not have dared to tackle this task. I am grateful to Pfizer Neusentis for generous studentship support, including several months working with the fantastic team at Granta Park.

For all the members of the Conway Group who made it a joy to go to the lab each day, thank you – great people, great work! And Hundred Acre Wood wouldn't have been half as much fun for Tigger if Eeyore wasn't around. I am also thankful to Andrew Hamilton for sparking an interest in research during my Masters placement, and the Hamilton group for encouraging me to pursue it and supporting me throughout. If my partner in crime and I knew where we'd be five years later...

Too many people have contributed to this work to name them all, but as a start I would like to thank: the Schofield Group for assistance in molecular biology, particularly Adam Hardy for tolerating too many questions and Michael McDonough for crystallisation advice, Anthony Bradley for contributions to virtual screening, Holger Kramer for access to MALDI facilities and training, Nisha Patel for native-MS experiments, and David Zollman for nucleosome studies.

We wish to thank the EPSRC funded network, PPI-net (EP/I037210/1 and EP/I037172/1) and GSK for making the PPI-net screening collection available, the Structural Genomics Consortium for provision of biological material and a control compound, and Professor Joan Massagué for plasmids containing DNA for the *TRIM33* gene. We would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (<http://dx.doi.org/10.5281/zenodo.22558>), and thank the helpful support staff there for assistance in scripts. We also thank Diamond Light Source for access to beamline I03 (proposal number mx9306) that contributed to the results presented here.

In the course of my DPhil I gained a valuable collaborator. Thanks to him I have been fortunate to have a variety of stunning backdrops to my writing, been encouraged to look at my research brightly, and still have a smile on my face by the end of it.

Najlepša hvala za vse, moj dragi mož.



When my father resolved during home-school to teach us how to learn, he probably didn't imagine how far that would get us; learning from my mother's improvised cooking style kept me, and many around me, well fuelled along the way. Growing up with such a wonderful (and large) family gave me much to aspire to, now it is a delight to see it continue to grow.

Having support in fellowship and prayer from family and friends, at church and across the University, has been a great blessing. The greatest joy of all this, has been to learn more about God's fascinating world, and be left in awe of Him.

For what can be known about God is plain to them, because God has shown it to them. For his invisible attributes, namely, his eternal power and divine nature, have been clearly perceived, ever since the creation of the world, in the things that have been made. *Romans 1:19-20*

III Abbreviations

2IA	2-iodoacetamide
2-TY	2×typtone-yeast extract growth media
a.u.	arbitrary units
A ₂₈₀	absorbance at 280 nm
Ac	acetyl
AlphaScreen	amplified luminescent proximity homogeneous assay screen
AML	acute myeloid leukemia
APS	ammonium persulfate
BAZ1B	bromodomain adjacent to zinc finger domain 1B
BCP	bromodomain containing protein
β-ME	β-mercaptoethanol
BPTF	bromodomain PHD finger transcription factor
BRD	bromodomain-containing protein
BSA	bovine serum albumin
CBP	CREB binding protein
CD	circular dichroism
CH ₂ Cl ₃	dichloromethane
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHCA	α-cyano-4-hydroxycinnamic acid
CHCl ₃	chloroform
cHLM	calculated human liver microsomal stability
CID	collision-induced dissociation
Cl-DMI	5-(chloromethyl)-3-methylisoxazole
CM5	carboxymethylated dextran
CNS	central nervous system
CREB	c-AMP-response element binding protein
Da	dalton(s): unit of mass
DMI	3,5-dimethylisoxazole
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside 5'-phopshate
DSF	differential scanning fluorimetry
DTT	dithiothreol
<i>E. coli</i>	<i>Escherichia coli</i> bacteria
EDTA	ethylenediaminetetraacetic acid
EP300	histone acetyltransferase p300
EPL	expressed protein ligation
ESI	electrospray ionization

ϵ_{280}	molar absorption coefficient at 280 ($M^{-1} \text{ cm}^{-1}$)
FAC	final assay concentration
FIH	factor inhibiting hypoxia-inducible factor
FPLC	fast protein liquid chromatography
FRAP	fluorescence recovery after photobleaching
FT-ICR-MS	Fourier Transform Ion Cyclotron Resonance Mass spectrometer
GFP	green fluorescent protein
GST	glutathione <i>S</i> -transferase
H	histone polypeptide
h	hour(s)
HAT	histone acetyltransferase
HDAC	histone deacetylase
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high-performance liquid chromatography
<i>hν</i>	microwave irradiation
IBiD	interferon response binding domain
IC ₅₀	half maximal inhibitory concentration
ICM	Internal Coordinate Mechanics
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
KAc	acetylated lysine
KIX	kinase-inducible domain interacting
KMe ₁₋₃	mono-, di-, trimethylated lysine
L	litre
LC	liquid chromatography
LE	ligand efficiency
LIC	ligation independent cloning
LIFT	
LLE	lipophilic ligand efficiency
MCS	multiple cloning site
MDCK	Madin-Darby canine kidney
Me	methyl
MeOH	methanol
min	minute(s)
MLL1	mixed lineage leukemia 1
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
NCL	native chemical ligation
NCS	<i>N</i> -chlorosuccinimide

NEt ₃	triethylamine
NMM	<i>N</i> -methylmaleimide
NTA	nitrilotriacetic acid
o/n	overnight
OH	hydroxyl
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PFI	property forecast index
PHD	plant homeodomain
PHF8	PHD finger protein 8
PPI(s)	protein-protein interaction(s)
PTM	post-translational modification
r.t.	room temperature
RID	receptor interaction domain
RING	really interesting new gene
RNA	ribonucleic acid
s.d.	standard deviation
s.e.m.	standard error of the mean
SDF	structure-data file
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	second(s)
SLIC	sequence and ligation independent cloning
SMILES	simplified molecular-input line-entry system
SPR	surface plasmon resonance
SUMO	small ubiquitin-like modifier
t	time
TAF1	TATA box binding protein-associated factor
TEV	tobacco etch virus
THF	tetrahydrofuran
TIC	total ion count
TLC	thin layer chromatography
T _m	denaturation ('melting') temperature
TPSA	topological polar surface area
TRIM	tripartite motif-containing
Tris	trisaminomethane
Water-LOGSY	water-ligand observed via gradient spectroscopy
WT	wild-type

IV Contents

I	Abstract.....	i
II	Acknowledgements.....	ii
III	Abbreviations.....	iii
IV	Contents.....	vi

Chapter 1: Introduction

1.1	Epigenetics.....	1
1.2	Chromatin structure and histone modifications.....	2
1.3	Epigenetic reader modules.....	5
1.3.1	Bromodomains.....	6
1.3.2	Plant homeodomain fingers.....	8
1.3.3	Combinatorial modifications.....	11
1.4	TRIM24 and TRIM33.....	14
1.5	Small molecule inhibition of reader domains.....	16
1.5.1	Bromodomains.....	16
1.5.2	PHD fingers.....	17
1.6	Utility of assays towards epigenetic reader inhibition.....	18
1.7	Assays used in the discovery of bromodomain inhibitors.....	19
1.8	Computational modelling.....	22
1.9	Clinical applications.....	23
1.10	Aims.....	25
1.11	References.....	27

Chapter 2: Biological Characterisation

2.1	Introduction.....	35
2.1.1	TRIM24 and TRIM33.....	35
2.1.2	Aims.....	36
2.2	Domain organisational predictions.....	37
2.3	Expression and purification.....	39
2.3.1	Plasmids.....	39

2.3.2	Expression trials	40
2.3.3	Purification	41
2.3.4	Anion-exchange purification	43
2.4	Assay development.....	45
2.4.1	Tryptophan fluorescence.....	45
2.4.2	Differential scanning fluorimetry.....	46
2.4.3	AlphaScreen	48
2.4.4	Inactivating mutations.....	55
2.4.5	Surface plasmon resonance.....	58
2.4.6	Isothermal titration calorimetry of selected peptides.....	60
2.5	Conclusions	62
2.6	References	64

Chapter 3: Lead Identification

3.1	Introduction	66
3.1.1	Aims	67
3.2	Computational approach.....	69
3.2.1	Receptor state definition	70
3.2.2	<i>In silico</i> screening.....	71
3.2.3	Compound selection	74
3.3	Screening cascade.....	77
3.3.1	Maybridge collection screening.....	79
3.3.2	PPI-net collection screening	89
3.4	Hit validation.....	91
3.4.1	AlphaScreen TruHits kit	91
3.4.2	Differential scanning fluorimetry.....	95
3.4.1	Maybridge fragment characterisation.....	97
3.4.2	WaterLOGSY	103
3.5	Conclusions	107
3.6	References	109

Chapter 4: Structural Studies

4.1	Introduction.....	111
4.1.1	Structural coverage of bromodomain family	111
4.1.2	Aims.....	114
4.2	Expression and purification.....	114
4.2.1	Cleavage from N-terminal His ₆ vector	114
4.2.2	N-terminal GST vector.....	116
4.2.3	Cloning into N-terminal SUMO vector	118
4.2.4	Purification and cleavage for crystallisation.....	120
4.3	Crystallisation of TRIM24	125
4.4	Association with nucleosome.....	128
4.5	Conclusions.....	131
4.6	References.....	132

Chapter 5: Acetyl-lysine-mimicking Residue

5.1	Introduction.....	134
5.1.1	Non-natural acetyl-lysine mimicking amino acids	135
5.1.2	BRD4 bromodomains.....	136
5.1.3	3,5-Dimethylisoxazoles as KAc mimics.....	137
5.1.4	Aims.....	138
5.2	Inhibitory values of free amino acids	139
5.3	Incorporation into peptides	142
5.3.1	IC ₅₀ value determination	143
5.3.2	Validation of controls.....	144
5.4	<i>In situ</i> alkylation	146
5.4.1	Reaction monitoring.....	149
5.4.2	Alkylation optimisation.....	153
5.4.3	Isolation and affinity measurements.....	159
5.6	Conclusions.....	161
5.7	References.....	162

Chapter 6: Conclusions

.....	164
-------	-----

Chapter 7: Materials and Methods

7.1	Biological methods	167
7.1.1	Materials and reagents	167
7.1.2	General methods	168
7.1.3	Microbiology	168
7.1.4	Constructs	175
7.1.5	Protein expression and purification.....	176
7.1.6	Assay procedures	182
7.1.7	Crystallisation.....	189
7.1.8	Non-denaturing mass spectrometry.....	190
7.2	Computational methods	191
7.2.1	Docking file preparation	191
7.2.2	Docking	193
7.2.3	Results analysis.....	193
7.3	Chemical methods	195
7.3.1	General experimental.....	195
7.3.2	Synthesis of alkylating agent	196
7.3.3	Alkylation	198
7.4	References	200

Chapter 8: Appendices

8.1	Computational processing scripts	202
8.2	Vectors	207
8.3	Protein characterisation data	209
8.4	Compound NMR spectra	243
8.5	AlphaScreen: biotinylated peptide signal response	246
8.6	References	250

Chapter 1: Introduction

Epigenetic regulation of gene transcription is affected by chemical modifications on DNA and post-translational modifications (PTMs) of amino acids in histone proteins¹ – the central unit of a chromosome’s dynamic storage structure. It is proposed that a certain pattern of epigenetic modifications can result in a defined downstream phenotype, leading to the concept of the “histone code”.² Recent work has shown that modulation of proteins that add (‘write’), bind (‘read’), or remove (‘erase’) the histone code is a viable therapeutic strategy,³ notably for certain types of cancer.⁴ Lysine residues are susceptible to PTMs, and addition of an acetyl or methyl group(s) can be recognised by bromodomains and plant homeodomains (PHDs) respectively, affecting downstream events. A sub-family of bromodomain-containing proteins, including tripartite motif family members TRIM24 and TRIM33, contain an adjacent bromodomain and PHD finger, raising the possibility of combinatorial signalling eliciting a range of responses.⁵ We aim to investigate multi-domain epigenetic readers to gain insights into subtle mechanisms of transcriptional modulation afforded by multiple reader domains in one protein.⁶

1.1 Epigenetics

The term ‘epigenetics’, literally meaning ‘above genetics’, was first introduced by C. H. Waddington in 1942 and defined as “*the causal interactions between genes and their products, which bring the phenotype into being*”.⁷ Historically epigenetics has implied memory of a cellular signal, and the broader use of this word as shorthand for various regulatory systems involving DNA or histone modifications, nucleosome location, or noncoding RNA can be debated.⁸ This terminology has been particularly contentious for histone modifications,⁹ since mechanisms for inheritance of these modifications through cellular division are as yet unclear.¹⁰ However, the fact that histone PTMs provide a mechanism for environmental factors to affect gene transcription, and the fact that at least some PTMs survive through cell division, has linked histone PTMs with a potential molecular mechanism for epigenetic memory.

After a chromatin-based epigenetics meeting, with the intention of creating a consensus definition of ‘epigenetics’, Berger *et al.* concluded that “an epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence”.¹¹ Under this, histone modifications were considered ‘epigenetic maintainers’, which do not possess absolute DNA sequence specificity, and are insufficient for initiation of an epigenetic state. More recently, Arrowsmith *et al.* define epigenetics as “heritable changes in gene expression or phenotype that are stable between cell divisions, and sometimes between generations, but do not involve changes in the underlying DNA sequence of the organism”.³ In an effort to characterise heritability, two separate yeast studies demonstrated that H3K9Me defines a transcriptionally silent state that can be epigenetically inherited, independently of DNA sequence or other positive-feedback loops. Maintenance of this state through many mitotic divisions, and transgenerationally through meiosis, was determined by the persistence of the H3K9Me modification: a result of the activity of the Clr4 methyltransferase and Epe1 demethylase, and dilution of histones during DNA replication.^{12,13} These data validate the consideration of histone PTMs as bona fide epigenetic modifications.

1.2 Chromatin structure and histone modifications

In eukaryotic cells, chromatin is the dynamic storage structure of genetic information, formed from multiple nucleosome units linked by flexible strands of DNA. These nucleosomes comprise two sets of the histone polypeptides H2A, H2B, H3 and H4, with ~146 DNA base pairs wrapped in ~1.7 super-helical turns around this globular octamer (Figure 1.1).¹ This essential feature of packing forms a template which is decorated with various PTMs containing a wealth of information about the activity state of the gene.

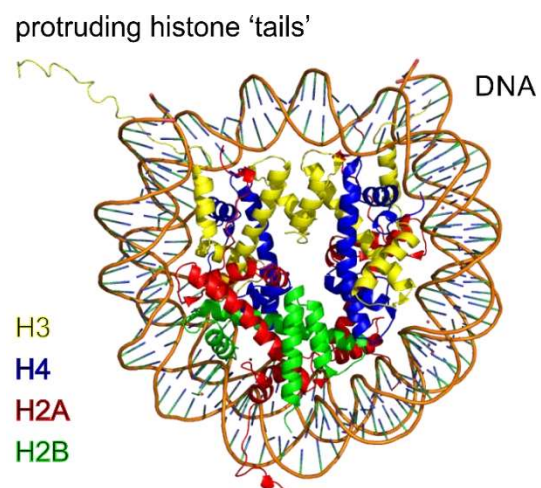


Figure 1.1 Structure of a chromatin unit, displaying internal H3-H4 tetramer complex with accompanying H2A and H2B dimers, as well as protruding histone 'tails' accessible to interaction with epigenetic modulator proteins (PDB: 1AOI).

The depth of the information accessible through chromatin modifications is multifaceted (Figure 1.2). Histone ‘tails’ are susceptible to at least eight distinct modifications, including methylation, phosphorylation, glycosylation and ubiquitination, meaning different groups with a variety of chemical properties can be appended.^{14–16} Additionally, these modifications can occur at multiple positions along a histone tail, including within the histone core, and to several degrees. Lysine residues, for example, can be found unmodified, acetylated (KAc), or in mono-, di- or trimethylated states (KMe₁₋₃). Histone modifications are commonly referred to by the histone subunit, followed by the identity and position of the modified residue e.g. H3K9Me₃ is histone H3 with a trimethylated lysine in the ninth position. This nomenclature will be used throughout.

Lysine acetylation, a ubiquitous PTM, was first applied directly to histone proteins and linked to transcriptional activity in 1988,¹⁷ prompted by chemical studies on acetylation of calf thymus chromatin.¹⁸ The effects of these PTMs were initially thought to be purely electrostatic: lysine acetylation neutralises the charge of the ϵ -nitrogen atom, reducing the interactions with negatively charged DNA phosphate groups. This reduction in charge gives rise to the less tightly packaged form of chromatin, euchromatin, which is associated with active gene transcription. Conversely, lysine bearing an alternative PTM of trimethylation is positively charged, and promotes the tightly packed heterochromatin structure, which is less accessible to transcription factors and hence is associated with gene silencing. Subsequent research has shown that these PTMs have further function, beyond a passive structural role, that also affects gene transcription through direct protein-protein interactions (PPIs). Genomic distributions of these modifications combine to form a complex, context-dependant language,¹⁹ which can manifest phenotypically.

The proposal that a ‘histone code’ exists that regulates activities associated with chromatin, was first presented in 2000.²⁰ It was suggested that a landscape of distinct modifications constitute a ‘map’ which defines spatial and temporal access to the genetic region. Since then the model has developed to describe the storage, interpretation and organisation of genetic information being achieved by the complementary action of ‘writers’, ‘readers’, and ‘erasers’: proteins which add, recognise and remove PTMs, respectively (Figure 1.3).

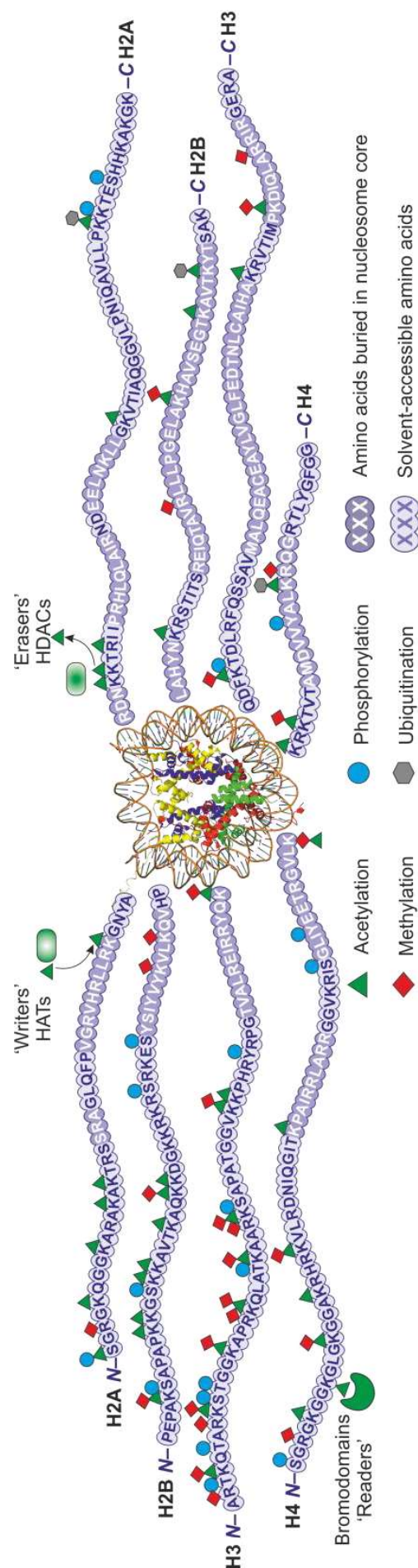


Figure 1.2 Schematic representation of post-translational modifications to histone proteins, displayed as amino acid sequences. Peptide 'tails' protrude through the gyres of DNA wrapped around the histone (central cartoon representation). Modifications occur to both solvent exposed regions and residues buried within the nucleosome core. Acetyl groups are a common modification to the ϵ -nitrogen atom of lysine residues, and are added, recognised or removed by histone acetyltransferases (HATs), bromodomains, and histone deacetylases (HDACs), respectively (shown in green). Additionally, methylation of lysines can occur to multiple degrees and is recognised by PHD fingers. Operation on this 'histone code' has been described as 'write', 'read', and 'erase' action. Reprinted with permission from Jennings *et al.*, Future Medicinal Chemistry, 2014.³⁰ Copyright Future Science Group.

‘Writer’ enzymes place covalent modifications on unstructured *N*-terminal histone ‘tails’, which protrude from nucleosomal DNA through grooves between super-helical segments. Distinctive patterns can relay epigenetic information to

modulators or ‘readers’ which, amongst other functions, control transcription; chromatin markers have been associated with both repression and activation of genes. This in turn can result in the action of ‘erasers’ removing units of modification to adapt to new requirements of the system.

The understanding of enzymes responsible for actively maintaining this dynamic equilibria has advanced, in both breadth of modifications and depth of knowledge in their functionality. Accumulating evidence suggests epigenetic readers may serve to initiate specific downstream events, including transcriptional regulation and DNA repair.²¹ However, the tractability of inhibiting enzymes ‘reading’ the code has only recently come to the forefront, particularly in terms of disease pathology.²² This concept is attractive from a medicinal chemistry perspective as it should, in principle, be possible to interfere with these proteins as an alternative to targeting the proteins whose expression they control.²³

1.3 Epigenetic reader modules

Several protein folds are able to recognise PTMs of chromatin, making them fundamental to transcriptional regulation. These reader domains take part in highly specific interactions with small, chemical modifications to accessible portions of the nucleosomal units, and mistranslated epigenetic signals have been implicated in disease. Complex spatial arrangements strike a fine balance between structural imposition required for specificity and flexibility necessary for diverse readership across the histone code.

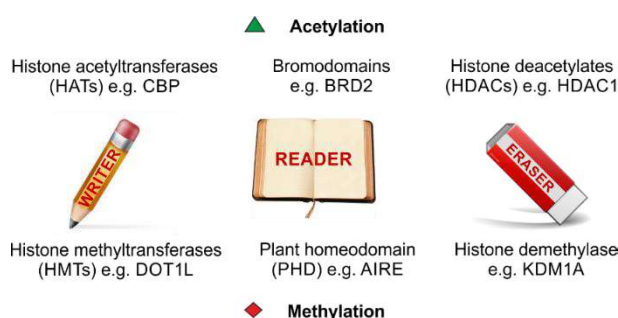


Figure 1.3 The model for the proposed ‘histone code’ for the regulation of gene expression from PTMs. Proteins which add, recognise, or remove modifications are referred to as ‘writers’, ‘readers’, and ‘erasers’, respectively.

1.3.1 Bromodomains

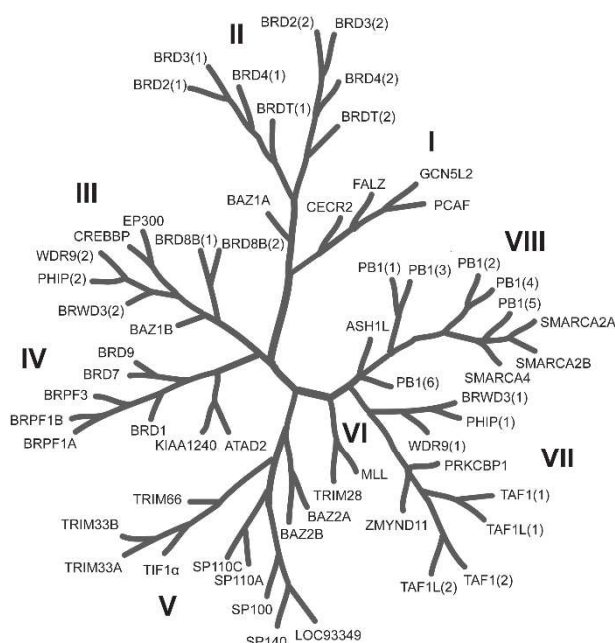


Figure 1.4 Human bromodomain phylogenetic tree constructed from structure based alignments including 61 bromodomains found within 46 proteins. Subdivision into ‘families’ indicated numerically. Reprinted with permission from Hewings *et. al*, J. Med. Chem.¹⁰² Copyright 2012 American Chemical Society.

Bromodomains are an important and well characterised protein domain of ~110 amino acids which bind to KAc modifications on histone tails, and other proteins, acting as readers of lysine acetylation states. Named after the *Drosophila* gene *brahma* where they were first identified,²⁴ there are 61 human bromodomains known within 46 proteins in the human proteome, with some proteins containing more than one distinct bromodomain (Figure 1.4).^{25,26} These have been allocated into eight families based on

three-dimensional structure-based alignments, followed by manual curation of the aligned sequences.²⁷ The most widely studied of these is the bromodomain and extra-terminal domain (BET) family, including the dual bromodomain containing proteins BRD2-4 and BRDT.

A left-handed, antiparallel four-helical bundle (α A, α B, α C, and α Z) dominates the tertiary structure of bromodomains²⁸ and is highly conserved across the family, despite low sequence identity (Figure 1.5: A).²⁹ Specificity for certain acetylated lysine residues is thought to arise, at least in part, from the variable length loops between helices, termed the ZA and BC loops, at the opening to the KAc binding site, which can interact with adjacent residues to observe contextual sequence.²⁷

The acetylated lysine residue binds in a pocket found in bromodomains that is predominantly hydrophobic, and contains four structurally-conserved water molecules which form the base of the pocket (Figure 1.5: B). There are often two additional conserved water molecules that reside in the ZA loop region. KAc donates a hydrogen bond from the ϵ -NH to one of the ZA channel water molecules and accepts a hydrogen bond from a conserved asparagine residue ((Figure 1.5: B –

BRD2(2) N156), that is replaced by a tyrosine or threonine in some bromodomains.²⁷ The KAc also forms a water-mediated hydrogen bond to a conserved tyrosine residue within the pocket (Figure 1.5: B – BRD2(2) Y113). Although these hydrogen-bonding interactions are key for KAc recognition, the inherent affinity a bromodomain for an KAc amino acid is low. In the case of histones, additional affinity comes from binding of the peptide tail at the entrance to the KAc-binding pocket.

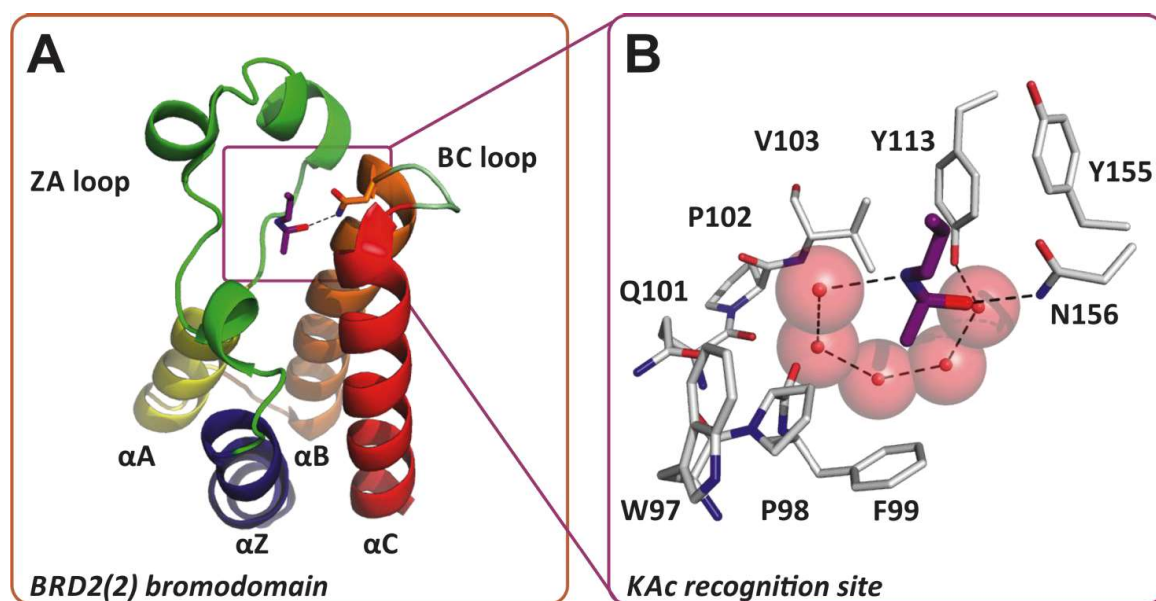


Figure 1.5 A. X-Ray crystal structure of a typical bromodomain (second bromodomain of BRD2) in complex with KAc (carbon = purple, PDB code 2DVQ). The bromodomain structural motif is formed of four helices αZ (blue), αA (yellow), αB (orange), and αC (red) and two loop regions known as the ZA loop (green) and the BC loop (light green). **B. Interactions of a KAc residue in a well-defined hydrophobic pocket.** A highly conserved asparagine residue and several tightly bound water molecules (red spheres), which form the base of the pocket, take part in hydrogen bonding interactions. In BRD2 hydrogen bonding occurs from the carbonyl oxygen to N156 and *via* water to Y113. The hydrophobic entrance to the pocket, W97-P98-F99, is common to BET bromodomains and termed the 'WPF shelf'. Reprinted with permission from Jennings *et. al*, Future Medicinal Chemistry.³⁰ Copyright 2014 Future Science Group.

1.3.2 Plant homeodomain fingers

Methylation of lysine residues comprise an important category of modification, which forms a prevalent 'ON' mark at start sites, as well as effecting gene repression in other contexts. Plant homeodomain (PHD) fingers are zinc finger-containing proteins that recognise lysine methylation states,^{31,32} and have demonstrated greater diversity in residue composition than other epigenetic reader domains. Since identification as a potential domain involved in chromatin-mediated transcriptional regulation in 1995,³³ PHD fingers have been the subject of various investigations into function and biological affinities. Mutations within this motif have been linked to immunological, neurological and developmental disorders.²² There are 90 annotated PHD fingers in the human genome, which can be grouped into 'kinships' based on sequence alignments, groupings which are likely to translate into similar biological activity (Figure 1.6).

Zinc finger-containing proteins are responsible for a range of biological functions, including epigenetic regulation, apoptosis, protein folding and lipid binding, with structures as diverse as their functionality. The classical zinc finger unit is composed of a Cys₂His₂ motif coordinating a Zn²⁺ ion.³⁴ The first natural known variant on this sequence was the friend of GATA (FOG) family CCHC zinc fingers,³⁵ which presented the apparent ability for this robust unit to tolerate chelating sequence variations without gross structural distortions. The PHD motif expands on this evidence, demonstrating that multiple cysteine or histidine substitutions can be made without loss of structural integrity.

Despite low sequence identity and frequent insertion or deletion events, PHD domains retain a consistent structural fold, which is spatially distinct from other similar zinc-binding regions associated with gene regulatory action, such as the Cys₃HisCys₄ RING finger and the Cys₂HisCys₅ LIM domain. The highly conserved coordination of zinc ions (Zn²⁺) by a characteristic Cys₄HisCys₃ motif³⁶ is required for conformational stability in these small, 50-80 residue long domains. There are notable exceptions to the typical architecture of a PHD finger, with several widely accepted domains featuring an alternative arrangement of metal binding residues. For example the

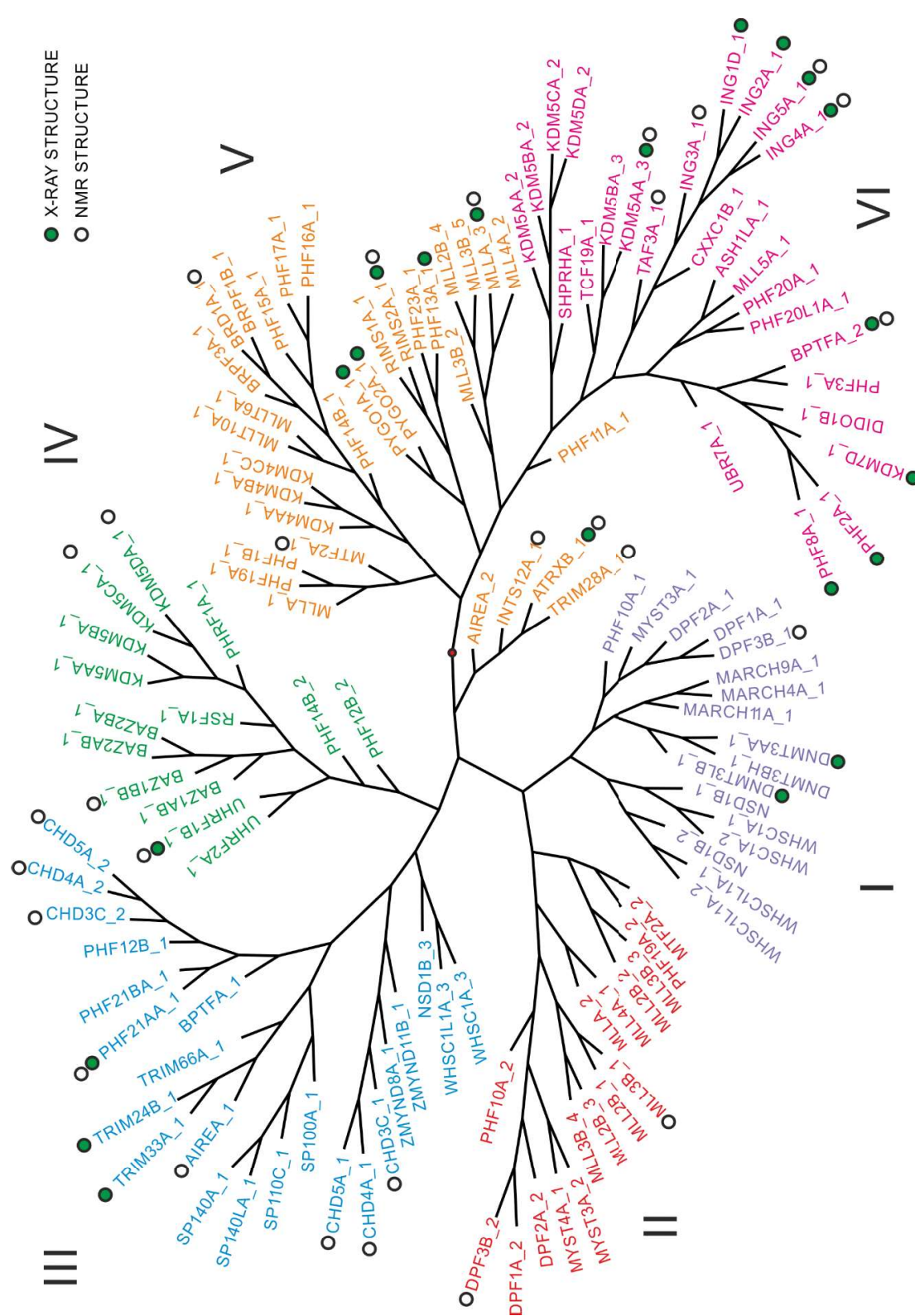


Figure 1.6 Phylogenetic tree for the human PHD family based on multiple sequence alignment. A neighbour-joining phylogenetic tree was generated from the full sequence alignment with ClustalW¹⁰³ and visualised with PhyloDraw,¹⁰⁴ before being manually arranged for clarity. Family divisions into Kinships (numbered I-VI) are presented based on sequence identifiers for each grouping. Structures for PHD domains were collated by a sequence identity based search against the PDB, and the availability of X-ray diffraction or NMR solution structures is indicated.

transcriptional regulator ATRX with a Cys₈ coordination pattern is also frequently referred to as a PHD domain.³⁷

PHD finger topology is dominated by a single super-secondary structural motif of two interlaced zinc fingers in a cross braced arrangement with a core of two antiparallel β -strands (Figure 1.7: A). The lack of an α -helical peptide binding surface is typical of this domain, as is the incorporation of large sections of disorder. The first solution structure of a PHD motif from bromodomain adjacent to zinc finger domain 1B (BAZ1B)³⁸ asserted the domain's reliance on Zn²⁺ ions for expression and organisation, with removal resulting in loss of chemical shift dispersions characteristic of folded proteins (Figure 1.7: B). This structure revealed two loops from R7-I16 and R31-Q43 which presented as likely candidates for PHD-ligand interactions. These two

highly variable loops imbue the protein with a diverse set of readership abilities, demonstrating specificity across a range of substrates.³⁹ PHD fingers have exhibited target interactions with lysine residues bearing a specific level of methylation, from un-methylated (KMe₀) to mono-, di-, and tri-methylated (KMe₁₋₃).

PHD fingers commonly bind methylated lysine residues in a pocket defined by three residues forming an aromatic cage. The diverse composition of this region includes many combinations of phenylalanine, tryptophan, histidine, and tyrosine residues. Positional specificity is achieved *via* an extended recognition surface, including an acidic or polar pocket for the nearby arginine sidechain, which can either be found adjacent (as in the 'RK' sequence beside H3K9/H3K27) or separated by one amino acid (as in the 'RXK' sequence formed from H3R2 preceding H3K4). The N-terminus

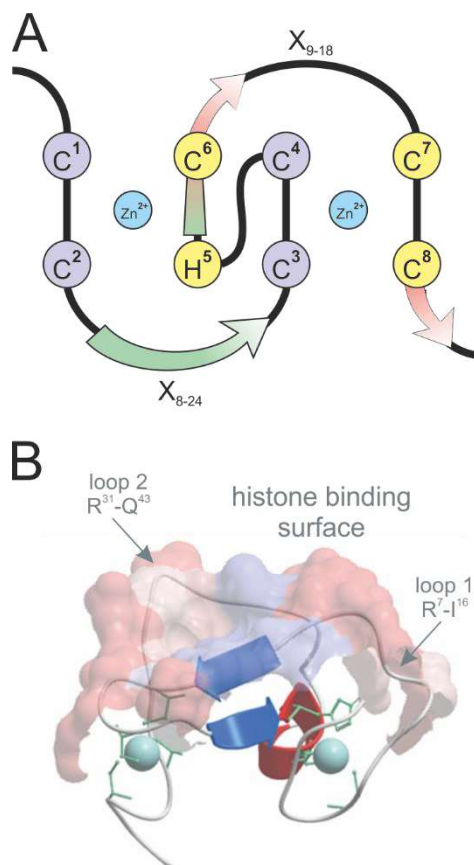


Figure 1.7 A: Metal ligation pattern for general PHD finger; complexation of two Zn²⁺ ions in a cross braced fashion. Arrows represent likely secondary structural regions: red – α -helices, green – β -strands. **B: Solution structure of BAZ1B (WHSF) PHD finger (PDB: 1F62)** demonstrates typical topology. Chains coloured according to secondary structure; histone binding surface coloured by residue hydrophobicity; stick form – metal binding residues.

of a histone chain is cradled in one of the variable loop regions of the PHD finger, encouraging hydrogen bonding and disfavoured elongation of the chain. These features give each specific PHD finger exquisite selectivity of readership, while allowing breadth of recognition across the family.

1.3.3 Combinatorial modifications

Epigenetic reader domains are found in the context of complex protein architectures, and are predicted to work in combination with various other effector domains. Of increasing scientific interest is the local organisation of multiple reader domains.⁴⁰ Increased affinities of reader domains for peptides with multiple modification sites imply a cooperative action in substrate recognition.⁶ It has

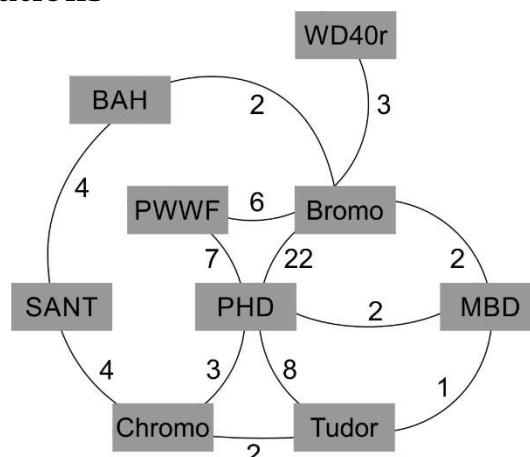


Figure 1.8 Network analysis of human epigenetic reader domain associations in the human proteome. The number of coincidences for any two domains are indicated on connecting lines.

also been suggested that intermediate modification profiles of histones exist as 'ready' states, to facilitate rapid responses to changing environmental conditions.^{5,41} According to the histone code hypothesis, multiple recognition domains are frequent in reader proteins to produce the diversity required for action across the breadth of the epigenome (Figure 1.8), with combinatorial marks that tether a protein partner to a specific nucleosomal region also enabling combined 'read/write' or 'read/erase' actions.^{31,40} The dominance of phenotypic responses from inhibition of proteins containing multiple bromodomains, could also imply that multiple reader domains may be necessary to initiate full phenotypical responses from combinations of modifications. For example use of bromodomain probes without high levels of selectivity over the first or second bromodomains of proteins in the BET family still demonstrate a 'BET-inhibition-like' phenotype.

Positioning of a PHD domain within a gene product is contextually varied. Many appear in close proximity to other histone 'reader' modules including bromodomains, chromodomains and further PHD domains, where cooperative interactions are necessary for functionality (Figure 1.9). Although few links have been made between a PHD finger's wider environment and resulting

function, the high incidence of multiple reader domains suggests an influence of compound or sequential modifications. In the case of DPF3b, tandem PHD domains are necessary for acetylation signal recognition,⁴² although many cases exist where a lone PHD finger is sufficient to recruit other complementary nuclear signalling proteins.

In a large-scale structural analysis of the bromodomain family²⁷ a subclass of domains was characterised as containing tandem PHD and bromodomain units (Figure 1.10). Members of this class exhibit differential recognition depending

on spatial constraints and the relative orientation of recognition sites. The later can dictate *cis*- or *trans*- histone recognition: the PHD and bromodomain in BPTF are linked by a rigid 15 amino

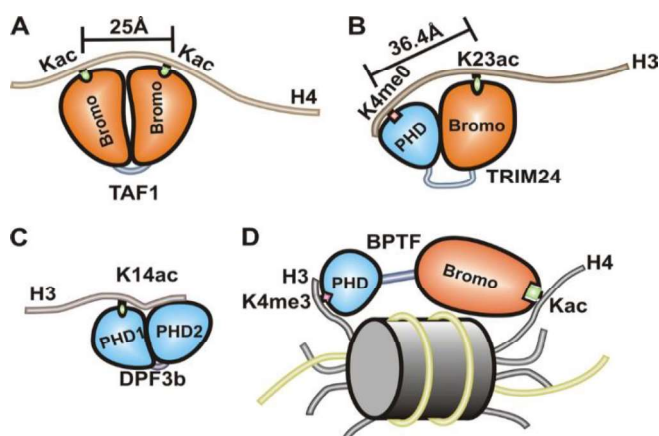


Figure 1.9 Examples of multivalent recognition of histone post translational modifications, by adjacent epigenetic reader domains:¹⁴ **A.** The TAF1 double bromodomain reads two acetylated lysines on the histone 4 tail; **B.** Specificity of modification sites may arise from spatial constraints and inter-site distances;¹⁶ **C.** Combinations of modules can provide alternative functionality;¹⁷ **D.** Anti-parallel arrangement of binding sites in BPTF, caused by a short rigid linker, forces *trans*-histone recognition. Reprinted with permission from Wang and Patel, J Biol Chem.⁶ Copyright 2011 American Society for Biochemistry and Molecular Biology.

acid long α -helix, which enables recognition of methylation and acetylation on different core histone peptides H3K4Me₃ and H4KAc;⁴³ TRIM24, with domains linked by a hydrophobic core of the α z helix, recognises modifications on the same histone H3K4Me₀ and H3K23Ac. Additionally, alignment of peptide orientation affects combined readouts: modifications in positions H3K4 and H3K23 are spatially aligned, whereas lysines H3K4 and H3K27 are presented on opposite sides of the peptide.

With progress in the epigenetic field from inhibition of enzymatic effectors, towards inhibition of reader domains,³ inhibition of multivalent modifications presents itself as a challenging future direction. Since two HDAC inhibitors (romidepsin and vorinostat) have been approved for use in refractory cutaneous and peripheral T cell lymphoma,⁴⁴ five bromodomain inhibiting compounds have been registered for clinical trials. Concurrent inhibition of several reader domains could result in a novel mode of action, with the possibility of discovery of therapeutic benefits.

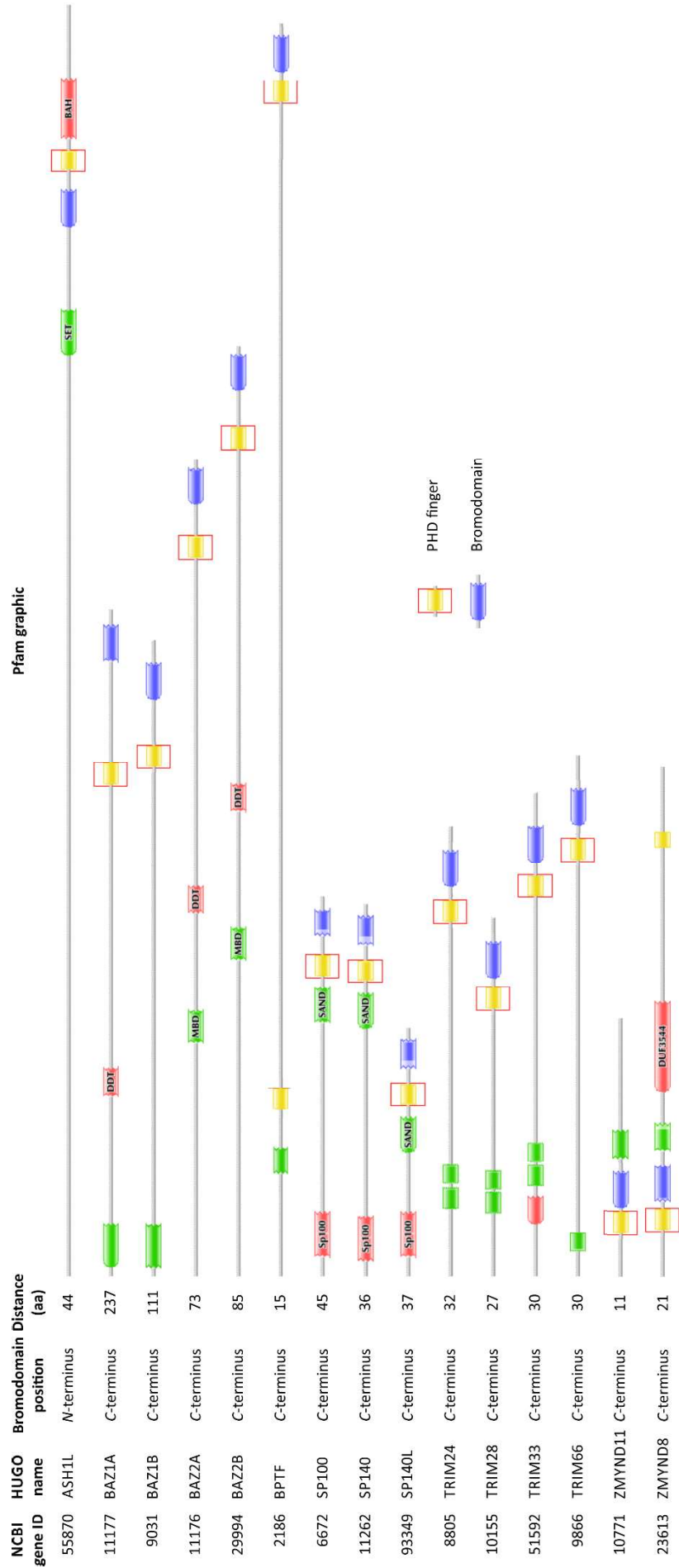


Figure 1.10 Pfam domain architectures of PHD (yellow) containing proteins with adjacent bromodomains (blue). Proteins named by Human Genome Organisation (HUGO) nomenclature. Collated from proteins with gene annotations, and predicted PHD domains based on sequence. Distance between the PHD and bromodomain are indicated (aa – number of amino acids).

1.4 TRIM24 and TRIM33

Tripartite motif-containing (TRIM) TRIM24

proteins contain an *N*-terminal TRIM33

tripartite motif (RBCC), consisting of a RING finger domain involved in

ubiquitination,^{45,46} one or two B-box zinc finger domains which are often

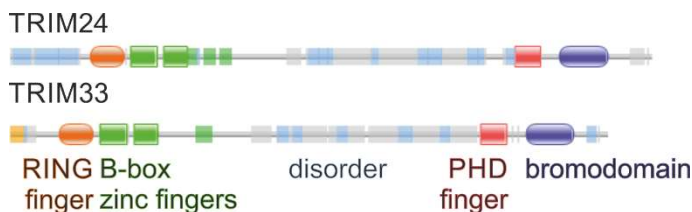


Figure 1.12 Pfam domain architecture of TRIM24 and TRIM33. *N*-terminal tripartite motif (RBCC): RING finger domain, two B-box zinc finger domains, and a coiled coil region. Proximal PHD fingers and bromodomains are found at the *C*-terminus.

found in transcription factors, and a coiled coil region (Figure 1.12). In addition to these domains, TRIM proteins also contain a conserved 25 amino acid-long, tryptophan- and phenylalanine-rich sequence downstream of the coiled coil motif. Associated PHD and bromodomains are a feature often found in the *C*-terminus. TRIM24, TRIM28, TRIM33 and TRIM66 form regulatory complexes which respond to non-canonical histone signatures.⁴⁷ These proteins also form part of the transcription intermediary factor 1 family and are otherwise known as TIF1 α , β , γ , and δ respectively. Members of this family have been related to transcriptional activation necessary for embryonic development, tissue homeostasis, and wound repair.^{48,49,50} TRIM33 was first identified as a novel member of the human TIF1 gene family,⁵¹ with 48% overall identity with respect to TRIM24. The longer α isoform of this protein contains a 17 amino acid insertion, with the shorter β isoform bearing more similarity to other bromodomain containing members of the TRIM family (Figure 1.11).

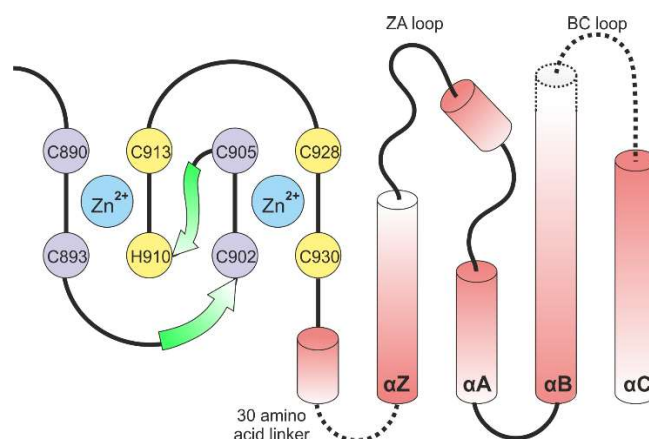


Figure 1.11 Schematic representation of topology of the tandem PHD bromodomain module in TRIM33 α , as observed in the single crystal x-ray structure (PDB: 3U5O). Residues responsible for zinc ligation are highlighted. Notably, when compared to other members of the TRIM family, this α -isoform has a 17 amino acid insertion which extends the BC loop, but many of these residues could not be located due to disorder. α -Helices and β -sheets are represented as red cylinders and green arrows respectively.

The tandem PHD-bromodomain cassettes of TRIM24 and TRIM33 have been shown to interact with histone-like peptides displaying a specific combination of unmodified, trimethylated and acetylated lysines (Figure 1.13: B).⁵ PHD binding of *N*-terminal H3 residues is optimal for affinity, with key interactions formed to lysines in positions 4 and 9 (Figure 1.13: C). Recognition of the K4Me₀ position, through coordination in an acidic pocket formed of glutamate and aspartate residues and the formation of an antiparallel β -sheet with a segment of the PHD finger, is common to members of the TIF1 α -like family.⁴⁶ Rather than the usual aromatic cage found in PHD fingers for KMe₃ recognition,¹⁸ a highly conserved tryptophan is involved in a cation- π interaction with the charged amine of a modified K9Me₃. Binding of KAc residues to TRIM bromodomains occurs in a typical fashion (Figure 1.13: D): the KAc residue binds in a well-defined hydrophobic pocket, forming water mediated hydrogen bonding interactions with a tyrosine residue (Y935 in TRIM24; Y993 in TRIM33). A direct hydrogen bond is formed to a highly conserved asparagine in TRIM24 (N980), but this interaction is not apparent for TRIM33 (N1039). Sequence recognition specificity of TRIM33 for H3K18Ac is contributed to by interactions between a glutamate (E981) found on the ZA loop at the entrance to the pocket and the nitrogen in the adjacent arginine residue in the cognate histone peptide (H3R17).

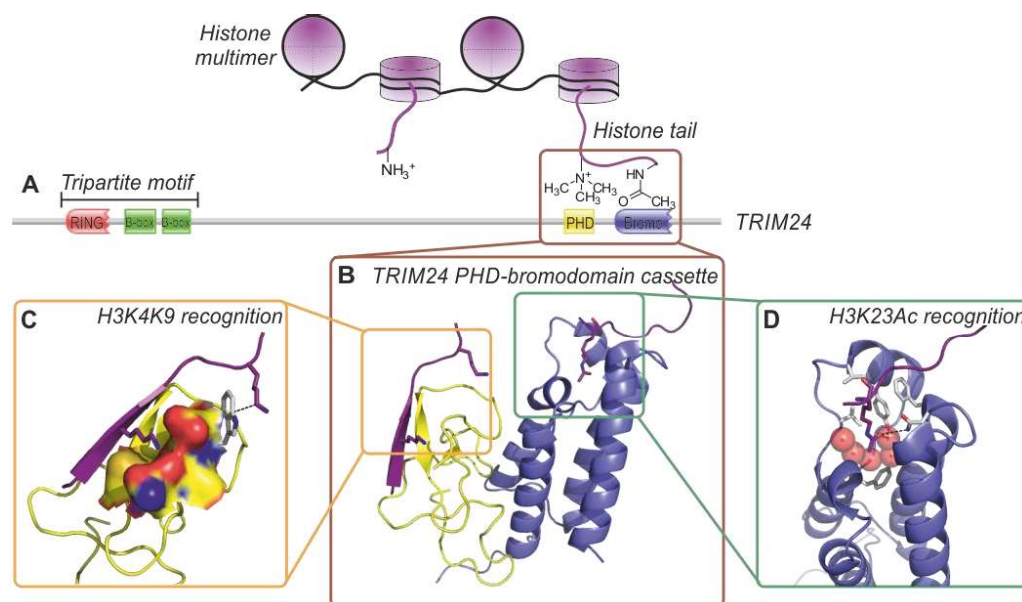


Figure 1.13 TRIM24 histone recognition pose (PDB: 3O34, 3O37). **A:** Domain architecture, represented by Pfam prediction of domains. **B:** Recognition of multiple modification sites on a H3 histone tail – lysine residues 4, 9, and 23. **C:** Surface of interaction with PHD domain, key interactions with K4 and K9 indicated. Potential for a cation- π interaction between trimethylated K9 and W828. **D:** Stabilising hydrogen-bonding recognition of acetylated K23 in the bromodomain binding pocket, including bridging water molecules.

1.5 Small molecule inhibition of reader domains

Given the dynamic plasticity of the epigenetic landscape, studying mechanisms by molecular biology techniques alone is insufficient. Small molecule probes are important tools in providing a more comprehensive understanding of epigenetic control. The recent development of many ligands for epigenetic reader domains have demonstrated the existence of defined surfaces which make them more ‘druggable’ than many classical PPIs.

1.5.1 Bromodomains

Some of the earliest work in this field focused on the p300/CBP-associated factor (PCAF)/Tat-K50Ac interaction which had been implicated in HIV,⁵² including the first assay used to quantify binding affinities to a bromodomain: an ELISA after NMR based chemical screening.⁵³ Since then, work has been carried out towards inhibition of a wide range of bromodomains, mostly focused on the BET family,⁵⁴ although recently an increasing number of ligands have targeted non-BET bromodomains including CREBBP,^{55,56} TRIM24,^{57,58} and BRD7/9.^{59,60}

An investigation was conducted on the tractability of developing small molecule inhibitors of 33 non-BET bromodomain KAc-binding sites by Vidler *et al.* using available structural information from 105 PDB entries.⁶¹ This used SiteMap, software that combines binding site identification with druggability assessment, based on physicochemical properties: size, degree of enclosure by the protein, hydrophilicity and hydrophobicity.⁶² Grouping according to common binding site features suggested possibilities for developing compounds that are selective for one bromodomain over another. Rather than any quantitative measure, the resultant druggability score (Dscore) was able to rank bromodomains in order of their relative druggability. Although, as pointed out by Zhang *et al.*, this approach is heavily dependent on the extent of our current knowledge of a bromodomain including prior ligand-bound structural availability,⁶³ and also discards the flexibility of bromodomains.⁶⁴ However, the only bromodomain which demonstrated a significant change in predicted druggability between the *apo* and ligand bound form was CREBBP: a result of the induced fit pocket highlighted by the structure of a dihydroquinazolinone bound to CREBBP.⁵⁵

In comparison to common PPIs with less well defined binding pockets (surface area $>1000 \text{ \AA}^2$), all bromodomains were classed as more druggable: 13 bromodomains, including the BET family, were classified as druggable (Dscore >0.85); five additional bromodomains were viewed as being intermediate in their druggability – BRD1, BRD9, CREBBP and two bromodomains of PB1; the remaining bromodomains, including TRIM24, were classified as being difficult (Dscore <0.75). Since there would be no reason to assume the therapeutic potential of a bromodomain would relate to druggability, rather than to disregard certain members, the intention of the study was to form realistic expectations of hit rates before any screening effort is undertaken. This may affect the diversity of a screening library, screening concentrations, or acceptable binding thresholds.

1.5.2 PHD fingers

With PHD fingers implicated in various human diseases, including cancer, immunological, and neurological disorders,⁶⁵ it is likely that specific inhibition of these domains will provide insight into disease mechanisms, presenting itself as a potential new therapeutic strategy.⁶⁶ However, there are currently many fewer ligands for these domains than for bromodomains.

A structural analysis of readers of methyl-lysine modifications, also used SiteMap to calculate Dscores for 42 different domains, covering 10 classes of structural fold.⁶⁷ The predicted druggability of these domains was more variable than that of the BET bromodomains (included as a reference) and was also affected by greater conformational changes for ligand bound structures.

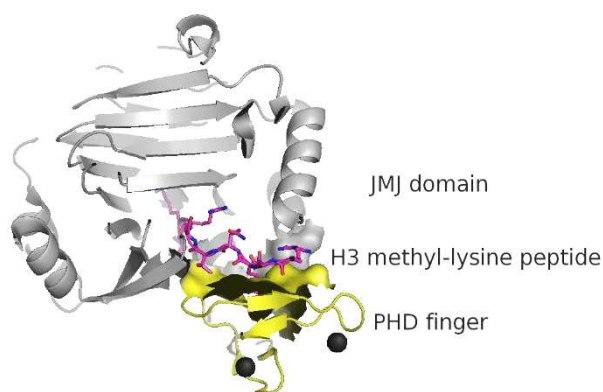


Figure 1.14 ‘Druggable’ site in peptide bound conformation of PHF8 (PDB: 3KV4). PHD finger (yellow) recognises a H3K4Me₃K9Me₂ peptide (purple) in a cleft formed with the adjacent JMJ domain (grey). SiteMap Dscore = 1.05.

For example the third PHD finger of MLL was undetectable by SiteMap in the *apo* form, but a loop formed upon methyl-lysine binding creates a highly druggable pocket (Dscore 0.97). Another PHD finger which stood out in this analysis of peptide bound conformations was that of the demethylase PHD finger protein 8 (PHF8) with an

enclosed cleft formed between the PHD finger and adjacent JMJ domain possessing a Dscore of 1.05 (Figure 1.14). This demonstrates that interactions between structural modules can have a significant impact on the calculated druggability of a binding site.

The third PHD finger of lysine-specific demethylase 5A (KDM5A aka JARID1A) has been implicated in acute myeloid leukemia (AML).⁶⁸ Analogues of the anti-arrhythmia drug amiodarone were shown to inhibit the KDM5A PHD3- H3K4Me₃ interaction with IC₅₀ values of 25-40 μ M, and also demonstrated selectivity over several other PHD containing proteins (Figure 1.15: A).⁶⁹

Many interactions of small molecule ligands with PHD fingers were characterised during the development of ligands for L3MBTL1, a member of the malignant brain tumor (MBT) family containing methyl-lysine readers. Nicotinamide UNC280 was also shown to inhibit the recognition of a H3K9Me₁ peptide by PHD finger protein 13 (PHF13) with an IC₅₀ of 48 $\pm$ 4 μ M (Figure 1.15).⁷⁰ Optimisation of this ligand for L3MBTL1 selectivity appeared to remove affinity towards PHD fingers (represented by PHF23 and KDM5A).⁷¹

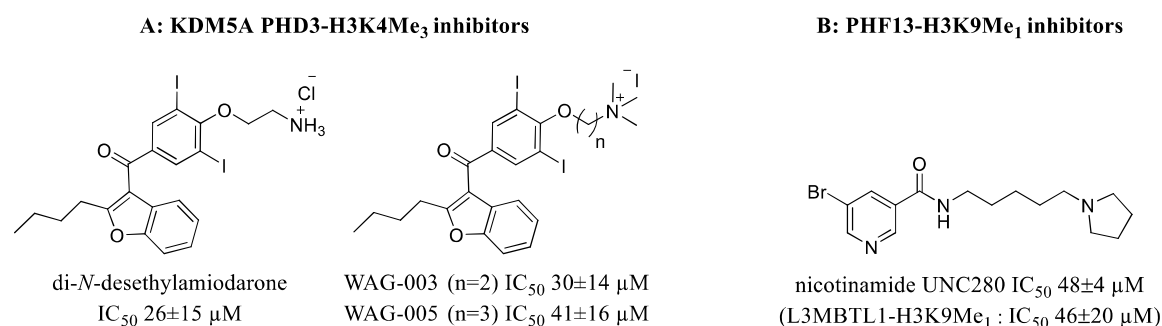


Figure 1.15 Molecules identified as inhibitors of PHD finger recognition of methylated lysines in histone peptides.^{70,71} IC₅₀ values given for protein-peptide recognition partners above, unless stated otherwise.

1.6 Utility of assays towards epigenetic reader inhibition

Validation of assays for proteins without enzymatic ability is challenging, particularly when binding partners are not well established. However, combinations of biophysical assays have successfully enabled identification and characterisation of lead molecules for epigenetic reader targets, leading to improvements in affinity and selectivity.

The elucidation of bromodomain function in KAc recognition, including the use of pull down assays to introduce histone selectivity to recognition by bromodomains in transcriptional co-activators,⁷² led to several assays based on displacement of a known binding partner, often a modified peptide based on the endogenous protein binding partner. These require use of a competitor peptide with sufficient affinity to enable signal detection, and subsequent studies have shown decreases in signal noise and increased accuracy of quantification after optimisation of competitor peptides,⁷³ or choice of buffer.⁷⁴

The availability of a range of assays to identify bromodomain inhibitors has been crucial to the rapid progress in this area. Most successful campaigns have been characterised by tiered complementary screening methods, to enable fast identification and validation of hits. Even within techniques, assay protocols may vary significantly due to differences in protein and detection parameters, hence SAR analysis is often best performed within a single comparative set which is subjected to the same conditions.⁷⁵

1.7 Assays used in the discovery of bromodomain inhibitors

Label free methods enable characterisation of ligands for targets without known native binding partners: the first small molecule inhibitors of bromodomains were discovered through nuclear magnetic resonance (NMR)-based chemical screening monitoring 2D ¹⁵N HSQC spectra.⁷⁶ Operationally simple differential scanning fluorimetry (DSF) assays allow rapid progression to screening platforms and have been used throughout ligand identification campaigns;⁷⁷ Surface plasmon resonance (SPR) has primarily been used as a secondary assay for hit validation.^{78,79}

Amplified Luminescent Proximity Homogeneous Assay Screen (AlphaScreen) is an indirect proximity based fluorescence assay with the ability to detect large protein interactions at distances of up to 200 nm. It can measure a broad range of affinities between pico- to millimolar. With the advantage of high-throughput and low resource requirements, AlphaScreen is frequently used as a primary screening method, despite a high rate of artefacts.⁸⁰

A fluorescence anisotropy (FA) assay developed for BET bromodomains by GSK provided a robust, low-resource and sensitive platform able to identify weak fragment hits,⁸¹ and to characterise naphthyridines as BET bromodomain inhibitors.⁸² Subsequently, FA has been used for identification and evaluation of inhibitors of the PCAF bromodomain, with IC₅₀ values that correlate to *in vitro* anti-HIV-1 activity.⁷⁴ Isothermal titration calorimetry (ITC) is still considered the ‘gold standard’ for quantifying binding affinity,⁸³ directly measuring thermodynamic contributions to affinity, and has frequently been used to characterise affinities of both peptides⁴⁷ and novel ligands⁸⁴ and determine the selectivity of ligands between bromodomains.

Cellular activity has been important for both preliminary studies into natural recognition modes of bromodomains, and demonstration of therapeutic potential for targets of probe compounds. Importantly, these demonstrate interactions of epigenetic reader proteins, and the efficacy of their inhibition, in an intact nucleosome environment. Visualisation of histone binding by BET proteins in cells was first achieved by fluorescence resonance energy transfer (FRET) combined with flow cytometry to assess selectivity.⁸⁵ The ability of the BET inhibitor (+)-JQ1 to compete with histone binding in a cellular environment was established through fluorescence recovery after photobleaching (FRAP),⁷⁷ an assay which has demonstrated the cellular efficacy of inhibitors of a diverse set of bromodomain targets.⁸⁶ Following identification by DSF, and IC₅₀ determination by AlphaScreen, the ability of triazolophthalazine bromodomain inhibitors to displace CREBBP from chromatin in cells was also assessed using FRAP.⁸⁷

Overall, the field demonstrates the utility of rapid hit identification using indirect assays, which provide information on the displacement of competitor ligands with a known binding site, followed by the importance of confirming target protein binding early in the ligand development process using direct assays. Further characterisation and optimisation of ligand binding can be achieved using combinations of biophysical and cellular assays, including screening panels and more precise direct assay techniques like ITC and NMR.

Table 1.1 Summary of biophysical assays employed in the development of bromodomain inhibitors. For details of the techniques, refer to Chung and Witherington 2011.⁸⁸

	Assay	Requirements	Throughput	Information	Notes
Label free (direct)	nuclear magnetic resonance (NMR)	¹⁵ N labelled protein Full NMR assignment High protein consumption	Medium: 100s	Affinity, especially weak: mM-nM range Residue resolved map of binding Integrity of protein	Water filled pockets reduce sensitivity (e.g. PCAF)
	differential scanning fluorimetry (DSF)	Few Stability in buffer conditions	High: 1000s	Measures protein stability: proportional to thermodynamics of interaction Detects all stabilising interactions, even if not inhibitory (up to low μ M)	Temperature dependence can result in false negatives
	surface plasmon resonance (SPR)	Functional immobilised protein (or ligand partner when used in competitive format) Low protein requirements	High: 1000s	Affinity and kinetics Signal is proportional to size ratios (protein:ligand)	Can perform sequential runs
	isothermal titration calorimetry (ITC)	Approximate affinity range for experimental design	Low: 10s	Direct binding observation Thermodynamic parameters: affinity, enthalpy, stoichiometry. Gibbs free energy and entropy subsequently determined	Highly sensitive to DMSO Challenging sample preparation
Competition based (indirect)	amplified luminescent proximity homogeneous assay screen (AlphaScreen)	Known binding partner Labelled protein and ligand to match donor acceptor beads (e.g. His ₆ for nickel chelating, biotin for streptavidin)	High: 10,000s	Chemiluminescent signal Can detect weak interactions: high sensitivity, low background	Avidity resulting from bead based technology High false negative rate
	fluorescence anisotropy (FA)	Suitable ligand (without interference in fluorescence) Knowledge of structure for fluorescent labelling	Medium: 100s	Kinetics (binding constants) Signal dependent on size (affecting rate of tumbling) < μ M range	Good correlation observed for <i>in vitro</i> ApoA1 reduction

1.8 Computational modelling

Fuelled by the increasing availability of high-resolution crystal structures of epigenetic reader domains, the development of inhibitors of these targets has been greatly expedited by the increasing use of computational techniques – molecular dynamics (MD),^{55,89} docking,⁹⁰ and quantum mechanics (QM) – which allow interrogation of a structure, and any associated small molecules.

Due to the structural similarities among KAc binding sites across bromodomain families, the design of selective inhibitors remains challenging,⁶³ increasing the importance of using computational approaches to analyse the key differences between these proteins. In order to facilitate the search for new inhibitors several databases have been developed to provide information collated from papers and public databases regarding epigenetic reader domains, including binding affinities, protein-ligand interactions, and structures. However, curation of these resources is often inconsistent: ChEpiMod appears to no longer be supported at its original domain and has poor search reliability,⁹¹ but the more recent EpiDBase does provide current links to many common directories including ChEMBL and DrugBank.⁹² The later also highlights the high proportion of research towards bromodomain ligands relative to PHD fingers, with 1393 entries across 25 bromodomains and only 114 entries for three PHD fingers (redundancy not removed). A Chemical Probes Portal (<http://www.chemicalprobes.org/>) has been established to provide community driven information on the appropriate use of chemical probes, although this currently only contains seven molecules.⁹³

Docking studies offer a much faster means to predict ligand-protein interactions than by MD simulation, encouraging their use in the creation of bespoke fragment-based libraries directed

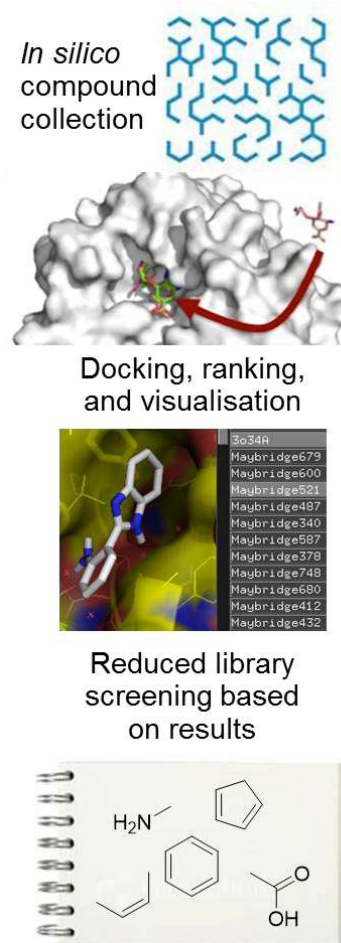


Figure 1.16 General docking methodology for reduction of a large compound collection to a small selection for focussed screening.

towards a target. Starting from a large library of more than 9000 compounds, Caflisch and co-workers used a fragment-based high-throughput docking approach, Anchor-based Library Tailoring (ALTA), to select 24 compounds for biological testing using DSF and identified 4 compounds which stabilise BRD4(1).⁹⁴ Vidler *et al.* took a virtual screening approach to identify fragments predicted to bind to BRD4(1). 143 of these compounds were assessed by AlphaScreen, and six novel hit compounds were identified, four of which included unprecedented KAc-mimicking groups.⁹⁰ Muvva *et al.* used an *in silico* approach to screen compounds from the NCI Diversity, Drug Bank and Toslab databases for predicted BRD4(1) affinity.⁹⁵ Although a number of predicted ligands were identified, no experimental evidence for BRD4(1) binding was presented.

1.9 Clinical applications

While assays provide the fundamental tools for compound identification and development, applications can only be confirmed after progression to clinical models and trials. Promising effects of BET bromodomain ligands have been observed in a range of pathologies, inducing very strong phenotypic responses across several indications,⁹⁶ including NUT midline carcinoma (NMC),⁷⁷ osteosarcoma,⁹⁷ glioblastoma multiforme,⁹⁸ and chronic inflammation.⁹⁹ It has been suggested this is due to the role of bromodomains in maintenance of cellular state, rather than activation of specific pathways.¹⁰⁰

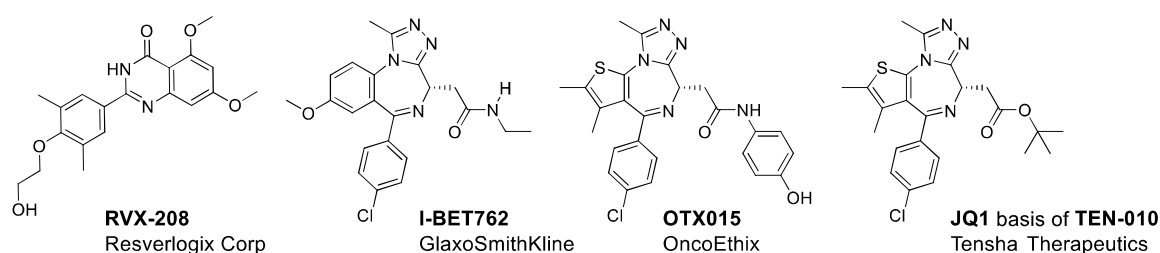


Figure 1.17 Several bromodomain ligands currently in clinical trials. JQ1 represents TEN-010: an optimised analogue. In addition to these the structure of CPI-0610 (Constellation Pharmaceuticals) remains undisclosed.

Several ligands of bromodomains have progressed to clinical trials, which should help identify benefits of this clinically-novel mode of action (Figure 1.17). Although not developed as a bromodomain inhibitor, RVX-208 became the first BET inhibitor in clinical trials investigating the

effect of the compound on atherosclerosis, coronary artery disease, dyslipidemia and type II diabetes (ClinicalTrials.gov identifiers: NCT01058018, NCT01067820, NCT01423188, and NCT01728467). I-BET762 (GSK525762) is currently undergoing two Phase 1 clinical trials, one for NUT midline carcinoma and the other for relapsed refractory hematological malignancies (ClinicalTrials.gov identifiers: NCT01587703 and NCT01943851). OTX015 is undergoing trials in acute leukemia or other hematological malignancies and selected solid tumors (ClinicalTrials.gov identifiers: NCT01713582 and NCT02259114) with further trials scheduled for patients with recurrent glioblastoma multiform, or acute myeloid leukemia in combination with azacitidine (ClinicalTrials.gov identifiers: NCT02296476 and NCT02303782). Another BET inhibitor CPI-0610 (Constellation Pharmaceuticals: structure undisclosed) is recruiting for trials in patients with progressive lymphoma, previously treated multiple myeloma, or other previously treated acute leukemia or myelodysplasia (ClinicalTrials.gov identifiers: NCT01949883, NCT02157636, and NCT02158858). TEN-010 (Tensha Therapeutics: structure undisclosed) is a JQ1 analogue, also known as JQ2, in early clinical development for oncology, with effects attributed to Myc inhibition. The compound will be administered subcutaneously in patients with advanced solid tumour malignancies (ClinicalTrials.gov identifier: NCT01987362).

Given the relatively sparse knowledge of biological pathways involving other epigenetic reader domains – in some cases little is known beyond a genetic prediction of a domain – and associated difficulties in prediction of phenotypic consequence,⁶⁶ it seems likely that determining any further clinical benefits of inhibition of histone readership will require deliberate attempts to generate probes for novel targets. However, recent developments in the field suggest that the increased risk of focusing on targets without identified therapeutic potential is mitigated by the rapid success in developing BET bromodomain inhibitors, coupled with the emerging importance of epigenetic reader domains in a range of biological processes.

1.10 Aims

The aim of this study was to probe the readership mechanisms of the tandem PHD and bromodomains of TRIM24 and TRIM33, to gain insight into mechanisms in epigenetic control. The development of small molecule inhibitors of PHD fingers is a challenging area, due to their extended peptide recognition surface and diverse interactions, with few currently known. However, in combination with an adjacent bromodomain, the new binding surface between structural units forms a cleft which could provide a more ligandable target, as has been observed for the PHF8 PHD-JMJ cassette. The ability to inhibit each reader domain separately or concurrently could give therapeutic candidates specificity for mode of action, rather than target.¹⁰¹

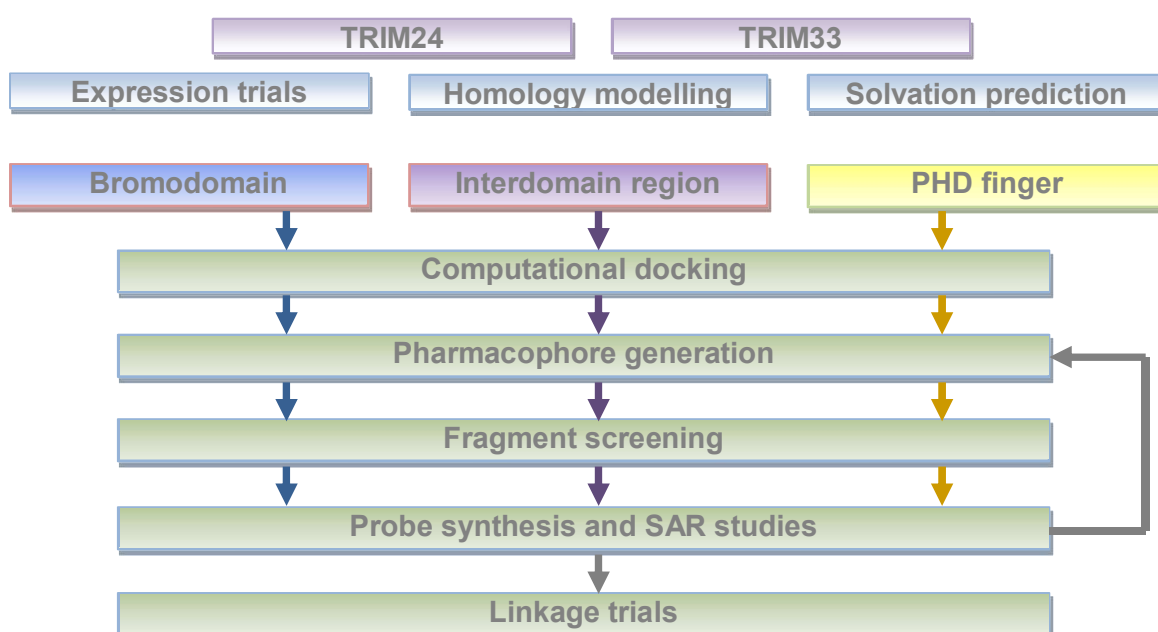


Figure 1.18 Inhibition strategy for TRIM24 and TRIM33. Concurrent screening will commence with expression and purification, and target familiarisation, including modelling and solvation predictions. Constructs containing both the bromodomain and PHD finger will present an additional interdomain region for inhibition. A cascade of *in silico* screening, followed by *in vitro* screening will lead towards probe development, for each domain in isolation, or in combination.

With an initial target of TRIM24, our convergent approach combined computational studies with refinement of expression systems and assay development, to enable design and synthesis of ligands (Figure 1.18). *In silico* screening based on available crystal structures facilitated lead identification through focused *in vitro* screens, which were carried forward through validation to provide lead molecules for structure activity relationship studies. Extension of this methodology towards the homologous TRIM33 expedited progress for this novel target.

Research could foreseeably extend towards tethered inhibitors, allowing binding to a bromodomain to act as an anchor for interactions with the PHD finger. This could incorporate a higher level of specificity for a given target reader module, and allow greater control for analysing phenotypical results of epigenetic marks.

The work described in this thesis aimed to address the following questions, which arise from the target proteins and the possibility of multivalent epigenetic recognition:

- Are there differential binding patterns of recognition between the homologous proteins TRIM24 and TRIM33?
- Can a computational approach enable identification of small molecule ligands for TRIM24 or TRIM33?
- Would ligands binding to the tandem PHD-bromodomain cassette of TRIM24 or TRIM33 inhibit both methyl- and acetyl-lysine binding?
- Are ligands of tandem domains able to capitalise on context by binding in the interdomain region, and does this effect their inhibition potential?
- Can we develop methodology to facilitate competition-based assays for bromodomains whose binding partners are either unknown, or of prohibitively low affinity?

1.11 References

- (1) Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- (2) Jenuwein, T., and Allis, C. D. (2001) Translating the histone code. *Science* (80-.). 293, 1074–1080.
- (3) Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K., and Schapira, M. (2012) Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.* 11, 384–400.
- (4) Dawson, M. A., Kouzarides, T., and Huntly, B. J. P. (2012) Targeting Epigenetic Readers in Cancer. *N. Engl. J. Med.* 367, 647–657.
- (5) Xi, Q., Wang, Z., Zaromytidou, A.-I., Zhang, X. H.-F.-F., Chow-Tsang, L.-F., Liu, J. X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., Studer, L., Mark, W., Patel, D. J., Massagué, J., and Massague, J. (2011) A Poised Chromatin Platform for TGF-beta Access to Master Regulators. *Cell* 147, 1511–1524.
- (6) Wang, Z., and Patel, D. J. (2011) Combinatorial Readout of Dual Histone Modifications by Paired Chromatin-associated Modules. *J. Biol. Chem.* 286, 18363–18368.
- (7) Waddington, C. H. (2012) The epigenotype. 1942. *Int. J. Epidemiol.* 41, 10–13.
- (8) Riddihough, G., and Zahn, L. M. (2010) What Is Epigenetics? *Science* (80-.). 330, 2010–2010.
- (9) Ptashne, M. (2013) Epigenetics: core misconception. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7101–3.
- (10) Probst, A. V, Dunleavy, E., and Almouzni, G. (2009) Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* 10, 192–206.
- (11) Berger, S. L., Kouzarides, T., Shiekhata, R., and Shilatifard, A. (2009) An operational definition of epigenetics. *Genes Dev.* 23, 781–783.
- (12) Ragunathan, K., Jih, G., and Moazed, D. (2015) Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* (80-.). 348, 132–135.
- (13) Audergon, P. N. C. B., Catania, S., Kagansky, A., Tong, P., Shukla, M., Pidoux, A. L., and Allshire, R. C. (2015) Restricted epigenetic inheritance of H3K9 methylation. *Science* (80-.). 348, 132–135.
- (14) Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* 128, 693–705.
- (15) Bycroft, M. (2011) Recognition of non-methyl histone marks. *Curr. Opin. Struct. Biol.* 21, 761–766.
- (16) Schones, D. E., and Zhao, K. (2008) Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* 9, 179–191.
- (17) Hebbes, T. R., Thorne, a W., and Crane-Robinson, C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* 7, 1395–1402.

- (18) Marushige, K. (1976) Activation of chromatin by acetylation of histone side chains. *Proc. Natl. Acad. Sci. U. S. A.* 73, 3937–3941.
- (19) Gardner, K. E., Allis, C. D., and Strahl, B. D. (2011) OPERating ON Chromatin, a Colorful Language where Context Matters. *J. Mol. Biol.* 409, 36–46.
- (20) Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications. *Nature* 403, 41–45.
- (21) Oliver, S. S., and Denu, J. M. (2011) Dynamic Interplay between Histone H3 Modifications and Protein Interpreters: Emerging Evidence for a “Histone Language.” *ChemBioChem* 12, 299–307.
- (22) Baker, L. a, Allis, C. D., and Wang, G. G. (2008) PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* 647, 3–12.
- (23) Muller, S., Filippakopoulos, P., and Knapp, S. (2011) Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* 13, e29.
- (24) Haynes, S. R., Dollard, C., Winston, F., Beck, S., Trowsdale, J., and Dawid, I. B. (1992) The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res.* 20, 2603.
- (25) Jeanmougin, F., Wurtz, J. M., Le Douarin, B., Chambon, P., and Losson, R. (1997) The bromodomain revisited. *Trends Biochem. Sci.* 22, 151–153.
- (26) Filippakopoulos, P., and Knapp, S. (2012) The bromodomain interaction module. *FEBS Lett.* 586, 2692–2704.
- (27) Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C. H., Knapp, S., and Mueller, S. (2012) Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* 149, 214–231.
- (28) Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., and Zhou, M. M. (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491–496.
- (29) Zeng, L., and Zhou, M. M. (2002) Bromodomain: An acetyl-lysine binding domain. *FEBS Lett.* 513, 124–128.
- (30) Jennings, L. E., Measures, A. R., Wilson, B. G., and Conway, S. J. (2014) Phenotypic screening and fragment-based approaches to the discovery of small-molecule bromodomain ligands. *Future Med. Chem.* 6, 179–204.
- (31) Biez, M. (2006) The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* 31, 35–40.
- (32) Mellor, J. (2006) It takes a PHD to read the histone code. *Cell* 126, 22–24.
- (33) Aasland, R., Gibson, T. J., and Stewart, A. F. (1995) The Phd Finger - Implications for Chromatin-Mediated Transcriptional Regulation. *Trends Biochem. Sci.* 20, 56–59.

- (34) Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003) Structural classification of zinc fingers. *Nucleic Acids Res.* 31, 532–550.
- (35) Laity, J. H., Lee, B. M., and Wright, P. E. (2001) Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* 11, 39–46.
- (36) Slama, P., and Geman, D. (2011) Identification of family-determining residues in PHD fingers. *Nucleic Acids Res.* 39, 1666–1679.
- (37) Gibbons, R. J., Bachoo, S., Picketts, D. J., Aftimos, S., Asenbauer, B., Bergoffen, J., Berry, S. a, Dahl, N., Fryer, a, Keppler, K., Kurosawa, K., Levin, M. L., Masuno, M., Neri, G., Pierpont, M. E., Slaney, S. F., and Higgs, D. R. (1997) Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. *Nat. Genet.* 17, 146–148.
- (38) Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor RID D-4566-2009. *J. Mol. Biol.* 304, 723–729.
- (39) Li, Y., and Li, H. (2012) Many keys to push: diversifying the “readership” of plant homeodomain fingers. *Acta Biochim. Biophys. Sin. (Shanghai).* 44, 28–39.
- (40) Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 8, 983–994.
- (41) Mukhopadhyay, S., and Sengupta, A. M. (2013) The role of multiple marks in epigenetic silencing and the emergence of a stable bivalent chromatin state. *PLoS Comput. Biol.* 9, e1003121.
- (42) Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J., and Zhou, M.-M. (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466, 258–U138.
- (43) Ruthenburg, A. J., Li, H., Milne, T. a, Dewell, S., McGinty, R. K., Yuen, M., Ueberheide, B., Dou, Y., Muir, T. W., Patel, D. J., and Allis, C. D. (2011) Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* 145, 692–706.
- (44) West, A. C., and Johnstone, R. W. (2014) New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest.* 124, 30–39.
- (45) Saurin, A. J., Borden, K. L. B., Boddy, M. N., and Freemont, P. S. (1996) Does this have a familiar RING? *Trends Biochem. Sci.* 21, 208–214.
- (46) Agricola, E., Randall, R. a R., Gaarenstroom, T., Dupont, S., and Hill, C. S. (2011) Recruitment of TIF1 gamma to Chromatin via Its PHD Finger-Bromodomain Activates Its Ubiquitin Ligase and Transcriptional Repressor Activities. *Mol. Cell* 43, 85–96.
- (47) Tsai, W.-W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M.-C., Patel, D. J., and Barton, M. C. (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927–932.
- (48) Herquel, B., Ouararhni, K., Khetchoumian, K., Ignat, M., Teletin, M., and Mark, M. (2011) Transcription cofactors TRIM24, TRIM28, and TRIM33 associate to form regulatory complexes that suppress murine hepatocellular carcinoma. *PNAS* 108, 8212–8217.

- (49) Herquel, B., Ouarrarhni, K., and Davidson, I. (2011) The TIF1 α -related TRIM cofactors couple chromatin modifications to transcriptional regulation, signaling and tumor suppression. *Transcription* 2, 231–6.
- (50) Hatakeyama, S. (2011) TRIM proteins and cancer. *Nat. Rev. Cancer* 11, 792–804.
- (51) Venturini, L., You, J., Stadler, M., Galien, R., Lallemand, V., Koken, M. H., Mattei, M. G., Ganser, a, Chambon, P., Losson, R., and de Thé, H. (1999) TIF1gamma, a novel member of the transcriptional intermediary factor 1 family. *Oncogene* 18, 1209–1217.
- (52) Mujtaba, S., He, Y., Zeng, L., Farooq, A., Carlson, J. E., Ott, M., Verdin, E., and Zhou, M.-M. (2002) Structural Basis of Lysine-Acetylated HIV-1 Tat Recognition by PCAF Bromodomain. *Mol. Cell* 9, 575–586.
- (53) Zeng, L., Li, J., Muller, M., Yan, S., Mujtaba, S., Pan, C., Wang, Z., and Zhou, M.-M. (2005) Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J. Am. Chem. Soc.* 127, 2376–7.
- (54) Brand, M., Measures, A. R., Wilson, B. G., Cortopassi, W. A., Alexander, R., Höss, M., Hewings, D. S., Rooney, T. P. C., Paton, R. S., and Conway, S. J. (2014) Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chem. Biol.* 10, 22–39.
- (55) Rooney, T. P. C., Filippakopoulos, P., Fedorov, O., Picaud, S., Cortopassi, W. a, Hay, D. a, Martin, S., Tumber, A., Rogers, C. M., Philpott, M., Wang, M., Thompson, A. L., Heightman, T. D., Pryde, D. C., Cook, A., Paton, R. S., Müller, S., Knapp, S., Brennan, P. E., and Conway, S. J. (2014) A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation- π interaction. *Angew. Chem. Int. Ed. Engl.* 53, 6126–6130.
- (56) Hay, D. a, Fedorov, O., Martin, S., Singleton, D. C., Tallant, C., Wells, C., Picaud, S., Philpott, M., Monteiro, O. P., Rogers, C. M., Conway, S. J., Rooney, T. P. C., Tumber, A., Yapp, C., Filippakopoulos, P., Bunnage, M. E., Müller, S., Knapp, S., Schofield, C. J., and Brennan, P. E. (2014) Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *J. Am. Chem. Soc.* 136, 9308–19.
- (57) Palmer, W. S., Poncet-Montange, G., Liu, G., Petrocchi, A., Reyna, N., Subramanian, G., Theroff, J., Yau, A., Kost-Alimova, M., Bardenhagen, J. P., Leo, E., Shepard, H. E., Tieu, T. N., Shi, X., Zhan, Y., Zhao, S., Draetta, G., Toniatti, C., Jones, P., Geck Do, M., and Andersen, J. N. (2015) Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J. Med. Chem.* 10.1021/acs.jmedchem.5b00405.
- (58) Bennett, J. M., Fedorov, O., Tallant, C., Monteiro, O. P., Meier, J., Gamble, V., Savitski, P., Nunez-Alonso, G. A., Haendler, B., Rogers, C., Brennan, P. E., Müller, S., and Knapp, S. (2015) Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *J. Med. Chem.* 59, 1642–1647.
- (59) Hay, D., Brennan, P., Fedorov, O., Tallant, C., Monteiro, O., Muller-Knapp, S., Knapp, S., Schofield, C. J., Martin, S., and Rogers, C. (2015) Design and synthesis of potent and selective inhibitors of BRD9 and BRD7 bromodomains. *Med. Chem. Commun.* 6, 1381–1386.
- (60) Clark, P. G. K., Vieira, L. C. C., Tallant, C., Fedorov, O., Singleton, D. C., Rogers, C. M., Monteiro, O. P., Bennett, J. M., Baronio, R., Müller, S., Daniels, D. L., Méndez, J., Knapp, S., Brennan, P. E., and Dixon, D. J. (2015) LP99: Discovery and Synthesis of the First Selective BRD7/9 Bromodomain Inhibitor. *Angew. Chemie Int. Ed.* 54, 6217–6221.

- (61) Vidler, L. R., Brown, N., Knapp, S., and Hoelder, S. (2012) Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* 55, 7346–59.
- (62) Halgren, T. a. (2009) Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* 49, 377–389.
- (63) Zhang, G., Sanchez, R., and Zhou, M.-M. (2012) Scaling the druggability landscape of human bromodomains, a new class of drug targets. *J. Med. Chem.* 55, 7342–7345.
- (64) Steiner, S., Magno, A., Huang, D., and Caflisch, A. (2013) Does bromodomain flexibility influence histone recognition? *FEBS Lett.* 587, 2158–2163.
- (65) Musselman, C., and Kutateladze, T. (2009) PHD fingers: epigenetic effectors and potential drug targets. *Mol. Interv.* 9, 314–323.
- (66) Brennan, P., Filippakopoulos, P., and Knapp, S. (2012) The therapeutic potential of acetyl-lysine and methyl-lysine effector domains. *Drug Discov. Today Ther. Strateg.* 9, e101–e110.
- (67) Santiago, C., Nguyen, K., and Schapira, M. (2011) Druggability of methyl-lysine binding sites. *J. Comput. Aided. Mol. Des.* 25, 1171–1178.
- (68) Van Zutven, L. J. C. M., Önen, E., Velthuisen, S. C. J. M., Van Drunen, E., Von Bergh, A. R. H., Van Den Heuvel-Eibrink, M. M., Veronese, A., Mecucci, C., Negrini, M., De Greef, G. E., and Beverloo, H. B. (2006) Identification of NUP98 abnormalities in acute leukemia: JARIDIA (12p13) as a new partner gene. *Genes Chromosom. Cancer* 45, 437–446.
- (69) Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B., and Denu, J. M. (2012) Identification and Characterization of Small Molecule Inhibitors of a Plant Homeodomain Finger. *Biochemistry* 51, 8293–8306.
- (70) Herold, J. M., Wigle, T. J., Norris, J. L., Lam, R., Korboukh, V. K., Gao, C., Ingberman, L. A., Kireev, D. B., Senisterra, G., Vedadi, M., Tripathy, A., Brown, P. J., Arrowsmith, C. H., Jin, J., Janzen, W. P., and Frye, S. V. (2011) Small-Molecule Ligands of Methyl-Lysine Binding Proteins. *J. Med. Chem.* 54, 2504–2511.
- (71) James, L. I., Korboukh, V. K., Krichevsky, L., Baughman, B. M., Herold, J. M., Norris, J. L., Jin, J., Kireev, D. B., Janzen, W. P., Arrowsmith, C. H., and Frye, S. V. (2013) Small-molecule ligands of methyl-lysine binding proteins: Optimization of selectivity for L3MBTL3. *J. Med. Chem.* 56, 7358–7371.
- (72) Hassan, A. H., Awad, S., Al-Natour, Z., Othman, S., Mustafa, F., and Rizvi, T. A. (2007) Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem. J.* 402, 125–33.
- (73) Philpott, M., Yang, J., Tumber, T., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Keates, T., Felletar, I., Ciulli, A., Knapp, S., and Heightman, T. D. (2011) Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* 7, 2899–2908.
- (74) Hu, P., Wang, X., Zhang, B., Zhang, S., Wang, Q., and Wang, Z. (2014) Fluorescence polarization for the evaluation of small-molecule inhibitors of PCAF BRD/Tat-AcK50 association. *ChemMedChem* 9, 928–931.

- (75) Bamborough, P., Diallo, H., Goodacre, J. D., Gordon, L., Lewis, A., Seal, J. T., Wilson, D. M., Woodrow, M. D., and Chung, C.-W. (2012) Fragment-based discovery of bromodomain inhibitors part 2: optimization of phenylisoxazole sulfonamides. *J. Med. Chem.* 55, 587–96.
- (76) Sachchidanand, Resnick-Silverman, L., Yan, S., Mutjaba, S., Liu, W. J., Zeng, L., Manfredi, J. J., and Zhou, M. M. (2006) Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chem. Biol.* 13, 81–90.
- (77) Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Felletar, I., Philpott, M., Munro, S., McKeown, M. R., Wang, Y., Christie, A. L., West, N., Cameron, M. J., Schwartz, B., Heightman, T. D., La Thangue, N., French, C. a, Wiest, O., Kung, A. L., Knapp, S., and Bradner, J. E. (2010) Selective inhibition of BET bromodomains. *Nature* 468, 1067–73.
- (78) Chung, C.-W., Coste, H., White, J. H., Mirguet, O., Wilde, J., Gosmini, R. L., Delves, C., Magny, S. M., Woodward, R., Hughes, S. a, Boursier, E. V., Flynn, H., Bouillot, A. M., Bamborough, P., Brusq, J.-M. G., Gellibert, F. J., Jones, E. J., Riou, A. M., Homes, P., Martin, S. L., Uings, I. J., Toum, J., Clement, C. a, Boullay, A.-B., Grimley, R. L., Blandel, F. M., Prinjha, R. K., Lee, K., Kirilovsky, J., and Nicodeme, E. (2011) Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J. Med. Chem.* 54, 3827–38.
- (79) Dawson, M. A., Prinjha, R. K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.-I., Robson, S. C., Chung, C., Hopf, C., Savitski, M. M., Huthmacher, C., Gudgin, E., Lugo, D., Beinke, S., Chapman, T. D., Roberts, E. J., Soden, P. E., Auger, K. R., Mirguet, O., Doehner, K., Delwel, R., Burnett, A. K., Jeffrey, P., Drewes, G., Lee, K., Huntly, B. J. P., and Kouzarides, T. (2011) Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* 478, 529–33.
- (80) Ferguson, F. M., Fedorov, O., Chaikuad, A., Philpott, M., Muniz, J. R. C., Felletar, I., von Delft, F., Heightman, T., Knapp, S., Abell, C., and Ciulli, A. (2013) Targeting low-druggability bromodomains: fragment based screening and inhibitor design against the BAZ2B bromodomain. *J. Med. Chem.* 56, 10183–7.
- (81) Chung, C., Dean, A. W., Woolven, J. M., and Bamborough, P. (2012) Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* 55, 576–586.
- (82) Mirguet, O., Lamotte, Y., Chung, C.-W., Bamborough, P., Delannée, D., Bouillot, A., Gellibert, F., Krysa, G., Lewis, A., Witherington, J., Huet, P., Dudit, Y., Trottet, L., and Nicodeme, E. (2014) Naphthyridines as Novel BET Family Bromodomain Inhibitors. *ChemMedChem* 9, 580–9.
- (83) Falconer, R. J., Penkova, A., Jelesarov, I., and Collins, B. M. (2008) Survey of the year 2008: applications of isothermal titration calorimetry. *J. Mol. Recognit.* 23, 395–413.
- (84) Lucas, X., Wohlwend, D., Hügle, M., Schmidtkunz, K., Gerhardt, S., Schüle, R., Jung, M., Einsle, O., and Günther, S. (2013) 4-Acyl pyrroles: mimicking acetylated lysines in histone code reading. *Angew. Chem. Int. Ed. Engl.* 52, 14055–9.
- (85) Kanno, T., Kanno, Y., Siegel, R. M., Jang, M. K., Lenardo, M. J., and Ozato, K. (2004) Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol. Cell* 13, 33–43.

- (86) Philpott, M., Rogers, C. M., Yapp, C., Wells, C., Lambert, J.-P., Strain-Damerell, C., Burgess-Brown, N. a, Gingras, A.-C., Knapp, S., and Müller, S. (2014) Assessing cellular efficacy of bromodomain inhibitors using fluorescence recovery after photobleaching. *Epigenetics Chromatin* 7, 1–12.
- (87) Fedorov, O., Lingard, H., Wells, C., Monteiro, O. P., Picaud, S., Keates, T., Yapp, C., Philpott, M., Martin, S. J., Felletar, I., Marsden, B. D., Filippakopoulos, P., Mu, S., Knapp, S., Brennan, P. E., Drive, R., and Ox, O. (2014) [1,2,4]Triazolo[4,3-a]phthalazines: Inhibitors of Diverse Bromodomains. *J. Med. Chem.* 57, 462–476.
- (88) Chung, C.-W., and Witherington, J. (2011) Progress in the discovery of small-molecule inhibitors of bromodomain--histone interactions. *J. Biomol. Screen.* 16, 1170–85.
- (89) Hopkinson, R. J., Walport, L. J., Münzel, M., Rose, N. R., Smart, T. J., Kawamura, A., Claridge, T. D. W., and Schofield, C. J. (2013) Is JmjC oxygenase catalysis limited to demethylation? *Angew. Chem. Int. Ed. Engl.* 52, 7709–7713.
- (90) Vidler, L. R., Filippakopoulos, P., Fedorov, O., Picaud, S., Martin, S., Tomsett, M., Woodward, H., Brown, N., Knapp, S., and Hoelder, S. (2013) Discovery of novel small-molecule inhibitors of BRD4 using structure-based virtual screening. *J. Med. Chem.* 56, 8073–88.
- (91) Meslamani, J., Smith, S. G., Sanchez, R., and Zhou, M.-M. (2014) ChEpiMod: a knowledgebase for chemical modulators of epigenome reader domains. *Bioinformatics* 30, 1481–1483.
- (92) Loharch, S., Bhutani, I., Jain, K., Gupta, P., Sahoo, D. K., and Parkesh, R. (2015) EpiDBase: a manually curated database for small molecule modulators of epigenetic landscape. *Database* 10.1093/database/bav013.
- (93) Arrowsmith, C. H., Audia, J. E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P. E., Brown, P. J., Bunnage, M. E., Buser-doeppner, C., Campbell, R. M., Carter, A. J., Cohen, P., Copeland, R. A., Cravatt, B., and Dahlin, J. L. (2015) The promise and peril of chemical probes. *Nat. Chem. Biol.* 11, 536–541.
- (94) Zhao, H., Gartenmann, L., Dong, J., Spiliotopoulos, D., and Caflisch, A. (2014) Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett.* 24, 2493–6.
- (95) Muvva, C., Singam, E. R. A., Raman, S. S., and Subramanian, V. (2014) Structure-based virtual screening of novel, high-affinity BRD4 inhibitors. *Mol. Biosyst.* 10, 2384–2397.
- (96) Gallenkamp, D., Gelato, K. a, Haendler, B., and Weinmann, H. (2014) Bromodomains and their pharmacological inhibitors. *ChemMedChem* 9, 438–64.
- (97) Lamoureux, F., Baud'huin, M., Rodriguez Calleja, L., Jacques, C., Berreur, M., Rédini, F., Lecanda, F., Bradner, J. E., Heymann, D., and Ory, B. (2014) Selective inhibition of BET bromodomain epigenetic signalling interferes with the bone-associated tumour vicious cycle. *Nat. Commun.* 5, 3511.
- (98) Cheng, Z., Gong, Y., Ma, Y., Lu, K., Lu, X., Pierce, L. A., Thompson, R. C., Muller, S., Knapp, S., and Wang, J. (2013) Inhibition of BET Bromodomain Targets Genetically Diverse Glioblastoma. *Clin. Cancer Res.* 19, 1748–1759.

- (99) Barrett, E., Brothers, S., Wahlestedt, C., and Beurel, E. (2014) I-BET151 selectively regulates IL-6 production. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1842, 1549–1555.
- (100) Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. a, Hoke, H. a, and Young, R. a. (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–47.
- (101) Lu, J.-J., Pan, W., Hu, Y.-J., and Wang, Y.-T. (2012) Multi-Target Drugs: The Trend of Drug Research and Development. *PLoS One* (Aggarwal, B. B., Ed.) 7, e40262.
- (102) Hewings, D. S., Rooney, T. P. C., Jennings, L. E., Hay, D. a, Schofield, C. J., Brennan, P. E., Knapp, S., and Conway, S. J. (2012) Progress in the Development and Application of Small Molecule Inhibitors of Bromodomain Acetyl-lysine Interactions. *J. Med. Chem.* 55, 9393–413.
- (103) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- (104) Choi, J. H., Jung, H. Y., Kim, H. S., and Cho, H. G. (2000) PhyloDraw: a phylogenetic tree drawing system. *Bioinformatics* 16, 1056–1058.

Chapter 2: Biological Characterisation

2.1 Introduction

The aim of this work was to characterise the natural binding mode and recognition pattern for the two homologous target proteins, TRIM24 and TRIM33, in order to gain an understanding of these more complex, dual-domain systems prior to small molecule screening. This investigation could identify any key points of difference between the two proteins or the two domains, which could enable the development of selectivity for small molecule probes. At the time of commencing this study there were no known small molecule approaches to inhibition of the tandem domains. Indeed, to date the only reported inhibitors of this family are of the TRIM24 bromodomain,^{1,2} with use of the PHD finger in the assay construct purely for solubility reasons.

Previous studies into cellular recognition modes of the TRIM family have focussed predominantly on their interactions with other signalling proteins: regulation of gene expression as transcription cofactors,³ role as an E3-Ubiquitin Ligase,^{4,5} and role in the cytokine TGFbeta pathway.⁶ These activities have also been linked to association of TRIM family proteins with chromatin.⁷ It is hypothesised that the involvement of two reader domains can elicit a more complex response to epigenetic signals, through responses to recognition by each domain in isolation or combination.

2.1.1 TRIM24 and TRIM33

As discussed in the Introduction, TRIM24 and TRIM33 display distinct recognition profiles towards histone modifications, including recognition of both methylation and acetylation state at various positions of a Histone 3 (H3) peptide (Figure 2.1). Availability of *apo*-crystal structures for TRIM24 and TRIM33 in the Protein Data Bank (PDB: 3O33 and 3U5O, respectively), make these proteins tractable targets for small-molecule probe development. Towards this aim, the former structure may be more useful, since data is of a higher resolution, including the positions of solvating water molecules. With a sequence similarity of 87% between the tandem PHD-bromodomain of these homologous proteins, studies will be conducted against both systems in

parallel. Work towards the novel TRIM33 target will be facilitated by extension of procedures from the more validated TRIM24, for which screening has been conducted at other institutions.¹

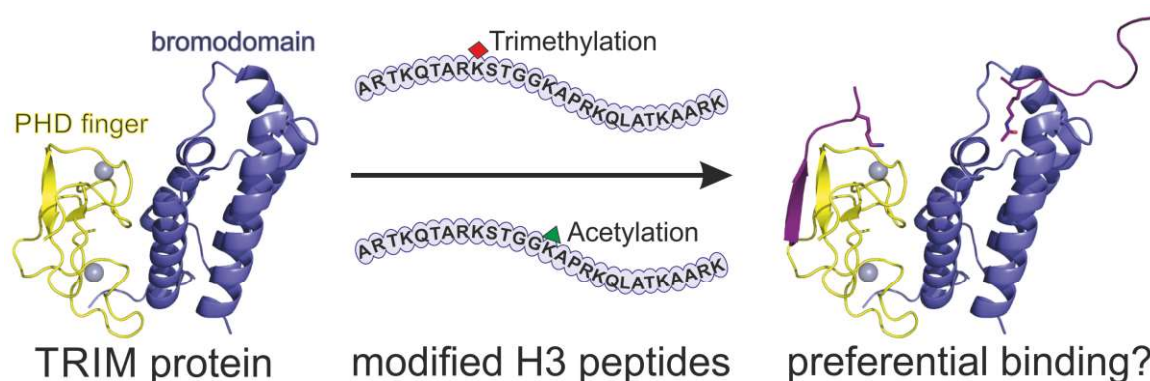


Figure 2.1 Studying natural recognition mode of TRIM proteins. The tandem PHD finger (yellow) and bromodomain (blue) cassette of TRIM proteins recognise H3 peptides bearing methylated or acetylated lysines, respectively. Characterising differential recognition of position of modification for TRIM24 and TRIM33 will further understanding of dual epigenetic reader domains, and may highlight potential for selective inhibitor development.

2.1.2 Aims

It has been shown that multiple modifications on a histone peptide can affect affinity towards TRIM proteins. Although previous studies have quantified selected peptide affinities, a systematic approach to identifying the positions of detection, and profiling the additive nature of modifications with the inclusion of K9Me₃, has not been carried out. Characterisation of the affinity of TRIM proteins to histone-mimicking peptides, by developing quantitative biophysical assays, would further our understanding of the complex nature of combinatorial epigenetic modifications.

The aims of this study were to:

- Express and purify the TRIM24 and TRIM33 PHD-bromodomain cassette, with an intact affinity label suitable for indirect assay development.
- Optimise an assay to quantify peptide affinity towards TRIM24 and extend this for use with TRIM33, for which no known screening capacity exists.
- Characterise the affinity of these tandem domains towards various H3 peptide modifications to profile their natural binding modes.

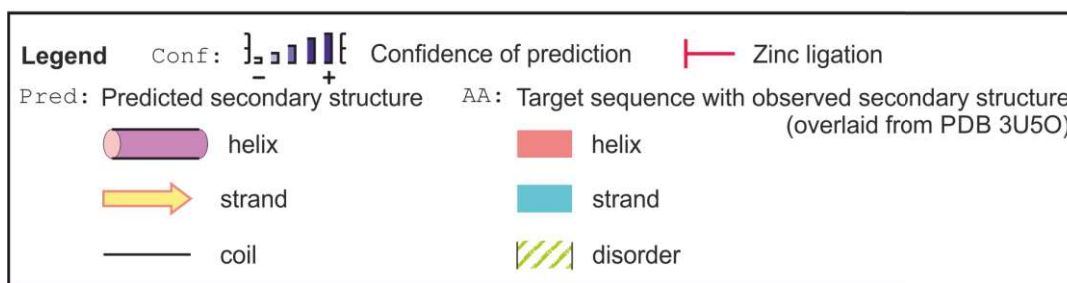
2.2 Domain organisational predictions

In vitro analyses of peptide affinities has been carried out more extensively for TRIM24 than TRIM33.⁸ Prior to assay development, a computational analysis of the predicted domain architecture of TRIM33 was carried out, with a comparison of significant non-homologous regions, in order to assess whether similar methodology would translate between the two systems. It was hypothesised that comparison might also provide possible rationalisations for differing selectivity in recognition sites between TRIM24 and TRIM33.

PSIPRED⁹ and GlobProt¹⁰ predictions were collated for TRIM33 and linked to putative domains (Figure 2.2). This analysis confirmed the presence of few elements of secondary structure within the PHD finger region, which poses a significant challenge for the design of expression constructs, since these defining features often serve as landmarks for sequence initiation and termination. As an alternative structural element, previous biological studies on TRIM24 and TRIM33 have used the zinc co-ordinating residues to define the extent of the PHD finger, with the expression sequences for both proteins starting almost immediately before the first cysteine of the Cys₄HisCys₃ motif.

The canonical form of TRIM33 (GenBank Accession: NP_056990) contains a 17 amino acid insertion, relative to TRIM24 (GenBank Accession: NP_148980), which extends the α B helix. The effect of this is two-fold: i) extension of the BC loop region in TRIM33 may directly affect proximal sequence recognition and positional preferences; ii) a crucial feature of KAc recognition – the asparagine residue involved in the primary hydrogen bond with the carbonyl of an acetyl group – is forced into a different orientation towards the binding pocket, which may reduce the ability to recognise this modification, and hence sensitivity in peptide-binding assays. Additionally, difficulties in purification of TRIM33 have been attributed to cleavage in this insertion region.

The observed similarities between TRIM24 and TRIM33 support concurrent assay development. However, low levels of secondary structure suggest that correct folding of the PHD finger topology may be challenging. In order to validate expression of PHD fingers within the tandem constructs, comparison of peptide recognition ability could be used to infer an active folded conformation.



38

2.3 Expression and purification

In order to initiate biological studies of this system, it was necessary to have access to homogenous, structurally intact proteins, which retained the ability to recognise histone modifications from the isolated PHD finger and bromodomain.

2.3.1 Plasmids

Plasmids containing DNA for the TRIM33 gene were received from Professor Joan Massagué's lab in pGEX-4T1 vectors (Addgene Plasmid 15734). After rehydration from filter paper,¹¹ direct transformation into *E. coli* expression strains was unsuccessful, potentially due to low plasmid concentrations. Successive transformation into highly competent NEB 10-beta (New England Biolabs) and XL10 (Gold) cells allowed uptake of gene-containing plasmids into bacterial chromosomal DNA. Amplification through cell growth and plasmid extraction, followed by subsequent transformation into BL21 DE3 (Gold) cells provided a system for expression trials.

TRIM24 plasmids were received from the Structural Genomics Consortium (SGC) in pNIC28-Bsa4 expression vectors.¹² Sequencing of this DNA revealed a missense mutation replacing an aspartic acid with a tyrosine (D872Y) in the PHD finger domain (Figure 2.3). This residue may be required for domain stability, by forming a hydrogen bond between a backbone carbonyl on the adjacent bromodomain. Site directed mutagenesis

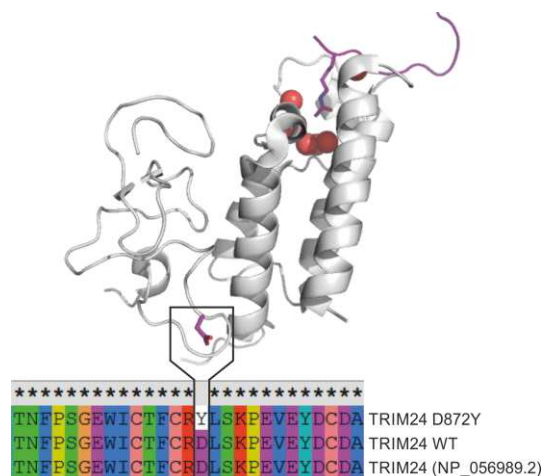


Figure 2.3 TRIM24 D872Y mutation. Position highlighted (black box) on structure (PDB: 3O34); sequence alignments shown after mutation to revert to wild type (WT) compared with NCBI sequence (NP_056989.2).

at this position was achieved to revert to the wild type sequence *via* PCR with primers containing one mismatched DNA base (see Materials and Methods for primers).

Ligation independent cloning (LIC)¹³ was used to transfer TRIM33 into the same pNIC28-Bsa4 vector (His₆ at *N*-terminus).¹² Protein expression would then be under the control of an inducible

T7 promoter based system, whereby expression could be induced on addition of isopropyl β -D-1-thiogalactopyranoside (IPTG), which removes transcriptional repression exerted by the LacI repressor (constitutively expressed from the pNIC28-Bsa4 plasmid). The pNIC28-Bsa4 vector was linearised using BseRI, and cohesive ends were generated using the chew-back action of T4 DNA polymerase. Annealing of TRIM33 gene inserts, amplified using PCR with inclusion of upper and lower LIC sequences, was followed by direct transformation into XL10 competent cells. In addition to facilitating protein purification, His₆-labels would allow the use of nickel chelation for development of indirect assays, for example SPR or AlphaScreen, which require immobilisation or coupling to a secondary element.

2.3.2 Expression trials

Trials were conducted at varying temperatures (18 or 28 °C), and concentrations of IPTG induction (0.1 to 0.5 mM), to achieve good overexpression of the protein of interest. A high ratio of induced protein expression, relative to background, would enable facile purification with a good yield. Cells were grown at 37 °C to an OD₆₀₀ of 0.8-1.0 before induction; for the structural Zn²⁺ ions in the PHD finger fold to be ubiquitously incorporated, ZnCl₂ was also added at this point (final concentration 0.1 mM). Whole cell samples (after one cycle of freeze-thaw lysis) were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to visualise expression levels. Optimal conditions for induction consisted of 0.25 mM IPTG at 18 °C overnight and were used thereafter for expression of all plasmid vectors generated.

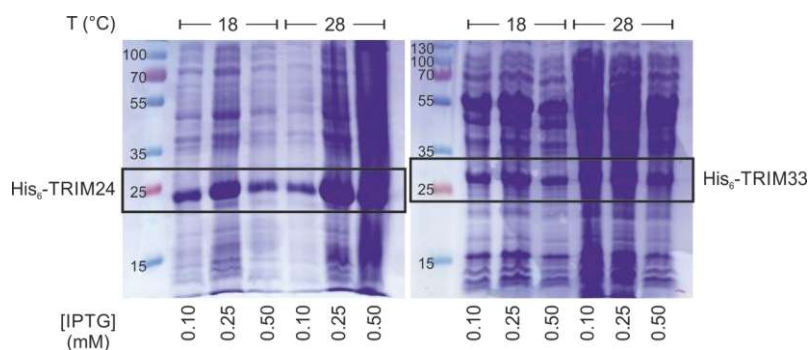


Figure 2.4 SDS-PAGE gels from expression trials for TRIM proteins in pNIC28-Bsa4 vector. Bands from overexpression highlighted: TRIM24 expected mass 23.92 kDa; TRIM33 expected mass 30.50 kDa. 100 mL cultures were incubated at 37 °C to an OD₆₀₀ of 1.0, before equilibration to 18 °C or 28 °C, and induction with 0.1 mM ZnCl₂ and IPTG at the concentrations indicated below.

2.3.3 Purification

Purification by immobilised metal affinity chromatography (IMAC; HisTrap HP, GE Healthcare), and size exclusion chromatography (Sephadex gel S75, GE Healthcare) yielded high quantities of 20-30 mg protein per litre of growth culture at >95 % purity (Figure 2.6). Analysis by SDS-PAGE after the first IMAC purification step suggested the protein of interest was the dominant species. However, both TRIM24 and TRIM33 eluted as two distinct species in size exclusion chromatography. Bands corresponding to higher molecular weight species, eluting soon after the void volume of the gel filtration column, were thought to be oligomeric species which were observed at the interface of the stacking and resolving sections of SDS-PAGE gels, and there was some indication of dimeric protein in the intermediary fraction (MW 48 kDa and 61 kDa, for TRIM24 and TRIM33, respectively).

Preliminary assay trials resulted in an unexpected inverse sensitivity of TRIM33 to DMSO: rather than its usual effect as an inhibitor of peptide recognition, titration increased the signal observed from interaction in assays. Comparison with dimeric FIH *via* analytical S200 column (Figure 2.5:B), native gel analysis of fractions from gel filtration (Figure 2.5:C), and characterisation by size-exclusion chromatography multiple angle light scattering (SEC-MALS) highlighted large oligomeric forms obtained during purification. It was thought that disruption of interaction between these monomers by DMSO liberates more protein resulting in the counterintuitive increase in signal. Further purifications isolated only the final eluting fractions of from gel filtration columns to minimise oligomer content in samples.

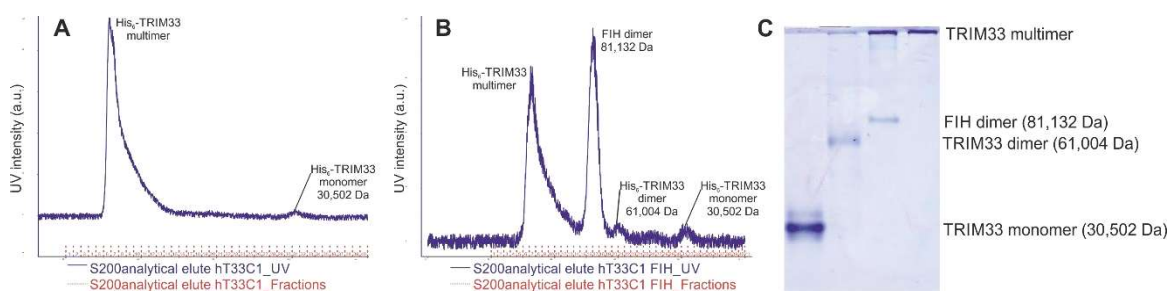


Figure 2.5: Analysis of multimeric forms of TRIM33. A: UV trace from elution of first collection fractions of TRIM33 from Sephadex 200 analytical column. B: UV trace from elution of first collection fractions of TRIM33 column doped with FIH, for size comparison, from Sephadex 200 analytical. C: Native gel analysis of isolated fractions from gel filtration of His₆-TRIM33 show monomeric form, and species eluting later containing dimer and multimers, with FIH for comparison.

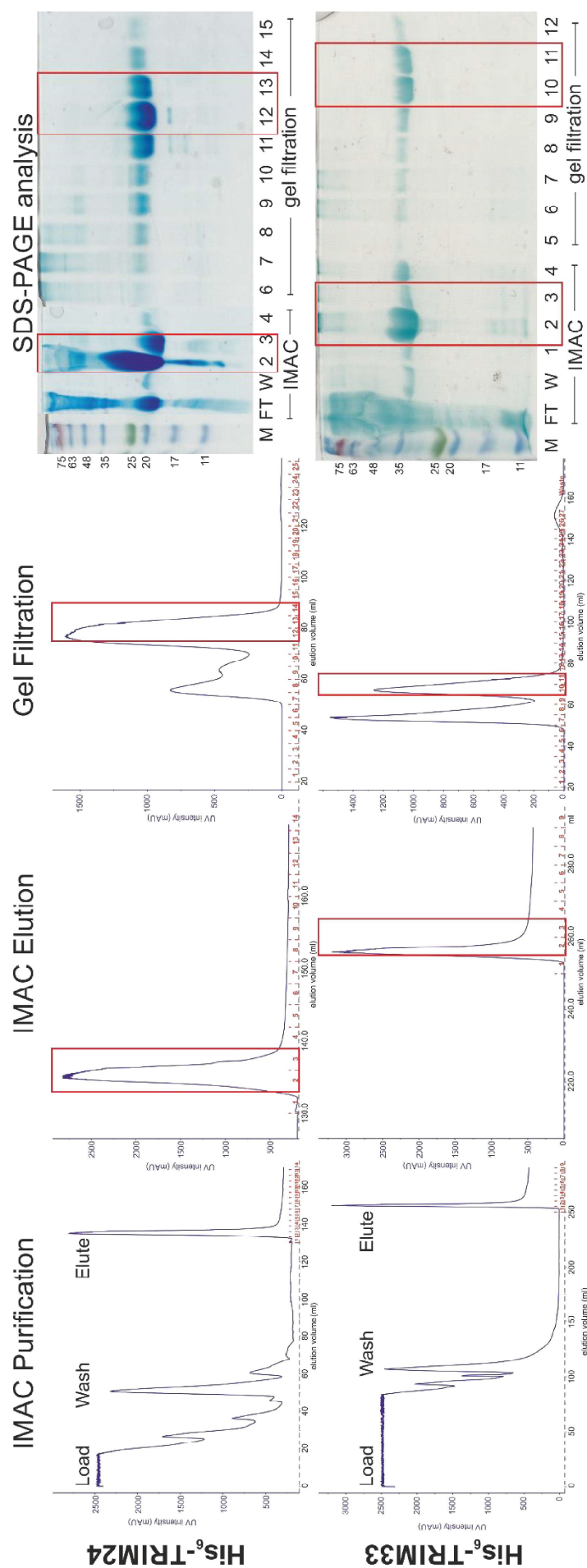


Figure 2.6 Purification of His₆-TRIM24 and His₆-TRIM33. Expression from pNIC-Bsa4 vectors. Purification by immobilised metal affinity chromatography (IMAC; HisTrap HP, GE Healthcare), followed by size exclusion chromatography (gel filtration; HiLoad Superdex 75 prep grade, GE Healthcare). Load, wash and elute steps are indicated, with the fraction collection from IMAC elution enlarged. Fractions containing the highest purity of protein, as assessed by SDS-PAGE (right), were collected and concentrated. Red box – isolated fractions; M – molecular weight ladder; FT – flow through; W – wash.

2.3.4 Anion-exchange purification

The necessity for separation of oligomeric forms led to an investigation of alternative purification methods. Any presence of incompetent oligomers in isolated protein could result in low recognition ability, and require high protein concentrations to generate signal in assays dependent on protein-peptide association. Anion-exchange purification separates molecules by overall surface charge, rather than size, and could result in better oligomeric separation, purity and hence recognition activity. UV traces from size exclusion chromatography suggested a higher oligomeric content for TRIM33 than TRIM24, and the former protein was selected for trials. Initial data demonstrated that TRIM33 bound well to a quaternary amine (QA) strong anion exchange (SAX) column, and eluted at relatively low salt concentrations, however, multiple oligomeric forms were observed across all fractions by non-denaturing PAGE. It was hypothesised that prolonged exposure to low salt conditions after separation increases oligomerisation – 500 mM NaCl was used in size exclusion purifications to stabilise the monomeric form.

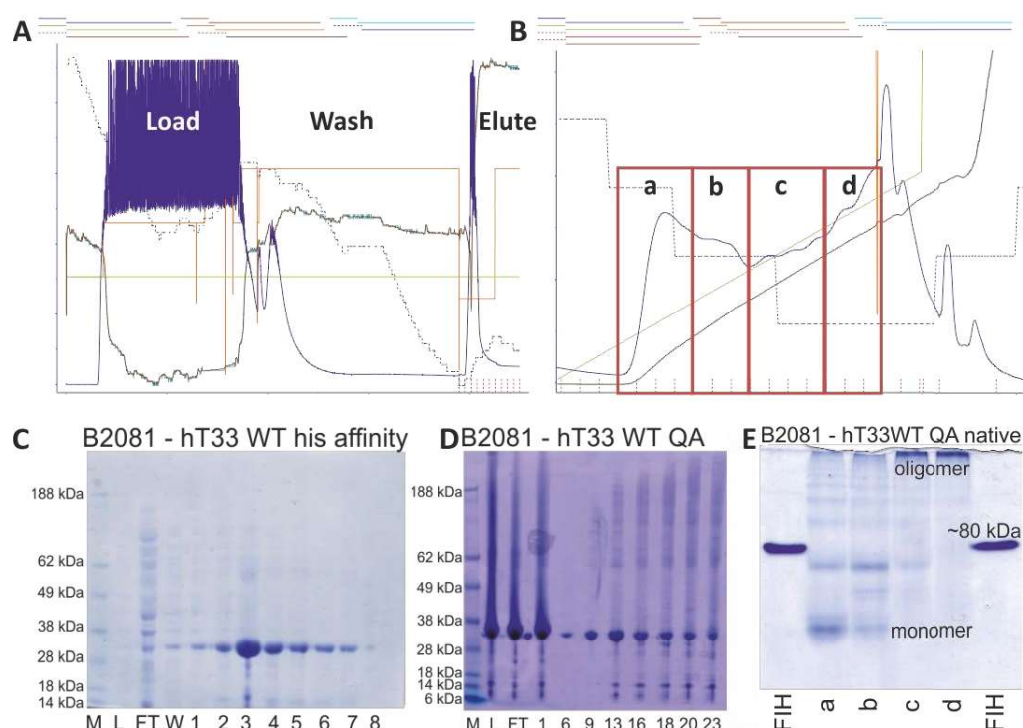


Figure 2.7 Summary of purification by strong anion exchange. **A** UV trace from his-affinity column. **B** UV trace of elution from QA-column. Buffer composition (Buffer A: 20 mM HEPES, 50 mM NaCl, 1 mM β -ME; Buffer B: 20 mM HEPES, 1 M NaCl, 1 μ M β -ME) shown as green trace. Combined fractions highlighted in red boxes. **C** SDS-PAGE gel of fractions from his-affinity column. **D** SDS-PAGE gel of fractions from QA-column show primary eluting species corresponds to MW of TRIM33 throughout. **E** Regular spaced bands on native PAGE gel of combined fractions from QA-elution show multiple oligomeric forms. Major species transitions from monomer to high order oligomers across elution (a-d).

Subsequently, a two-step gradient elution in HEPES buffer was performed, with a long elution of the monomer at low salt concentrations, followed by elution of higher order oligomers (Figure 2.8). Optimisation of the elution gradient enabled greater peak separation, as analysed by UV, with SDS-PAGE analysis suggesting the first eluting species was primarily monomer. Non-denaturing analysis was unsuccessful, however, oligomer could be observed on standard gels at the boundary between stacking and resolving gel sections (e.g. QA gradient C fractions 17-18).

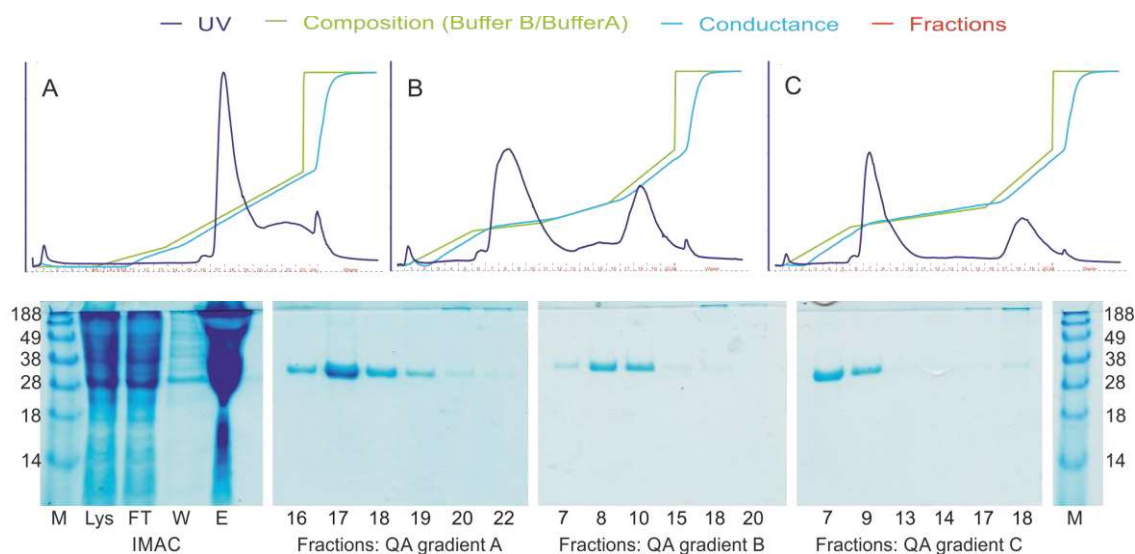


Figure 2.8 Top: FPLC UV traces (blue) from QA purification of His₆-TRIM33. Monomeric species elutes separately from oligomeric species. Conductivity (cyan) tracks well with buffer composition (green) from Buffer A (20 mM HEPES, 50 mM NaCl, 1 mM β -ME) to Buffer B (20 mM HEPES, 1 M NaCl, 1 mM β -ME). **Gradient A:** 0% B, 5 mL; 0-10% B, 10 mL; 10-50% B, 20 mL; 100% B, 10 mL. **Gradient B:** 0-18% B, 10 mL; 18-22% B, 10 mL; 22-32% B, 10 mL; 32-60% B, 10 mL; 100% B, 10 mL. **Gradient C:** 0-20% B, 10 mL; 20-30% B, 20 mL; 30-60% B, 10 mL; 100% B, 10 mL. **Bottom:** SDS-PAGE gels from IMAC purification, and selected fractions from QA purification trials. Oligomeric forms were observed at the boundary between stacking and resolving gel sections (≥ 100 kDa).

Although the trials conducted demonstrated that anion-exchange methods could be used as a ‘polishing step’ in purification after IMAC columns, there was no significant improvement in purity compared to size exclusion chromatography. Multiple injections were required to avoid breakthrough of sample during the load step on a 1 mL column, and the longer duration of purification carried the risk of protein degradation, so this method was not pursued further.

2.4 Assay development

The SGC provided a probe compound for the TRIM24 bromodomain, which was in development at the time this work was undertaken (SGC-TIF1 α 53: structure undisclosed). This was utilised as a positive control for assay development, and was compared with signals generated from exposure to modified peptides bearing acetylated or methylated lysine residues.

2.4.1 Tryptophan fluorescence

Proteins contain three aromatic amino acid residues – tryptophan, tyrosine, and phenylalanine – which contribute to their intrinsic fluorescence. Of these, tryptophan has by far the strongest fluorescence and highest quantum yield. Monitoring tryptophan fluorescence is a valuable tool which can be used to make inferences regarding local structure and dynamics.¹⁴ This technique has previously been used to characterise binding of small molecules to bromodomains,¹⁵ and is very sensitive to changes in electrostatic interactions. TRIM-KMe₃ recognition occurs through a cation- π interaction with a solvent exposed tryptophan residue, one of only two tryptophans in the protein. These residues present a potential novel method for differentiation of binding across the tandem domain, to enable focussed targeting of the PHD finger and interdomain regions.

To test the hypothesis that TRIM-KMe₃ recognition would affect intrinsic tryptophan fluorescence, and enable an assay based on this technique to detect inhibition of this interaction, fluorescence spectra for TRIM24 were obtained under various conditions. Signal specifically generated from aromatic residues within TRIM24 should increase in a protein-concentration dependent manner, and was indeed observed for broad signals resulting from excitation at 280 or 295 nm (Figure 2.9).

Titration of a trimethylated peptide binding partner (H3K9Me₃) into a protein sample resulted in slight dose-dependent increases in emission (~60% increase in emission after addition of 40 μ M H3K9Me₃ to 2 μ M TRIM24), and a ~10 nm shift in peak emission wavelength (maximum emission from excitation of 2 μ M TRIM24 at 295 nm: in the absence of peptide λ_{max} = 341.5, 287 a.u.; in the presence of 40 μ M H3K9Me₃ λ_{max} = 351.5, 455 a.u.). Similar variations of 10-20 nm were seen for

the shift in tryptophan fluorescence when its environment changes from one on a protein surface, to being shielded in a hydrophobic core.¹⁴ Although this technique was promising for the detection of TRIM-KMe₃ recognition, sensitivity of tryptophan fluorescence was low, with poor resolution for increases in fluorescence at low concentrations of peptide (only ~15% increase in fluorescence for 1:1 mixture of TRIM24 and H3K9Me₃). Additionally, relatively high sample volumes were required (600 μ L) so this low-throughput method was deemed unsuitable for peptide affinity characterisation or small-molecule screening. Utility for quantification of inhibition values for ligands that interfere with KMe₃ recognition could be established once they are identified.

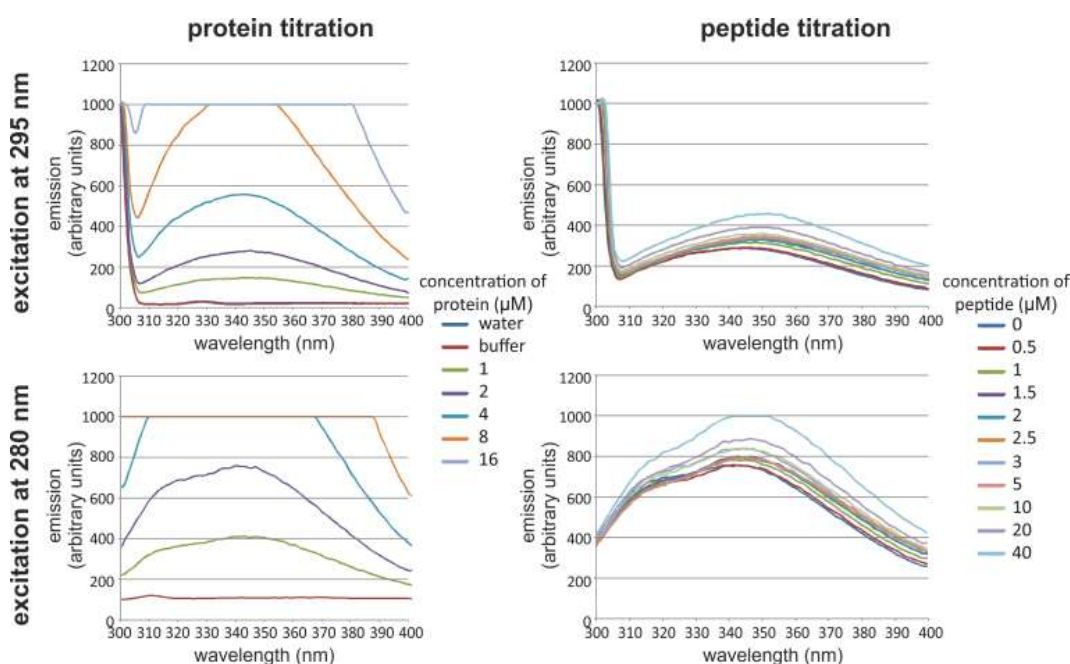


Figure 2.9 Tryptophan fluorescence of TRIM24 shows concentration dependence in isolation and upon titration with peptide binding partner (H3K9Me₃). Variations in emission measured after excitation at both 280 nm (tryptophan and tyrosine excitation) and 295 nm (tryptophan excitation), with a slight Stokes shift to longer wavelengths, demonstrate its potential utility in monitoring tryptophan environment. Signal from excitation at 280 nm reached saturation after 8 μ M.

2.4.2 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a rapid and inexpensive screening method to identify low-molecular-weight ligands which stabilise purified proteins with respect to thermal denaturation.¹⁶ Validation of this technique for TRIM24 and TRIM33, with either the provided control compound or acetylated peptides, might provide a useful platform for novel binding partner identification.

Data were obtained for melting temperatures (T_m) of TRIM24 and TRIM33 in the presence of varying concentrations of the positive control provided by the SGC (SGC-TIF1 α 53). Melting curves were well formed (Figure 2.10), with very low standard deviations between T_m of replicates (average SD 0.2 °C), and protein samples tolerated 1%

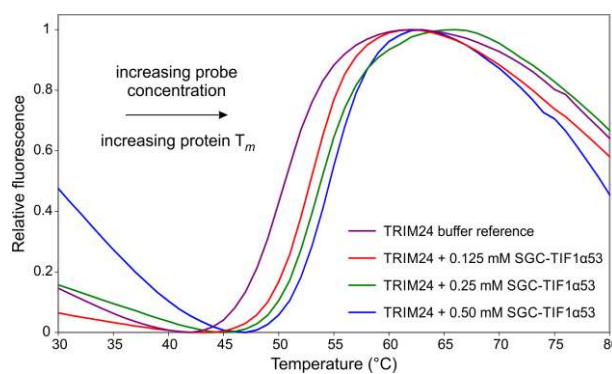


Figure 2.10 Representative melting curves of TRIM24 with varying concentrations of a bromodomain probe (SGC-TIF1 α 53). Curves are well formed and show probe-concentration dependent increase in protein stabilisation (T_m).

final assay concentration of DMSO without any artifactual signal. T_m shift values showed good concentration dependence for the control compound against TRIM24, but no significant shifts were observed with TRIM33, for either the control or the H3K14Ac peptide (Table 2.1). It could not be determined whether the absence of stabilisation of TRIM33 was due to a lack of sensitivity in this assay, or the selectivity of recognition, because no positive controls for TRIM33 were available at the time. These results suggested that DSF could be used to identify ligands for TRIM24, however, relatively low shifts suggest ligands would have to have high affinity to be observed, and sensitivity for ligands binding to TRIM33 was unknown.

Table 2.1 T_m shift data from control compound against TRIM24 and TRIM33 measured from 20-80 °C in 1 °C intervals with a hold of 1.5 min. Concentration dependent shifts in T_m were observed for TRIM24, but no significant shifts were observed with TRIM33, either due to low sensitivity, or selectivity of recognition.

Sample	T_m values (°C)			Average T_m (°C)	Average T_m shift (°C)	Standard deviation of T_m
His ₆ -TRIM24 buffer reference	50.4	51.9	50.9	51.1	-	0.7
His ₆ -TRIM24 + 100 μ M SGC-TIF1 α 53	54.6	54.5	54.3	54.5	3.4	0.2
His ₆ -TRIM24 + 50 μ M SGC-TIF1 α 53	53.8	53.5	53.6	53.6	2.6	0.2
His ₆ -TRIM24 + 25 μ M SGC-TIF1 α 53	52.9	52.9	53.1	52.9	1.9	0.1
His ₆ -TRIM24 + 100 μ M H3K14Ac	51.7	51.5	51.6	51.6	0.6	0.1

Sample	T _m values (°C)			Average T _m (°C)	Average T _m shift (°C)	Standard deviation of T _m
His ₆ -TRIM24 + 1% DMSO	-	51.3	51.0	51.1	0.1	0.2
His ₆ -TRIM33 buffer reference	54.5	54.1	54.0	54.2	-	0.2
His ₆ -TRIM33 + 100 µM SGC-TIF1α53	54.4	54.0	54.2	54.2	0.0	0.2
His ₆ -TRIM33 + 50 µM SGC-TIF1α53	54.2	54.0	54.0	54.1	-0.1	0.1
His ₆ -TRIM33 + 25 µM SGC-TIF1α53	54.2	54.0	54.1	54.1	-0.1	0.1
His ₆ -TRIM33 + 100 µM H3K14Ac	-	54.0	54.0	54.0	-0.2	0.0
His ₆ -TRIM33 + 1% DMSO	54.5	54.2	54.1	54.2	0.1	0.2

2.4.3 AlphaScreen

Amplified luminescent proximity homogeneous assay screen (AlphaScreen) is a competition assay that detects interactions in a proximity dependent fashion using singlet oxygen transfer between beads (Figure 2.11). Developed from a diagnostic assay technology known as luminescent oxygen channelling assay (LOCI),¹⁷ it uses simple ‘mix and read’ protocols, but requires a suitable affinity binding partner, and optimisation of signal for each system. This technique can be used for screening or characterising binding partners through titration curves.

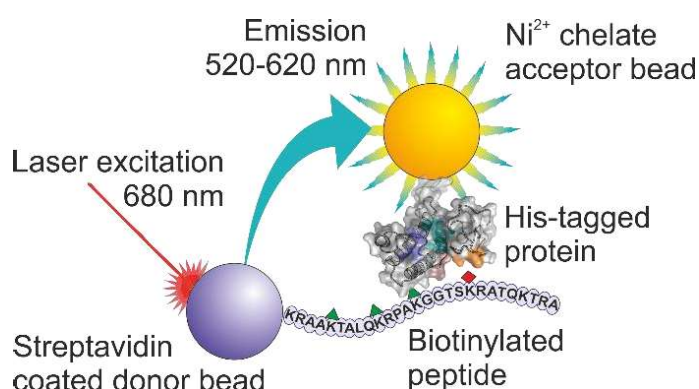


Figure 2.11 Principles of AlphaScreen assay. Streptavidin-coated donor beads associate with biotinylated peptides. Nickel chelating acceptor beads associate with polyhistidine tagged proteins. Singlet oxygen transfer from donor to acceptor bead leads to emission when beads are in close proximity (<200 nm).

2.4.3.1 Assay conditions

Assay protocols were based on those developed by the SGC for use in bromodomain screening.¹⁸ Use of a streptavidin-biotin system was advantageous due to the very high affinity of this binding interaction. As a result, most of the biotinylated component will be associated with the bead, rather than being free in solution, and the IC₅₀ for the corresponding peptide will be less affected by this equilibrium. The commercial recommendation for final assay concentrations of AlphaScreen donor and acceptor beads is 20 µg/mL. However, it was observed that an initial dilution of 1:400 or 1:300, to a final assay concentration of 5.0-6.7 µg/mL, was sufficient for the required levels of signal.

Assay buffer contained additives to prevent aggregation, ensure solubility and stability of components, and to reduce non-specific signal. These were within the limits reported for their use in AlphaScreen: CHAPS (detergent) ≤0.02%, bovine serum albumin (BSA) ≤0.5%. Correct handling of assay plates proved vital for reliable signal. Trials in which the plate was continuously mixed during incubation resulted in very low signals. Subdued laboratory lighting of less than 100 lux is recommended, and beads had to be protected from exposure to excessive ambient light at all times. Final protocol for incubation was to seal the plates with foil (Aluminium StarSeal, Starlab), mix briefly on an orbital shaker (10 sec, 600 rpm), and then incubate in the dark at room temperature.

AlphaScreen signals were always corrected for crosstalk before analysis. ProxiPlate-384 Plus microplates (Perkin Elmer) have white well walls for maximum reflection of light, which results in higher signal for luminescent assays, but medium levels of cross-talk, depending on the concentration of titanium dioxide used as whitener. The software crosstalk correction tool measures the contributions from glow and afterglow of adjacent samples and reduction in signal from bleaching, in order to correct for these effects in the calculation of the final results.

2.4.3.2 Signal response curves for biotinylated peptides

The dominant hypotheses for selectivity of TRIM PHD-bromodomain-peptide recognition include acetylation at the K18 and K23 positions.¹⁹ There is currently no systematic approach to analysing binding partner preferences between TRIM24 and TRIM33 with the inclusion of K9Me₃. It was sought to identify the highest affinity endogenous partners of these two homologous proteins by measuring the affinity of various H3-like peptides bearing combinations of acetylation at positions 14, 18 and 23, and methylation at position 9.

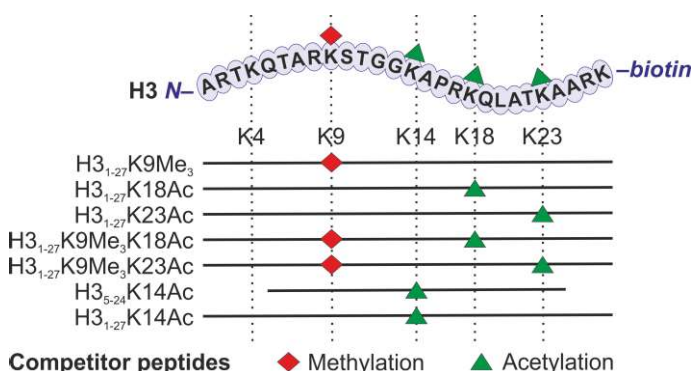


Figure 2.12 Peptide sequences of biotinylated competitor peptides used for development of AlphaScreen assays with TRIM33 and TRIM24. Seven peptides bearing various combinations of methylation (red diamond) at K9, and acetylation (green triangle) at K14, K18 or K23. Comparison of the different length peptides bearing the K14Ac modification demonstrated the effect of inclusion of K4Me₀.

Peptide sequences were designed to incorporate the K4Me₀ residue, which is recognised by the β -sheet surface of the PHD finger (Figure 2.12). It was hoped that inclusion of this residue could facilitate extension of AlphaScreen procedures to TRIM33, since the literature suggested that the

eight *N*-terminal residues of the H3 peptide interact with the PHD finger and control affinity.²⁰ Comparison of signal generated with the truncated H3 peptide (lacking four *N*-terminal residues), which has been used in an AlphaScreen for TRIM24, would confirm the contribution of this portion of the peptide to affinity for TRIM33. It had been well established that methylation of the K4 position reduces affinity to the PHD finger: affinities for TRIM24 have been measured by ITC as 8.6 μ M for H3₁₋₁₅K4Me₀, 41 μ M for H3₁₋₁₅K4Me₁, 198 μ M for H3₁₋₁₅K4Me₂, and >400 μ M for H3₁₋₁₅K4Me₃.⁸ This residue was exclusively incorporated in an un-methylated state (K4Me₀), to avoid affinities falling below detection limits. Biotin groups were incorporated into peptides through the side chain of a *C*-terminal lysine, either the naturally occurring K27 or an additional A25K substitution.

Generation of an AlphaScreen signal is most efficient when a high affinity binding partner is used as the competitor peptide. Screening has been carried out on TRIM24 using H3K14Ac,² however, this position of acetylation is not preferentially recognised. Most recently, hit identification has been performed for TRIM24 using the biotinylated H3₁₋₃₃K23Ac peptide.¹

Optimisation for final assay concentrations of protein and peptide was performed for a range of biotinylated competitor peptides with each batch of protein or beads used, by obtaining signal response curves for varying peptide concentrations (Figure 2.13). These enabled identification of which peptides produced sufficient signal, and the potential for assays responsive to inhibition at either the bromodomain or PHD finger. Final assay parameters were selected on the basis of minimising protein concentration, to avoid shifting the apparent K_i , and using a corresponding peptide concentration to provide a suitable signal to noise ratio. Lowering the protein and peptide concentration would also result in a more sensitive assay.

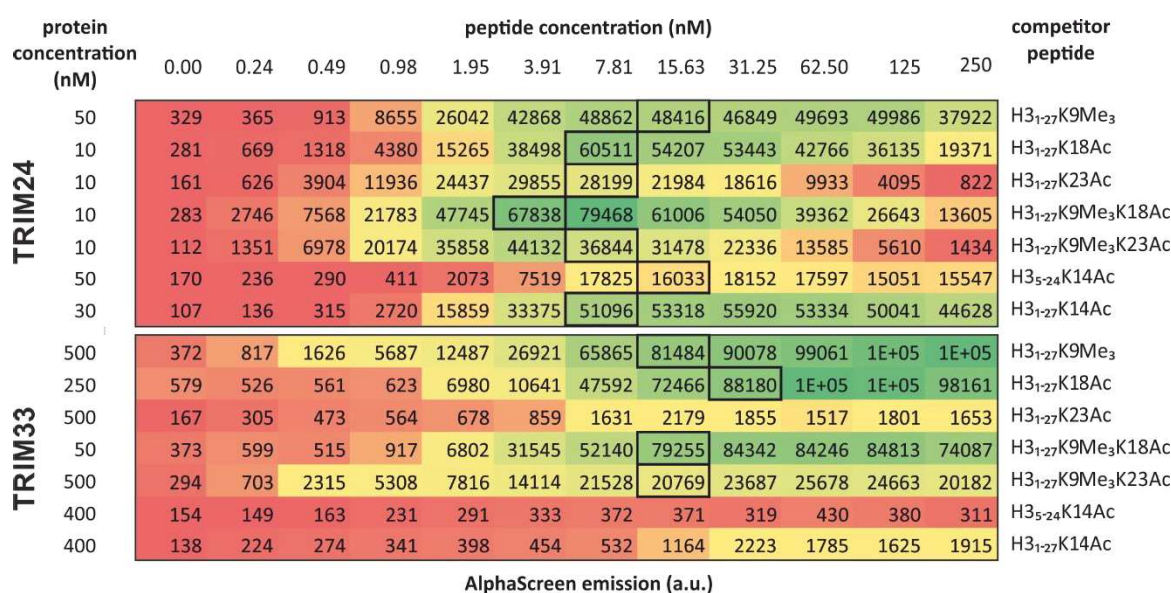


Figure 2.13 AlphaScreen signal response for biotinylated peptides bearing various modifications. Protein final assay concentration indicated on left; competitor peptide identity indicated on right. Colour scale from low emission (red) to high emission (green) showing emission after incubation with peptide in a 1:2 serial dilution series from 250 μ M. Black squares highlight optimised conditions for each protein and peptide pair; rows without selection did not yield sufficient signal for use in assay.

As expected, the signal from H3₅₋₂₄K14Ac was much lower than that from H3₁₋₂₇K14Ac. In the case of TRIM24 these both required higher protein concentrations of 50 μ M and 30 μ M, respectively, and in the case of TRIM33, neither of these concentrations reached sufficient levels of signal for

use as a competitor peptide. TRIM33's preferential recognition of the K18Ac modification, as reported in the literature, was immediately evident from the fact this was the only position of acetylation which resulted in sufficient signal from a singly modified peptide: maximum emission from K14Ac and K23Ac was 2223 a.u. and 2179 a.u., respectively.

Theoretical binding capacities of streptavidin coated donor and nickel (Ni^{2+}) chelate coated acceptor beads are 30 nM and 0.3-1 μM , respectively. If the capacity of the system is exceeded the beads become saturated and a 'hook effect' is observed i.e. reduction of signal with increasing concentrations of binding partners. This bead capacity is influenced by size of the protein and the affinity of its association with the bead. Although the theoretical maximum capacity of a Ni-NTA bead for a polyhistidine protein is ~ 30 nM, because the His-Ni-NTA interaction is relatively weak, and more protein is dissociated at equilibrium, there is a higher operational limit. This was observed for optimised protein concentrations of TRIM33, which were in the much higher range of 50-500 nM. Optimised peptide concentration were always ≤ 30 nM, and the 'hook effect' was clearly observed when low protein concentrations were used in combination with higher affinity competitor peptides, e.g. 10 μM TRIM24 with $\text{H3}_{1-27}\text{K23Ac}$ (emission 822 a.u. with 250 μM peptide) or $\text{H3}_{1-27}\text{K9Me}_3\text{K23Ac}$ (emission 1434 a.u. with 250 μM peptide).

2.4.3.3 IC_{50} value determination for modified peptides

Optimised conditions for the AlphaScreen assay format were used to probe modification recognition preferences for TRIM24 and TRIM33, by quantifying the ability of H3-like peptides, bearing single or multiple modifications, to displace competitor peptides (Figure 2.14). The IC_{50} value

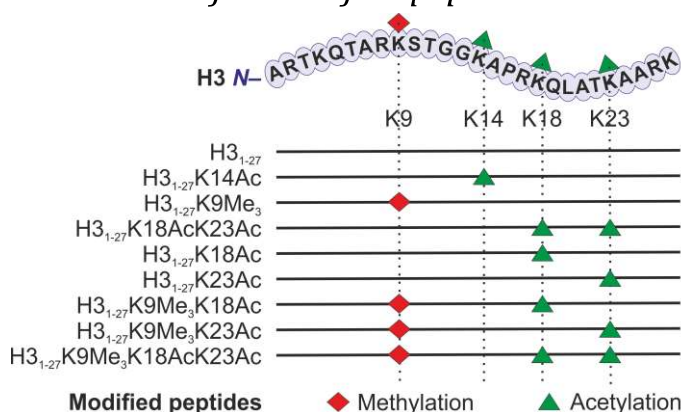


Figure 2.14 Peptides under investigation to probe recognition modes of TRIM33 and TRIM24. Nine peptides bearing various combinations of methylation (red diamond) at K9, and acetylation (green triangle) at K14, K18 or K23, including an unmodified H3 peptide.

was defined as the concentration of peptide required to reduce the AlphaScreen emission signal to

half that of its maximum, after subtraction of base signal. Data fitting was performed according to a full 4-parameter equation, where the x axis represents the (logarithmically scaled) concentration of analyte, range is the maximum y range (response minus background), and s is a slope factor:

$$y = \frac{\text{range}}{1 + \left(\frac{x}{IC_{50}}\right)^s} + \text{background}$$

Since it was known that the affinity of the competition peptide can have a large effect on measured IC_{50} values, we used a range of protein-peptide pairs to perform titration experiments and obtain dose response curves. In the case of TRIM24 it was possible to conduct the assay with all seven biotinylated peptides which had been designed, however, the lower response of TRIM33 meant that only three ($H3_{1-27}K18Ac$, $H3_{1-27}K9Me3K18Ac$, and $H3_{1-27}K9Me3K23Ac$) produced sufficient signal to be used in a displacement assay.

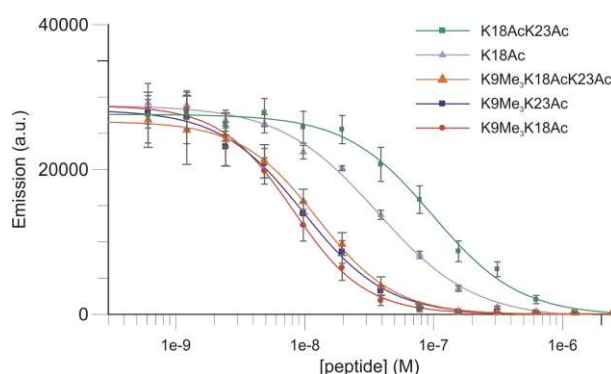


Figure 2.16 Representative AlphaScreen dose response curves for peptide titrations with TRIM24 PHD-bromodomain. Modified peptides (legend) titrated against TRIM24 with $H3_{1-27}K23Ac$ competitor peptide. Curves span IC_{50} well and plateau at maximum and minimum. Concentration ranges were modified for weaker binding peptides to ensure coverage.

The majority of the titration curves spanned the range of inhibition well (Figure 2.16), with the exception of unmodified $H3_{1-27}$ which had a very low affinity to both proteins. Curve fitting to approximate an IC_{50} for $H3$ was not possible for TRIM33, since there was only ~52% inhibition at the maximum concentration evaluated of 50 μM (Figure 2.15).

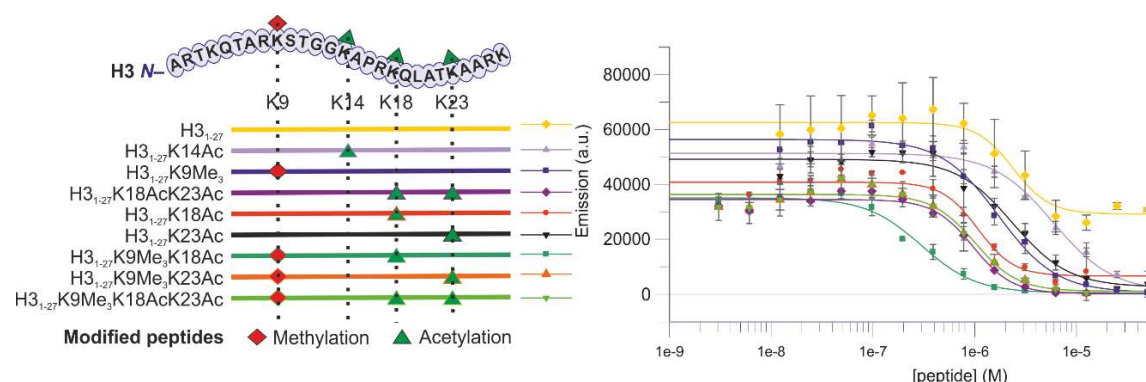


Figure 2.15 Dose response curves for modified peptides displacing $H3_{1-27}K9Me3K23Ac$ from TRIM33 in AlphaScreen assay. Titration curves span range well, but $H3$ peptide did not result in full inhibition. Error bars show standard deviation of results in triplicate; curve fitting was performed according to a 4-parameter fit in GraFit 7.0.²³

Affinities to each of the modified peptides could be clearly grouped into weak, medium, and tight binders (Figure 2.17). This correlated to the number of recognised modifications, with K9Me₃ only having a small affect in isolation, but a substantial cumulative effect in addition to K18/23Ac. Peptide displacement data seem to suggest the possibility of a switch in preferential recognition site, in conjunction with trimethylation. Although assay sensitivity might be limiting, TRIM24 demonstrated a higher affinity towards the K23Ac position in isolation (TRIM24 IC₅₀ of 13 nM for K23Ac vs 26 nM for K18Ac), but towards the K18 position when in combination with K9Me₃ (TRIM24 IC₅₀ of 2 nM for H3₁₋₂₇K9Me₃K18Ac vs 5 nM for H3₁₋₂₇K9Me₃K23Ac).

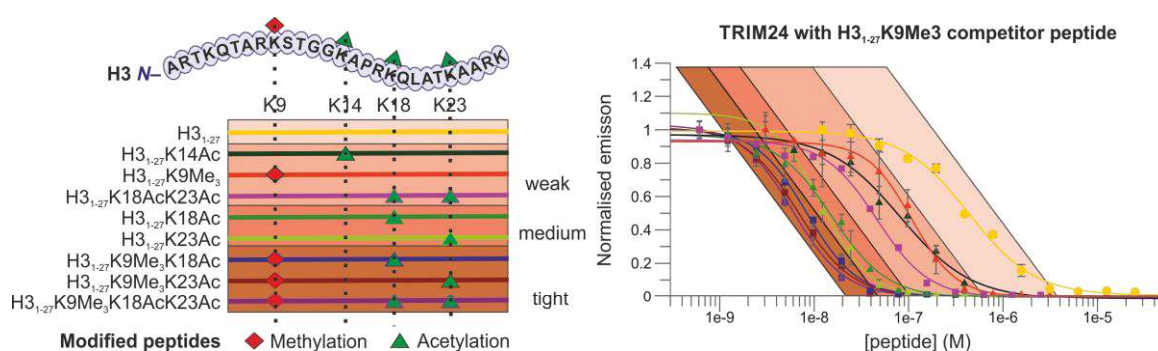


Figure 2.17 Categorisation of histone peptides as weak, medium, or tight binders. Correlated well with number of modifications the histone peptide contained. Normalised AlphaScreen dose response curves (right) from titration into TRIM24, displacing biotinylated H3₁₋₂₇K9Me₃ peptide

A screening window coefficient has been defined as a statistical characteristic parameter for the quality of assays.²² Termed the Z-factor (Z'), this takes into account variations associated with sample measurement (σ_s , the standard deviation for the samples), reference measurements (σ_c , the standard deviation for controls), and the dynamic signal range ($\mu_s - \mu_c$, the assay window or average maximum minus average minimum signal): $Z' = 1 - \frac{(3\sigma_s + 3\sigma_c)}{|\mu_s - \mu_c|}$. A value between 0.5 and 1.0 indicates that the separation band is large and it is classified as an excellent assay. For peptide IC₅₀ curve assay formats, these were ~0.5 and with the exclusion of peptides which did not result in 100% inhibition over the concentration range observed, (e.g. unmodified H3 peptide) was up to ~0.6. Geometric means of the affinities of each peptide were calculated, based on dose response curves with each competitor peptide used (Table 2.2).

Table 2.2 Summary of affinities of modified H3 peptides to TRIM24 and TRIM33. Calculated as geometric means from AlphaScreen dose response curves with various competitor peptides. Curve fitting was performed using four parameter fit in Grafit 7.0.²³

Peptide	TRIM24 IC ₅₀ (μM)	TRIM33 IC ₅₀ (μM)
H3 ₁₋₂₇ K14Ac	0.531	6.654
H3 ₁₋₂₇ K18Ac	0.026	1.163
H3 ₁₋₂₇ K23Ac	0.013	2.881
H3 ₁₋₂₇ K9Me ₃	0.222	1.840
H3 ₁₋₂₇ K9Me ₃ K18Ac	0.002	0.234
H3 ₁₋₂₇ K9Me ₃ K23Ac	0.005	1.070
H3 ₁₋₂₇ K18AcK23Ac	0.014	0.709
H3 ₁₋₂₇ K9Me ₃ K18AcK23Ac	0.005	1.070
H3 ₁₋₂₇	0.501	n.d.

2.4.4 Inactivating mutations

Generation of constructs containing inactive mutations in either the bromodomain or PHD finger, as well as the cross mutations designed to be inactive across both domains, was carried out for both TRIM24 and TRIM33. These mutants could be used to confirm specific recognition by each domain, through loss of affinity to either the inactivated PHD or bromodomains. After mutagenesis by PCR (see Materials and Methods for experimental procedures and primers used), plasmids containing modified protein constructs were expressed and purified with the same protocol as the wild-type (WT), with comparable yields and purity.

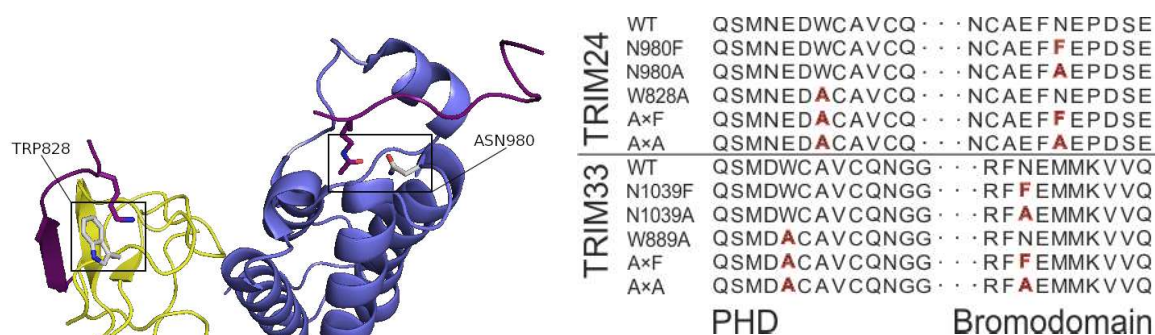


Figure 2.18 Inactivating mutations for PHD and bromodomains in TRIM24 and TRIM33. Left: position of mutations shown for TRIM24 (PDB 3O34,3O37), mutated residues in box shown as grey sticks. Right: sequence alignments of plasmids generated through mutagenesis resulting in single, and double inactive mutations (highlighted in red).

Secondary structures of purified proteins bearing the various mutations were compared by circular dichroism (CD). The typical bromodomain structure appeared to dominate spectra, with all constructs displaying a high proportion of α -helical content (Figure 2.19). Spectra for proteins with inactivating mutations in the bromodomain (TRIM24 – N980A, N980F; TRIM33 – N1039F, W889A) were highly similar, suggesting these mutations do not have a structural effect: average standard deviations 0.7 millidegrees and 0.4 millidegrees for TRIM24 and TRIM33 bromodomain mutations, respectively. Mutations of the PHD finger seemed to show a greater deviation of secondary structure from the WT than those of the bromodomain, particularly at lower wavelengths. This would suggest that substitution of this residue results in some change in structure of the PHD finger, although the relative forms of these curves were still similar, and slight differences could have been due to variations in sample concentration during data collection.

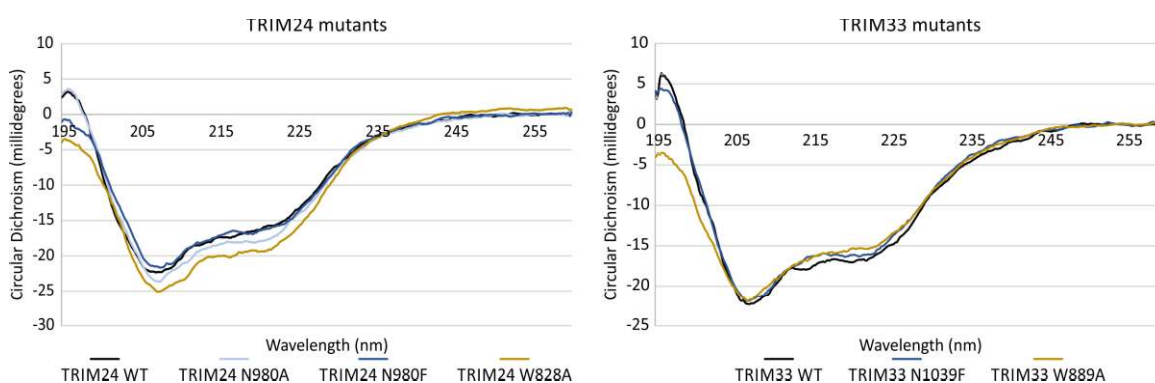


Figure 2.19 Circular Dichroism spectra from TRIM24 (left) and TRIM33 (right) proteins bearing bromodomain or PHD inactivating mutations. Smoothed average of results in triplicate with 0.2 nm step size, 0.5 sec/step, 10 mm path length, 0.2 mg/mL protein in NaF buffered phosphate buffer (pH 7.8).

Signal response curves for inactive mutants with biotinylated peptides in the AlphaScreen assay provided interesting insight into contributions of each domain to recognition. Mutation of the cation- π -forming tryptophan residue completely abolished PHD mediated recognition of KMe₃, and generated no signal in the AlphaScreen assay when incubated with up to 250 nM of a peptide bearing this modification (H3₁₋₂₇K9Me_s). The maximum AlphaScreen emissions with biotinylated H3₁₋₂₇K9Me_s were 675 a.u. and 3191 a.u. for TRIM24 W828A and TRIM33 W889A, respectively.

An alanine mutation of the conserved asparagine which forms a hydrogen bond to KAc residues (TRIM24 N980A) seemed insufficient to block TRIM24 association to peptides bearing this

modification, suggesting broader interactions within the KAc pocket contribute highly to affinity. However, a phenylalanine mutation (TRIM24 N980F) did hinder isolated KAc recognition, presumably by restricting access to the KAc recognition pocket of the bromodomain, while still allowing recognition of KMe₃. Surprisingly, a corresponding N1039F mutation did not hinder the ability of TRIM33 to recognise KAc-containing peptides to the same degree. This result supports the idea that the α B helical extension in this isoform does indeed shift the asparagine orientation away from the binding pocket, allowing mutations to bulkier residues to be better tolerated. Furthermore, the reduced effect of bromodomain inactivation, compared to PHD finger inactivation, also suggests that peptide affinity to this α -isoform of TRIM33 is dominated by interactions with the *N*-terminus of the H3 tail, and PHD recognition of KMe₃ modifications may exert a larger influence than affinity of a bromodomain to KAc.

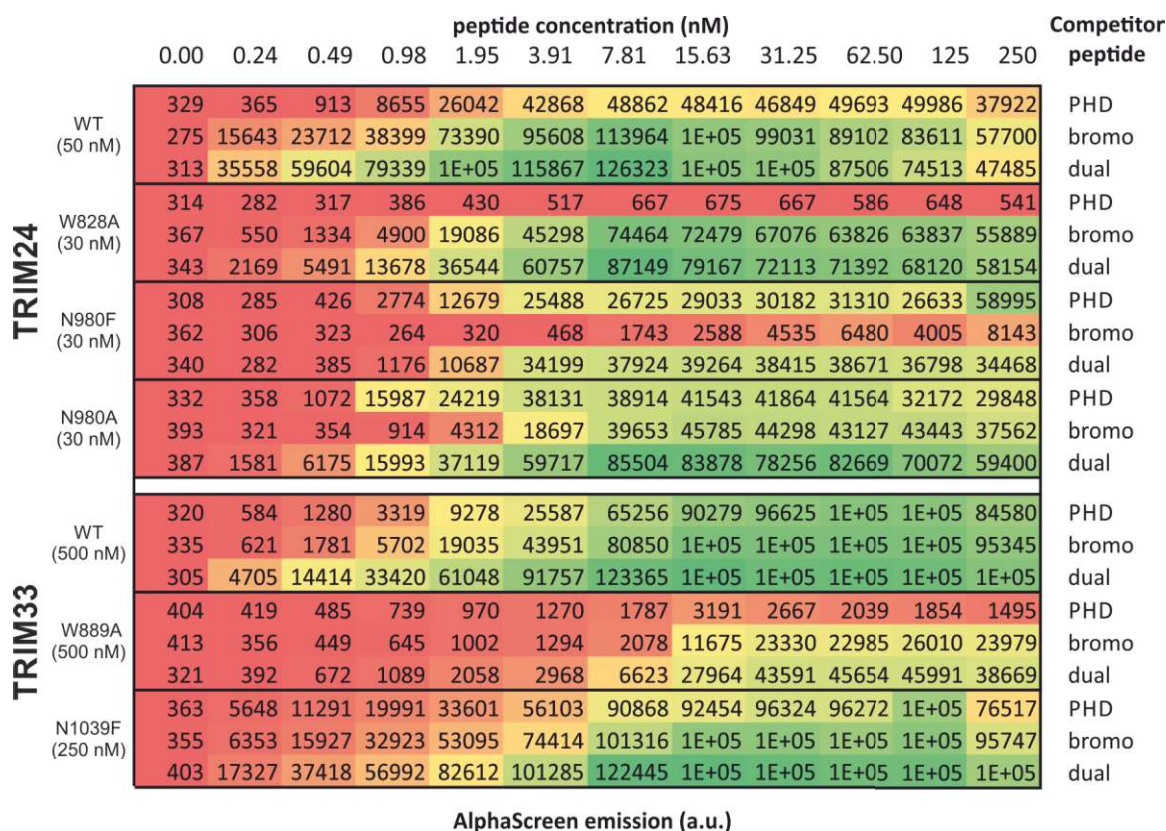


Figure 2.20 AlphaScreen signal response for biotinylated peptides with proteins bearing inactive mutations. Protein mutations indicated on left with final assay concentration. Competitor peptides indicated on right: PHD – H3₁₋₂₇K9Me₃; bromo – H3₁₋₂₇K18Ac; dual – H3₁₋₂₇K9Me₃K18Ac. Colour scale from low emission (red) to high emission (green) showing emission after incubation with varying peptide in a 1:2 serial dilution series from 250 nM. Red bars across the titration indicates lack of association i.e. TRIM24 W828A with PHD peptide, TRIM24 N980F with bromo peptide, TRIM33 W889A with PHD peptide.

2.4.5 Surface plasmon resonance

Surface plasmon resonance (SPR) is a highly informative technique, which was attractive for profiling interactions between TRIM proteins and modified H3-like peptides, because kinetic information could be obtained as well as affinities of interactions. Additionally, relatively low volumes of samples are typically employed, hence it could make efficient use of material.

It was preferable to maintain an assay format as similar as possible to the AlphaScreen, to obtain comparable results. The setup deemed most suitable, involved the use of a CM5 chip (carboxymethylated dextran coating on a gold surface), with attachment of a biotinylated peptide (*via* streptavidin) as the surface bound ligand, and various concentrations of TRIM24 as the analyte flowing over the cell in solution.

All four cells of the chip were coated with streptavidin following standard procedures, and immobilisation resulted in a signal of ~4000 response units (RU). Given the maximum response for an interaction $R_{max} = \frac{MW_{analyte}}{MW_{ligand}} \times R_L \times N$ (where N = stoichiometry, R_L = response for ligand, MW = molecular weight), this would give an expected response from biotinylated peptide binding of $\frac{3000}{13000} \times 4000 \times 1 \approx 920$ RU. Indeed, subsequent addition of biotin-H3₁₋₂₇K9Me₃ peptide generated a stable surface (subject to weak acid washes of Glycine pH 3.0) with an increase in baseline signal of ~400 RU.

Comparison of signal generated from protein analyte (120 sec injection, 30 µL/min), revealed higher levels of non-specific binding to the streptavidin coated reference cell than the flow cell containing the peptide, resulting in a negative signal in the sensogram. Moreover, a 30 sec Glycine wash (pH 3.0) was insufficient to remove all bound species, and starting baseline signal increased for each cycle (data not shown).

In an attempt to increase the signal to noise ratio, a higher affinity peptide was employed (biotin-H3₁₋₂₇K9Me₃K18Ac) along with supplementing running buffer with 0.05% BSA to reduce background non-specific binding. Additionally, it was hoped that two cycles of a stronger acid wash

(Glycine pH 2.0) might more efficiently recondition the surface. Although there was still a high level of binding to the reference cell, and an increasing baseline over the course of the experiment due to incomplete surface reconditioning, the traces were more informative (Figure 2.21). Maximum binding equilibrium was reached within the injection period at high concentrations (>6.4 μM), and demonstrated good K_{on} kinetics. However, a sharp drop in response immediately after injection, followed by a persistently high response which was stable for the entire dissociation period, suggested two independent binding events, with a low K_{off} and very high K_{off} , respectively. Fitting of a single site binding mode to the data yielded a K_{D} of 4.83 μM : since the drift in baseline was uncorrected, we would expect this value to be higher than that of the true constant.

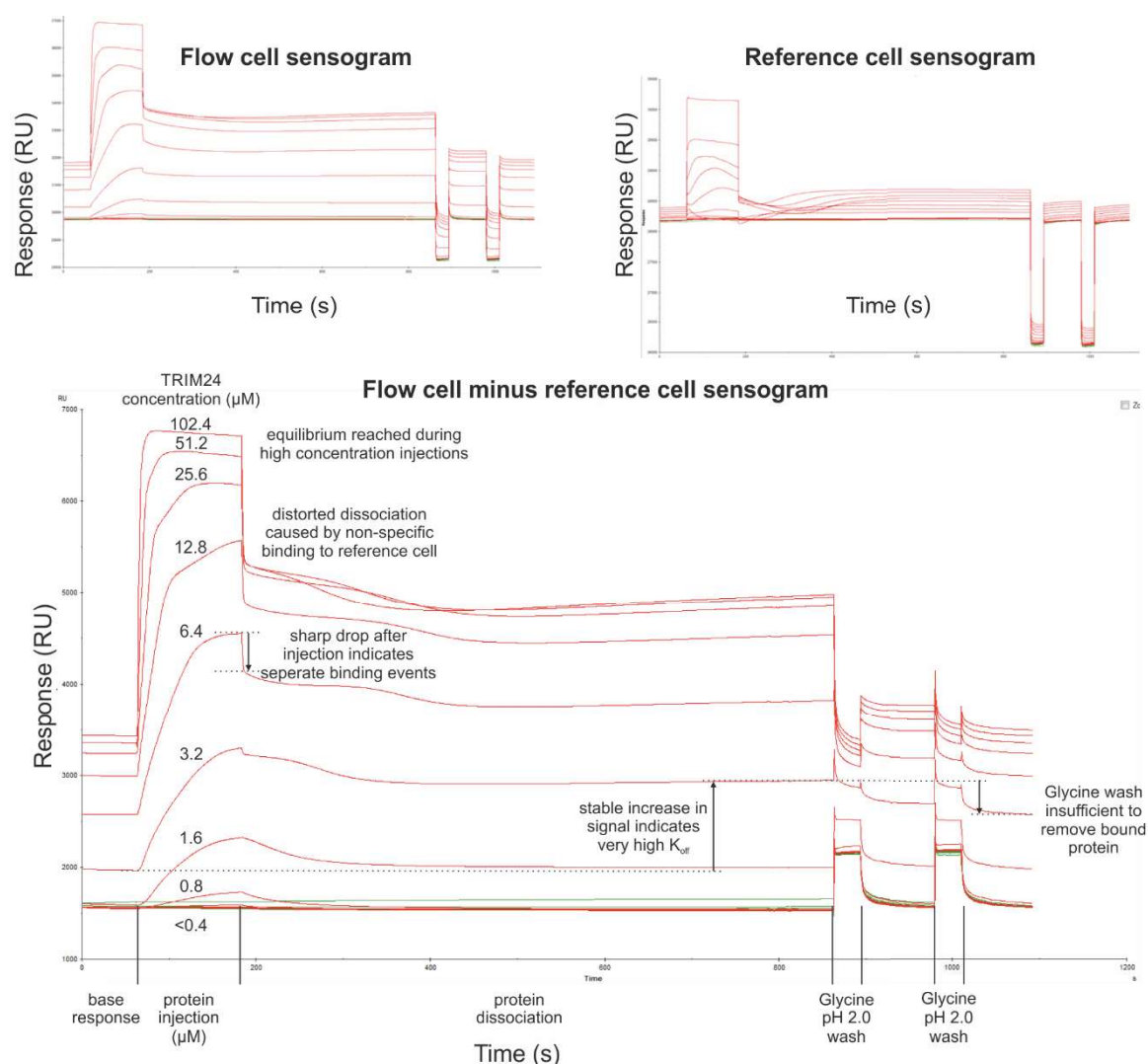


Figure 2.21 Sensograms from surface plasmon resonance experiments: CM5 gold chip; streptavidin covalently coupled with NHS/EDC; biotin-H3₁₋₂₇K9Me₃K18Ac ligand; His₆-TRIM24 analyte. Five start-up cycles (buffer) were followed by two-fold increases in analyte concentration. Raw sensograms are shown above as flow cell and reference cell: the later consisted of streptavidin coating in the absence of peptide ligand. Annotations on lower sensogram highlight key features of analysis.

An alternative NTA Sensor Chip (carboxymethylated dextran pre-immobilized with nitrilotriacetic acid), which can immobilise a His₆- tagged biomolecule on a nickel chelating surface, and a CAP Sensor Chip (carboxymethylated dextran matrix pre-immobilized with a ssDNA oligo), which can be more reliably regenerated through immobilisation of a biotinylated ligand *via* a ssDNA oligo-streptavidin agent, were also used but suffered from similar difficulties in signal-background ratio from high levels of non-specific binding.

2.4.6 Isothermal titration calorimetry of selected peptides

In order to confirm the affinities of the H3-mimicking peptides for TRIM24 and TRIM33 suggested by AlphaScreen, we sought to quantify the strength of binding interactions using an orthogonal technique. As discussed in the Introduction, isothermal titration calorimetry (ITC) is considered highly valuable because it can detect direct binding interactions with high precision, and define enthalpic and entropic contributions to affinity. Since this technique is highly resource intensive, it was used to profile each of the positions of modification which had been used as competitor peptides in AlphaScreen assays independently: K9Me₃ and K18Ac. Exhaustive dialysis of both components in the same container is ideal, however, this was not possible due to the inability to quantify the peptide concentration after analysis. Dialysis of the protein, followed by dissolution of lyophilised peptide in the same buffer, was able to produce a sufficient buffer match for the titration. Although HEPES buffer has a relatively high enthalpy of ionisation (20.4 kJ/mol), protein stability had been established in this buffer system, and attempted use of phosphate buffers resulted in higher heats of dilution.

Titration into cells containing TRIM24 produced good sigmoidal binding isotherms, with a clear plateau for high concentrations of peptide (Figure 2.22). After subtraction of signal resulting from the heat of dilution (titration of peptide into dialysed buffer), the Origin built-in curve fitting function (one set of sites) was used to fit data and calculate constants. Binding of K9Me₃ was entropically driven, demonstrated by the positive value of $\Delta S = 4.62$ cal/mol/deg and common for hydrophobic interactions. Binding of K18Ac, had a higher enthalpic contribution to Gibbs free

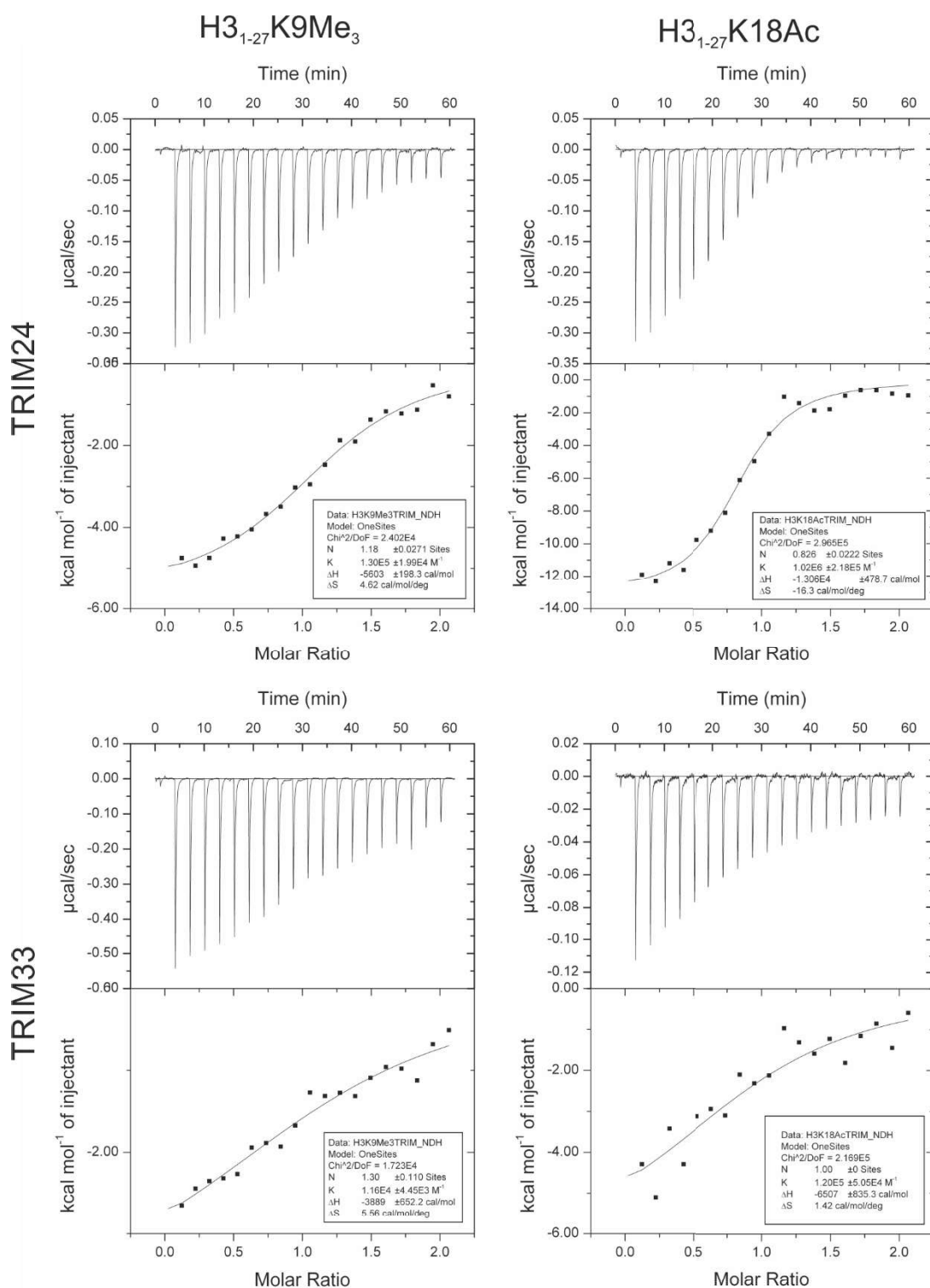


Figure 2.22 Isothermal titration calorimetry data for competitor peptides used in AlphaScreen against TRIM24 and TRIM33. Peptide ligand was titrated into the cell containing the protein in a series of 19×2 μL injections (first injection 0.4 μL). Concentrations for titrations: 750 μM H3K9Me₃, 50 μM TRIM24; 200 μM H3K18Ac, 20 μM TRIM24; 1.67 mM H3K9Me₃, 150 μM TRIM33; 200 μM H3K18Ac, 20 μM TRIM33. Low affinities of TRIM33 to peptides resulted in inaccurate fitting due to range of binding curve observed. Heat of dilution (from titration into buffer) was subtracted, and the first data point excluded from analysis. Error values represent errors of the nonlinear least-squares fit. Buffer conditions: 20 mM HEPES (pH 7.8), 150 mM NaCl.

energy ($\Delta H = -13$ kcal/mol), which is as expected from hydrogen-bonding and van der Waals interactions. Calculated K_D values were $7.69 \mu\text{M}$ for H3₁₋₂₇K9Me₃ and 980 nM for H3₁₋₂₇K18Ac, which are approximately 30-fold higher than IC_{50} values obtained through AlphaScreen.

Titration into cells containing TRIM33 showed a clear binding curve, but the peptide was not of sufficient concentration to reach saturation, and the data were hard to interpret with a sigmoidal fit (Figure 2.22). In the case of H3K18Ac, the stoichiometry was fixed to $N = 1.0$ for analysis in a one site model. Nonetheless, K_D values reinforced the higher affinity for recognition of acetylation relative to methylation, which had been previously observed: $86.2 \mu\text{M}$ for H3₁₋₂₇K9Me₃ and $1.16 \mu\text{M}$ for H3₁₋₂₇K18Ac.

2.5 Conclusions

Overall these data provide strong evidence to suggest that the PHD finger and bromodomain of TRIM proteins operate co-operatively, with a very strong increase in affinity towards peptides bearing multiple modifications. Both TRIM24 and TRIM33 demonstrated a significant increase in binding interaction with peptides bearing the K9Me₃ modification in addition to acetylation. Since the domains can also function independently, i.e. recognition of either acetylation or trimethylation in isolation, this suggests that combined readout of histone modifications could result in alternative downstream effects.

The remarkable affinity of TRIM24 for multiple modifications to H3 proteins cannot be accounted for by dual recognition alone: TRIM33 containing a highly homologous PHD and bromodomain has approximately 10 fold lower affinity. Either the relative strength of affinity towards modified peptides is an inherent property of each TRIM protein, or the specific combination of histone modifications resulting in the highest affinity towards TRIM33 has yet to be identified.

Clearly *N*-terminal tail recognition has a considerable effect on affinity, and in the case of TRIM33 is absolutely necessary for a tight binding interaction to be observed by AlphaScreen. The requirement of inclusion of both the terminal K4Me₀ and the K9Me₃ position in competitor peptides, in order to generate signal in the AlphaScreen assay, suggest TRIM33-H3 interactions are dominated by PHD finger recognition. However, it should be noted that contribution of affinity for peptides to TRIM33 seems to be less dependent on the conserved asparagine residue, than for TRIM24. This was demonstrated by the reduced effect of a bromodomain inactivating mutation in TRIM33. Domain inactive mutants generated in this work could be used to study contributions from each domain for affinity to multiply modified peptides, by measuring whether any decrease in IC₅₀ for peptides correlates to which domain is inactivated.

It is hypothesised that the less well studied β -isoform (GenBank Accession: NP_148980) might be crucial in dissecting the relative contributions of each domain in TRIM33. It seems likely that removing the extension to the α B-helix would allow the asparagine residue (N1039) to lie in a more conventional orientation towards the KAc binding pocket, greatly increasing affinity towards KAc modifications. This has been affirmed by homology modelling of the alternative isoform (Figure 2.23), and similar studies on TRIM33 β might result in affinities to modified peptides in the same order of magnitude as for TRIM24.

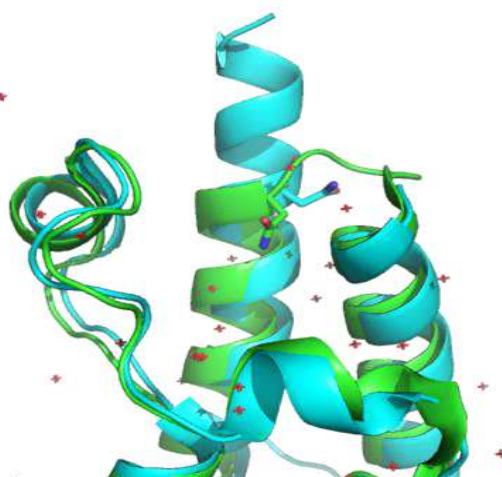


Figure 2.23 Homology model of TRIM33 α carried out by Lewis Vidler (Sven Hoelder Group, ICR). Partial view of structure of TRIM33 β from PDB code 3U50 (blue) overlaid with homology model of TRIM33 α (green). This suggests that removal of the α B-helical extension would allow the key asparagine residue to face the binding pocket in a more standard orientation. This could result in peptide binding affinities closer to the range observed for TRIM24.

2.6 References

- (1) Palmer, W. S., Poncet-Montange, G., Liu, G., Petrocchi, A., Reyna, N., Subramanian, G., Theroff, J., Yau, A., Kost-Alimova, M., Bardenhagen, J. P., Leo, E., Shepard, H. E., Tieu, T. N., Shi, X., Zhan, Y., Zhao, S., Draetta, G., Toniatti, C., Jones, P., Geck Do, M., and Andersen, J. N. (2015) Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J. Med. Chem.* 10.1021/acs.jmedchem.5b00405.
- (2) Bennett, J. M., Fedorov, O., Tallant, C., Monteiro, O. P., Meier, J., Gamble, V., Savitski, P., Nunez-Alonso, G. A., Haendler, B., Rogers, C., Brennan, P. E., Müller, S., and Knapp, S. (2015) Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *J. Med. Chem.* 59, 1642–1647.
- (3) Herquel, B., Ouararhni, K., Khetchoumian, K., Ignat, M., Teletin, M., and Mark, M. (2011) Transcription cofactors TRIM24, TRIM28, and TRIM33 associate to form regulatory complexes that suppress murine hepatocellular carcinoma. *PNAS* 108, 8212–8217.
- (4) Li, Y., Wu, H., Wu, W., Zhuo, W., Liu, W., Zhang, Y., Cheng, M., Chen, Y.-G., Gao, N., Yu, H., Wang, L., Li, W., and Yang, M. (2014) Structural insights into the TRIM family of ubiquitin E3 ligases. *Cell Res.* 24, 762–5.
- (5) Weng, L., Mitoma, H., Tricot, C., Bao, M., Liu, Y., Zhang, Z., and Liu, Y.-J. (2014) The E3 Ubiquitin Ligase Tripartite Motif 33 Is Essential for Cytosolic RNA-Induced NLRP3 Inflammasome Activation. *J. Immunol.* 193, 3676–3682.
- (6) He, W., Dorn, D. C., Erdjument-bromage, H., Tempst, P., Moore, M. A. S., and Massagué, J. (2006) Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. *Cell* 125, 929–41.
- (7) Herquel, B., Ouararhni, K., and Davidson, I. (2011) The TIF1 α -related TRIM cofactors couple chromatin modifications to transcriptional regulation, signaling and tumor suppression. *Transcription* 2, 231–6.
- (8) Tsai, W.-W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M.-C., Patel, D. J., and Barton, M. C. (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927–932.
- (9) Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.
- (10) Linding, R. (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31, 3701–3708.
- (11) Owens, C. B., and Szalanski, A. L. (2005) Filter paper for preservation, storage, and distribution of insect and pathogen DNA samples. *J. Med. Entomol.* 42, 709–711.
- (12) Savitsky, P., Bray, J., Cooper, C. D. O., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A., and Gileadi, O. (2010) High-throughput production of human proteins for crystallization: The SGC experience. *J. Struct. Biol.* 172, 3–13.
- (13) Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* 18, 6069–6074.
- (14) Vivian, J. T., and Callis, P. R. (2001) Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.* 80, 2093–109.
- (15) Sachchidanand, Resnick-Silverman, L., Yan, S., Mutjaba, S., Liu, W. J., Zeng, L., Manfredi, J. J., and Zhou, M. M. (2006) Target structure-based discovery of small molecules that block human p53 and CREB binding protein association. *Chem. Biol.* 13, 81–90.
- (16) Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* 2, 2212–21.

- (17) Ullman, E. F., Kirakossian, H., Singh, S., Wu, Z. P., Irvin, B. R., Pease, J. S., Switchenko, a C., Irvine, J. D., Dafforn, a, and Skold, C. N. (1994) Luminescent oxygen channeling immunoassay: measurement of particle binding kinetics by chemiluminescence. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5426–5430.
- (18) Philpott, M., Yang, J., Tumber, T., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Keates, T., Felletar, I., Ciulli, A., Knapp, S., and Heightman, T. D. (2011) Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* 7, 2899–2908.
- (19) Xi, Q., Wang, Z., Zaromytidou, A.-I., Zhang, X. H.-F.-F., Chow-Tsang, L.-F., Liu, J. X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., Studer, L., Mark, W., Patel, D. J., Massagué, J., and Massague, J. (2011) A Poised Chromatin Platform for TGF-beta Access to Master Regulators. *Cell* 147, 1511–1524.
- (20) Agricola, E., Randall, R. a R., Gaarenstroom, T., Dupont, S., and Hill, C. S. (2011) Recruitment of TIF1 gamma to Chromatin via Its PHD Finger-Bromodomain Activates Its Ubiquitin Ligase and Transcriptional Repressor Activities. *Mol. Cell* 43, 85–96.
- (21) Halfman, J. (1981) protein followed by addition of labeled analyte). The competitive immunoassay response is most commonly determined by measuring the de- crease in bound labeled analyte associated with increasing concentrations of analyte from standards or specimens . T 74.
- (22) Zhang, J.-H., Chung, T. D. Y., and Oldenburg, K. R. (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* 4, 67–73.
- (23) Leatherbarrow, R. J. (2009) GraFit Version 7. Erithacus Software Ltd. Horley, U.K.

Chapter 3: Lead Identification

3.1 Introduction

With the intention of developing small molecule chemical probes for the tandem reader domains of TRIM24 and TRIM33, work was carried out towards lead identification. A focussed library compiled using structure guided computational docking, as well as a larger collection of compounds, was screened using the assays developed in Chapter 2. Assays based on displacement of peptides containing both acetylated and trimethylated lysines (KAc and KMe₃, respectively), were developed to identify inhibition of each domain. Potential lead molecules were validated using the AlphaScreen technique, and with orthogonal assays.

At the time of commencing of this study on TRIM proteins, there were no small molecule ligands known to interact with these proteins in the literature.^{1,2} Furthermore, there were no known small molecule approaches to inhibiting multiple epigenetic reader domains, or assay protocols which could orthogonally detect binding to two reader domains. Indeed, there are still no reported screening methods of any format for use with TRIM33.

Inhibition of PHD fingers has been shown to be challenging, through both a ‘druggability’ analysis³ and the limited inhibitor development.⁴ However, the wealth of literature on bromodomains has demonstrated that the well-defined KAc binding pocket confers this domain with an exceptionally high ‘druggability’ relative to common protein-protein interaction (PPI)

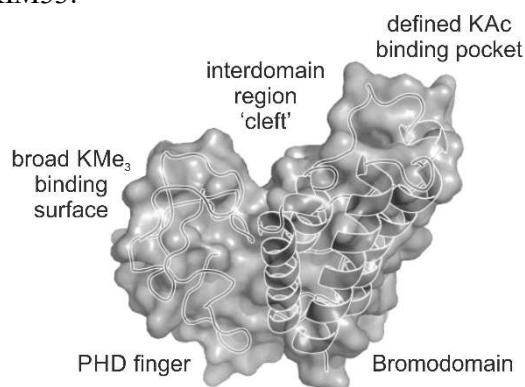


Figure 3.1 Use of dual-domains in screening increases surface available for potential binding partners. The interdomain region presents itself as a novel area for potential binding: the highly “druggable” KAc pocket could act as a handle for inhibition across domains, the broad binding surface of the PHD finger is adjacent to the more clearly defined topology of the interdomain region.

surfaces.⁵ It was hypothesised that by attempting to inhibit the tandem PHD-bromodomain cassette the likelihood of discovering inhibitors of the PHD finger would be greatly improved (Figure 3.1). A dual construct results in the potential for developing ligands that bind not only to the domains

themselves – the KAc binding pocket in the bromodomain, or the β -sheet surface of the PHD finger – but also the new surface created at the interdomain region, which appears to form a ‘cleft’ in the x-ray crystal structure. An extended binding surface, with a more distinct topology at the interface, would provide additional opportunities for affinity-generating interactions in the proximity of the PHD finger.

Identification of a lead compound depends on the ability to reliably detect interactions of putative binding partners. This often requires a package of multiple biophysical assays, in order to test a range of binding hypotheses. It was envisioned that the AlphaScreen assay could be adapted to detect inhibition of binding to either, or both, domains through use of varyingly modified peptides. Confirmation *via* direct biophysical assays could confirm interactions of ligands with TRIM24 or TRIM33.

3.1.1 Aims

With the overall aim of developing small molecule chemical probes of the tandem reader domains of TRIM24 and TRIM33, high-throughput small-molecule screening was carried out using the assays developed in Chapter 2 for profiling recognition of modified peptides (Figure 3.2).

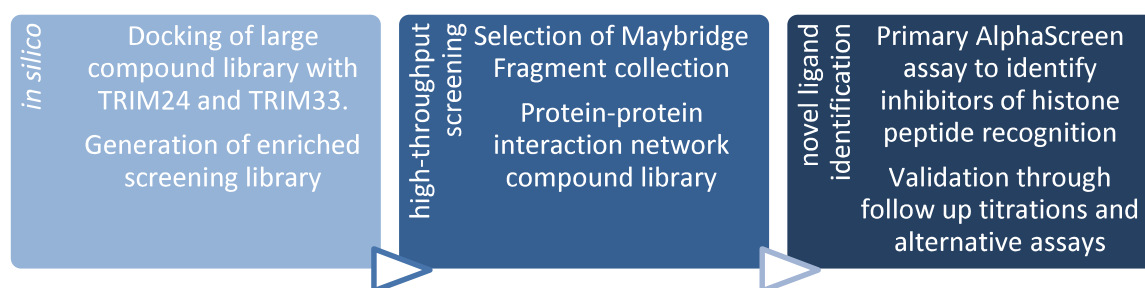


Figure 3.2 Planned work for identification of binding partners of TRIM24 and TRIM33. Stages of work to be undertaken, presented with the main components of each work.

It has been observed that *in silico* screening greatly enriches the possibility of identifying hit compounds through subsequent screening rounds.⁶ In order to increase the efficiency of ligand discovery, this approach will be used to reduce the number of compounds subjected to *in vitro* testing. Conducting docking studies would require: definition of TRIM protein “receptor” states, the structure of the protein of interest and hydration states, which will remain fixed during simulations; generation of appropriate ligand files, incorporating partial charges on atoms for

energetic analysis in addition to shape complementarity; and definition of docking parameters, including the region of interest.

The Maybridge portfolio (Fisher Scientific) contains compound collections tailored to the drug discovery and biotechnology sector. A fragment screening library of ~30,000 small compounds (MW <350) taken from the complete Maybridge collection (53,000 hit-like and lead-like molecules) is available to accelerate identification of lead compounds. Further reduction of this collection, through molecular docking simulations, will be used to bring forward a feasible number of compounds for screening.

The Protein-Protein Interaction Network (PPI-Net; EPSRC funded network EP/I037210/1 and EP/I037172/1), with support from GSK, provides a PPI-net screening collection available on request of 1534 compounds (<http://ppi-net.org/>). These have been selected from an Asinex compound library with chemical properties to favour inhibition of PPIs: a MW limit of <450 Da is higher than the usual 'fragment' limit to enable broader binding surfaces; a cLogP (calculated partition coefficient between n-octanol and water) of <4.5 would fit in the Lipinski guidelines for drug-like molecules;⁷ stereogenic centres create a more complex topology for surface binding, similar to natural products; property forecast index (PFI: $\text{cLogD}_{7.4} + \# \text{ of aromatic rings}$) <6. A study of 240 oral drugs with >30% bioavailability, revealed that 89% of them have a PFI < 7. Since this collection is well suited towards binding of extended surfaces, as presented by the PHD-bromodomain cassette, the entire collection will be screened, as a complement to the more focussed screening of the Maybridge collection.

The aims of this study were to:

- Generate a docking system for TRIM24 and TRIM33 to create an enriched screening library.
- Optimise the AlphaScreen format developed for peptide recognition profiling in Chapter 2, to use in high-throughput compound screening.
- Screen compounds for inhibitory activity of TRIM24 and TRIM33 histone peptide recognition, to identify potential novel ligands.

3.2 Computational approach

For efficient identification of lead hits which bind to TRIM24 or TRIM33, the Maybridge ‘Rule-of-three’ set of fragments, designed to cover diverse chemical space, was docked using AutoDock Vina.⁸ Discovery of many diverse ligands of epigenetic proteins have been initiated with computational docking,^{6,9,10} and AutoDock Vina is an open-source program developed at The Scripps Research Institute with a high average accuracy of binding mode predictions, based on comparisons of docked structures with those available from ligand-bound x-ray crystal structures.

AutoDock Vina uses a rigid conformation of a protein, which can be extracted from a PDB structure, or specified through homology modelling, and is termed the ‘receptor’. Molecules to be docked, termed ‘ligands’, are specified in separate files which incorporate atomic connectivity, and partial charges on atoms. After specification of the search space around the binding site, each run consists of a number (determined heuristically based on the size and flexibility of the ligand) of steps which randomly perturb the conformation, optimise locally (Broyden-Fletcher-Goldfarb-Shanno algorithm¹¹), and select whether to accept the step or not based on an energetic evaluation including the weighted sum of both inter- and intramolecular contributions. Many independent runs, starting from random conformations, aim to identify the global rather than local minimum.

Results for each ligand are merged, refined, clustered and sorted within the implementation to produce the final results. Each potential binding mode is reported as the predicted binding affinity (kcal/mol), along with the bound conformation (.pdbqt file) and the associated root mean square deviation (RMSD) value (relative to the lowest energy mode calculated from only movable heavy atoms).

3.2.1 Receptor state definition

Prior to commencing *in silico* screening methods, accurate configuration of receptor states is vital for reliability, since protein flexibility is not incorporated into docking simulations. For bromodomains, with several highly conserved waters bound in the KAc pocket, hydration can have significant impacts on the binding modes observed in docking. Furthermore, the topology of the shallow surface for interaction across the interdomain cleft region is greatly affected by inclusion of water molecules. A docking model for TRIM24 was refined to incorporate an analysis of water displaceability, as evaluated by WaterClassifier.¹² The structure of TRIM24 (PDB: 3O33) was partially dehydrated to facilitate interactions with molecules that could bind in these positions

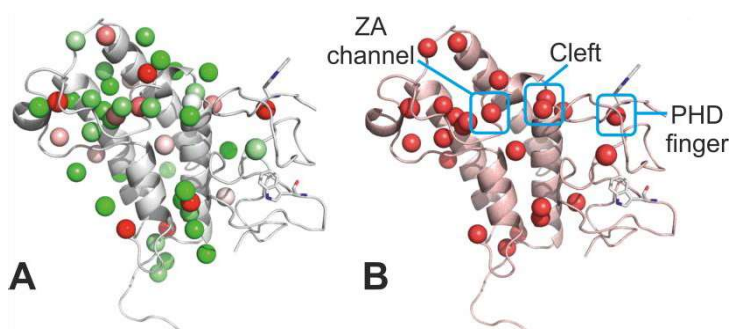


Figure 3.3 **A.** Water molecules (sphere) in the crystal structure of TRIM24 (PDB: 3O33) coloured by displacement probability from 0 (red) to 1 (green). **B.** Hydration state of TRIM24 displaying all water molecules included in molecular docking (red sphere), with conserved water molecules in key positions highlighted (blue).

(Appendix 8.1: colour by displaceability). The refined structure contained key conserved waters in the ZA channel, upper ridge of the cleft between the PHD finger and bromodomain, and beside the KMe₃ region of the PHD finger (Figure 3.3: B).

The low resolution of the available structure of TRIM33 (PDB: 3U5O), and an inability to locate several disordered regions, meant water molecules could not be accurately resolved for the *apo*-structure. Probable water locations were generated with WaterDock,¹² which confirmed the typical water-mediated hydrogen-bonding network at the base of the KAc pocket (data not shown). However, reliability of these predictions were insufficient for subsequent use in docking studies. For comparable methodology, the hydration state of TRIM24 was superimposed onto a water-free *apo*-structure of TRIM33, using alignment of individual domains in the crystal structures with the `cealign` command.

3.2.2 *In silico* screening

The receptor files defined above were utilised in an *in silico* docking screen conducted with the Maybridge Fragment Collection, a database of 31,973 compounds with a molecular weight <350 Da. Configuration files were specified to dock compounds into both the KAc pocket, and the cleft region (Figure 3.4 – Appendix 8.1: `config.txt`). The former specified a 20 Å search area centred on a conserved water in the base of the KAc pocket (HOH109: TRIM24 x=18.18, y=30.80, z=-3.30; TRIM33 x=8.56, y=31.42, z=-3.13), the later used a 20 Å search area centred on a water bound in the cleft region to avoid inclusion of the KAc pocket (HOH144: TRIM24 x=27.19, y=42.64, z=-2.44; TRIM33 x=29.59, y=41.30, z=-1.51).

Definition of individual ligand files from the Maybridge Fragment Collection structure-data file (.sdf) required processing through internal coordinate modelling (ICM: Molsoft) due to poor curation of the data file.¹³ Both

receptor and ligand files were converted into .pdbqt files which define atomic coordinates, partial charges and AutoDock atom types, for use with AutoDock Vina. This conversion was achieved with the use of the `prepare_ligand4.py` Python script in the AutoDockTools (ADT) utility.

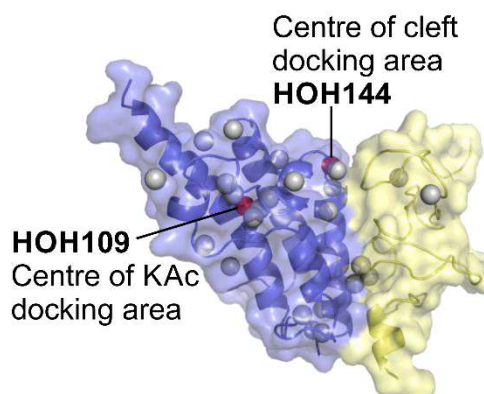


Figure 3.4 Overview of configuration file definition for docking. Water molecules (denoted in PDB files as atom type HOH) were used as centre points for 20 Å search areas in AutoDock Vina

Capability for the screen was provided by use of the Arcus hub of the University of Oxford Advanced Research Computing Facility. Scripts to run AutoDock Vina were submitted to a Torque Resource Manager using a Bourne-again shell (BASH) script (Appendix 8.1: `vina_screen_local.sh`). Large volumes of information generated were reduced through compilation of scripts to parse results into easily visualised formats. Quantification of relative binding affinities was used to extract initial results packages of the top 1000 docking results (Appendix 8.1: `zip_results.sh`, `vina_screen_get_top_quant.py`).

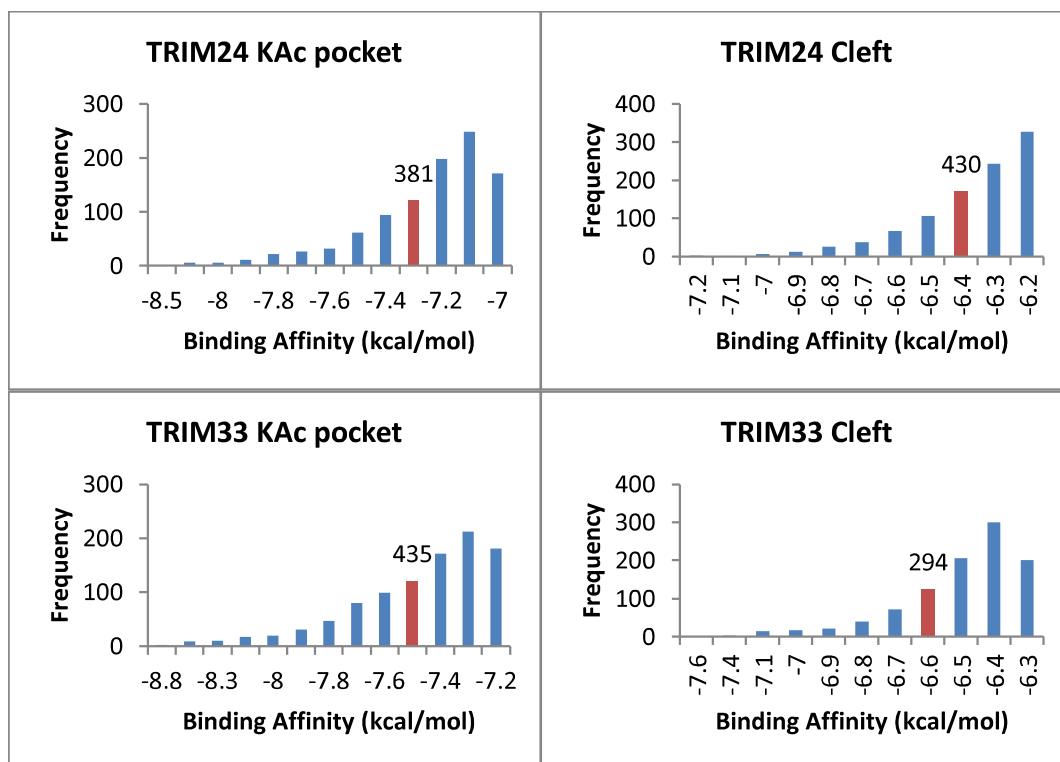
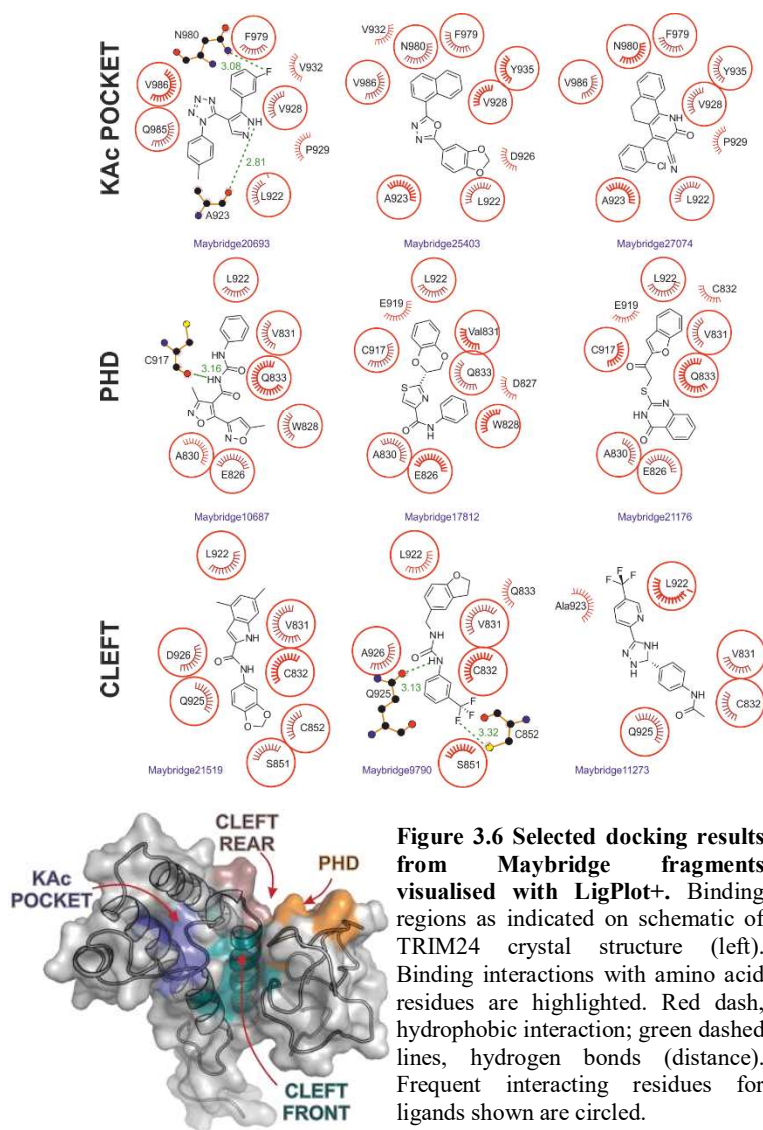


Figure 3.5 Histograms of binding affinities as result of docking of Maybridge Fragment Collection onto TRIM24 or TRIM33 using AutoDock Vina. Bar highlighted in red indicates the level chosen as an energetic cut-off for further analysis, labelled with the number of molecules satisfying this criterion.

Energetic cut-offs were enforced on these results, to reduce the number of candidates in the manual visualisation and selection process. Histograms highlight the spread of affinities, with low frequencies of compounds at high predicted affinity, and were used to define cut-offs before further compound selection (Figure 3.5). Reduction criteria and resultant numbers of molecules were: TRIM24 KAc, -7.3 kcal/mol, 381 molecules; TRIM24 cleft, -6.4 kcal/mol, 430 molecules; TRIM33 KAc, -7.5 kcal/mol, 435 molecules; TRIM33 cleft, -6.6 kcal/mol, 294 molecules. Results with binding affinities in the lower ranges of the distribution – above the selected cut-offs – are more likely to contain artefacts. Hence, exclusion of these would provide a more tractable selection of molecules for manual curation, with a relatively low probability of excluding ligands for TRIM24 or TRIM33.

Visual analysis of binding modes *via* PyMOL (Appendix 8.1: `make_pml.sh`) was complemented by two dimensional comparisons of interacting residues using LigPlot⁺¹⁴ (Figure 3.6), and chemical profiles in structure-data files (Appendix 8.1: `make_sdf.sh`). The combination of these representations formed a handle for grouping of results by various specifications, including binding



region and Tanimoto similarity coefficients. Trends in binding orientations were analysed to generate a pharmacophore.

Despite defining a large search area (20 Å) with loose constraints on the position of binding, results clustered in several key loci: the KAc binding pocket, a shallow cavity beside the PHD finger containing a conserved water molecule, and both sides of the groove formed by the domain interfaces referred to as the cleft (Figure 3.6). The extended conformations of these areas favoured elongated polycyclic scaffolds with polar linkages including amides

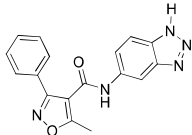
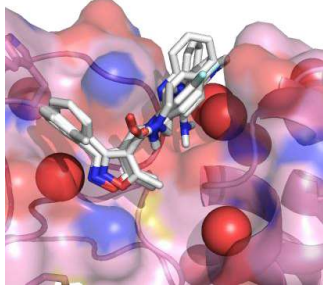
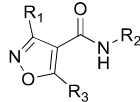
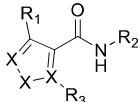
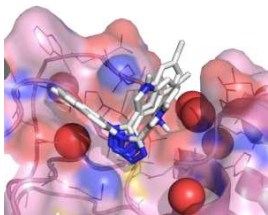
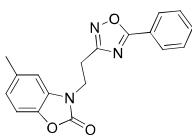
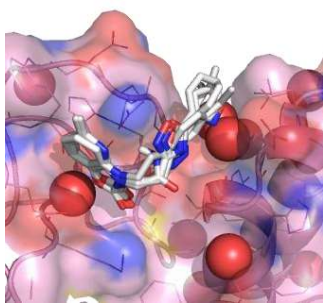
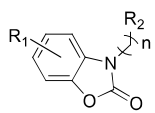
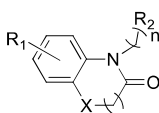
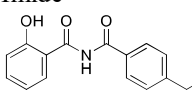
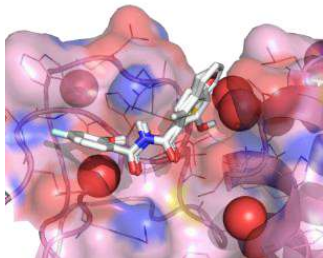
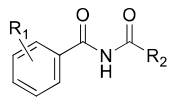
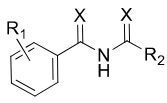
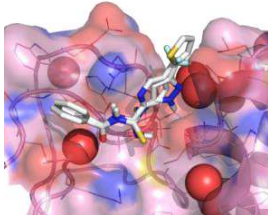
(e.g. Figure 3.6: Maybridge10687, 21519, 9790) and 5-membered heteroaromatics (e.g. Figure 3.6: Maybridge20693, 25403, 17812). Common protein-ligand interactions highlighted potentially significant residues: hydrophobic or polar interactions through the backbone carbonyl of L992 that is found across the entrance to the KAc pocket were common; several cysteines, other than those involved in Zn^{2+} ligation, formed hydrophobic interactions with ligand docking modes and this close proximity could be utilised in covalent interactions with ligands, or future labelling studies; the highly conserved asparagine (N980) within the KAc pocket preferentially formed hydrogen bonds with the ligands as expected.

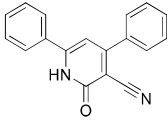
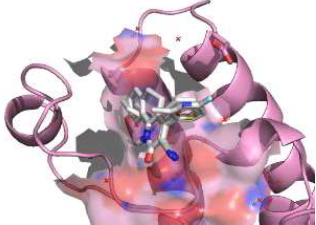
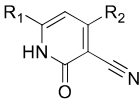
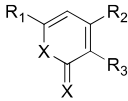
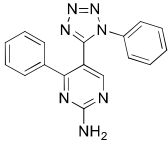
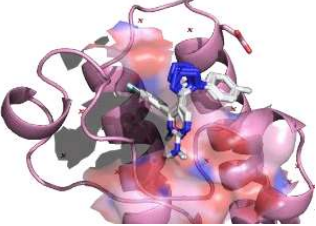
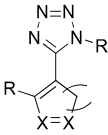
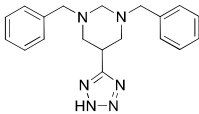
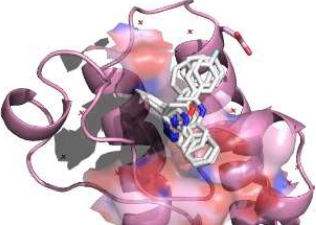
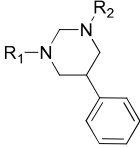
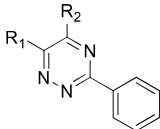
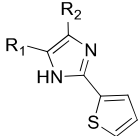
3.2.3 Compound selection

Cross-comparison of compounds identified in screening, across the two sites of both proteins, highlighted many ligands which were consistently identified as interacting with high binding affinities. Analysis of similarities in binding mode suggested that some of these hits may be artefacts, driven by overall hydrophobicity of a molecule, since high affinity interactions occurred in multiple, non-specific regions. However, other motifs were positively identified across several screens as partaking in specific interactions that are likely to generate affinity. The collection of all identified compounds was combined (1540), reduced to remove duplicates (1312), and manually reduced to a collection of the top 400 putative hits based on predicted affinity, assessed reliability of binding mode, and synthetic tractability (CD appendix: `Docking_results\top400maybridge.sdf`).

A script generated in PyMOL to load all lowest energy conformations for these 400 compounds, enabled visual analysis of docking poses, which highlighted several common binding modes (CD appendix: `Docking_results\top400maybridge.pse`). The 400 putative hits were reduced to six lead molecules, which represent groups of similar hit molecules (Table 3.1). *In silico* substitution of side groups, and resubmission to docking procedures gave some insight into functional group requirements, and pharmacophores were developed for structure activity relationship (SAR) studies. These pharmacophores included hypotheses on side chain functional groups, and possible scaffold variations depending on the observed docking.

Table 3.1 Summary of lead molecules identified from *in silico* screening. Common docking poses observed from results of AutoDock Vina with Maybridge screening collection: multiple molecules adopting similar conformations suggests a more likely binding mode. Primary substitution shows variations which could be taken forward for SAR. Core substitutions may more significantly change interactions with protein, but allow a similar binding mode.

Maybridge lead	Observed docking	Primary substitution	Core substitution
Isoxazole amide  MaybridgeF4079	Rear cleft pocket. Variants of side chains dock in similar conformation Pattern of hydrophobic/hydrophilic residues bind along cleft. 	 R₁ = Ph, Arhet, Bn, small alkyl (e.g. <i>c</i>-Pr) R₂ = widest variation, larger groups tolerated (preferred for peptide displacement)	 5-het cores maintaining R₁ and R₂ with optional R₃ e.g. tetrazole e.g. MaybridgeF24650 
Benzoxazalone  MaybridgeF24105	Polar pocket formed by E826, Q833, K846 could be targeted by carbamate of benoxazalone. 	 R₁ = <i>o</i>-Me, <i>p</i>-X (optional) R₂ = Arhet with n=2, extended	 Benzoxazole bioisosteres e.g. MaybridgeF10849
Imide  MaybridgeF25620	Imide hits dock in homologous pose to benoxazalone 	 R₁ = <i>o</i>-Me, <i>p</i>-X (optional) R₂ = widest variation (Ar; <i>o</i>-Me/Et/Pr) larger groups tolerated (preferred for peptide displacement)	 Polar unsaturated motifs maintaining nitrogen proton donor e.g. X=N MaybridgeF10951, X=O MaybridgeF16299 

Maybridge lead	Observed docking	Primary substitution	Core substitution
Oxo-dihydropyridine  MaybridgeF22434	1,3-substituted heteroaromatic compounds dock into entrance to KAc pocket  Common motif of oxo-dihydropyridine presents R₁ towards pocket, R₂ towards external E985, and charge complementarity to entrance ridge	 R₁ = Ar(het) direct, linked or fused R₂ = Arhet projected towards glutamate	1,3-disubstituted heteroaromatic cores with optional R₃  R₁ = benzylic groups to probe depth of KAc pocket. Hydrogen bonding to displace bound water.
Linked heteroaromatics  MaybridgeF4709 20 highly ranked compounds contained a 1,5-substituted tetrazole	Adjacent heteroaromatic scaffold creates linkage between KAc pocket and E985  Polar interactions with entrance to pocket: backbone carbonyl H-bonding to P927	 Arhet = pyrazole, pyrimidine, indole, isoxazole. R = Ph, <i>o</i>/<i>m</i>-X, <i>o</i>-Me	
Trisubstituted rings  MaybridgeF15226	2,4,5-substitution on 5/6 membered heteroaromatic rings to target the KAc, E985 and polar interactions with entrance (usually involved in 'RK' step) 	 Ar(het): potential for solubilising group polar region of ZA channel R₁ = Ph, Bn, <i>o</i>-Me/X, isoxazole (towards KAc pocket) R₂ = Ph, Bn: <i>o</i>-Me/X (towards glutamate)	Alternative polar cores, capable of hydrogen bonding to backbone carbonyl 1,2-substitution e.g. MaybridgeF7759, MaybridgeF28955  5-membered e.g. MaybridgeF27477 

3.3 Screening cascade

AlphaScreen technology has been used to identify and optimise chemical probes in many high-throughput screening (HTS) platforms,¹⁵ including those for both bromodomains and PHD fingers.^{16,17} Through use of liquid handling systems, the assay protocol developed for peptide IC₅₀ determination was transferred into a high-throughput hit detection format. This was used to screen compounds in duplicate at two concentrations, to detect any concentration-dependent inhibition of cognate peptide binding.

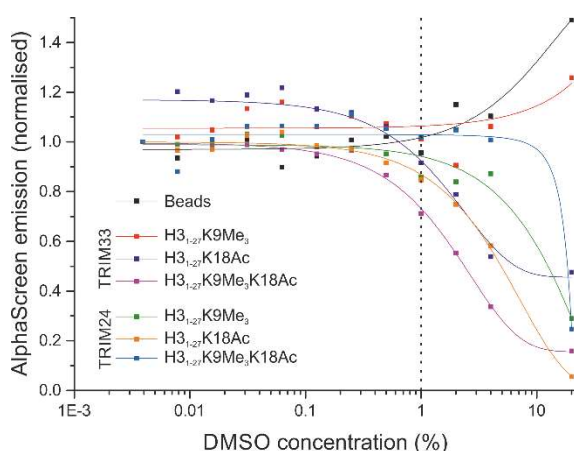


Figure 3.7 DMSO sensitivity of the AlphaScreen assay evaluated through variation in signal with percentage DMSO. Signals were normalised to the maximum emission from each competitor peptide. Emission remained at high levels up to 1% DMSO (dashed line) before a sharp drop when increasing to 20% DMSO. Legend: protein-peptide pair generating signal.

It has been recorded that DMSO, commonly used to solubilise compounds for aqueous assay conditions, can itself bind to the KAc pocket and inhibit peptide binding to bromodomains.¹⁸ Indeed, bromodomain screening in the AlphaScreen format uses high concentrations of DMSO to standardise assay results: 20% final assay concentration DMSO is used as a reference for ‘100%’ inhibition. Therefore, in order to reliably screen compounds, the DMSO

sensitivity of the assay was profiled (Figure 3.7). Together with compound solubility, this would dictate the upper limit of compound concentration possible for this assay. DMSO concentrations of up to 1% were tolerated with little effect on signal, and 20% DMSO generated an average reduction in signal of 60%. Addition of DMSO to wells containing beads in the absence of protein, to determine any bead effects, showed only a slight increase in signal for high DMSO concentrations, but this corresponded to a negligible level of ~200 a.u. compared with ~50,000 a.u. for high signal. Maximum information on potential binders to these tandem domains would be obtained if the localisation of binding, relative to the PHD finger and bromodomain, could also be inferred from screening results. Using the control compound provided by the SGC (SGC-TIF1 α 53: structure

undisclosed), selectivity between TRIM24 and TRIM33, as well as between domains within each protein, was compared (Figure 3.8). There was some selectivity between the two homologous proteins, with binding to TRIM33 >10-fold weaker (green: curve fitting was at the limit of the concentration curve) than to TRIM24.

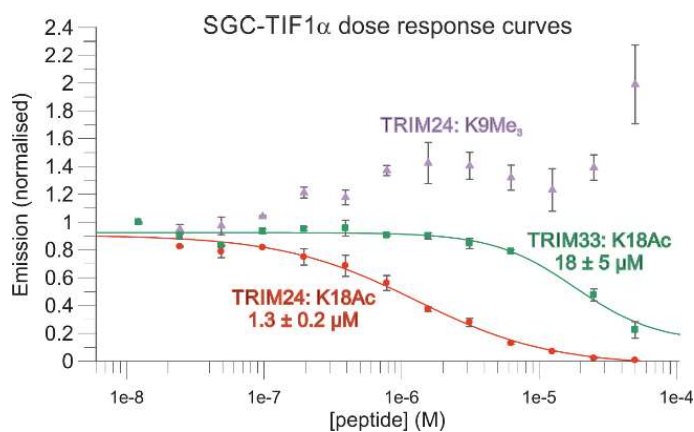


Figure 3.8 Dose response curves for SGC-TIF1 α control compound obtained by AlphaScreen. Compound demonstrated >10-fold selectivity for TRIM24 (red), over TRIM33 (green). The bromodomain binding compound was able to displace a K18Ac competitor peptide (red) but not a K9Me₃ competitor peptide (lilac).

Most notably, the control compound, which was in development as a bromodomain inhibitor, was not able to displace a peptide bearing exclusively the K9Me₃ mark (lilac). This result demonstrates that it is possible to distinguish between binding to the bromodomain and the PHD finger using this assay.

Use of a competitor peptide bearing either the K9Me₃ or KAc modification, would enable orthogonal screening of the bromodomain and PHD finger. To ensure all potential hits were discovered, including those which may bind in the interdomain region and prevent a peptide from binding across both domains, the doubly modified competitor peptide (K9Me₃ and KAc) was also used in screening. For consistency between proteins, the K18Ac position will be incorporated, since this was a requirement for generation of adequate signal from TRIM33, and produced high levels of signal in TRIM24 as well.

In this manner, screens were conducted against both TRIM24 and TRIM33 using peptides that can be displaced from the PHD finger, bromodomain, or spanning both domains: H3₁₋₂₇K9Me₃ (PHD); H3₁₋₂₇K18Ac (bromo); and H3₁₋₂₇K9Me₃K18Ac (dual). To our knowledge, this is the first example of use of a complementary set of competitor peptides in a displacement assay to probe a tandem domain, and gain information about binding mechanisms (Table 3.2). Reference wells containing buffer or 20% DMSO (maximum or minimum signal, respectively) were used to determine the sample deviations and the Z' value for each assay plate, as a measure of assay quality.

Assay component	Final assay concentration					
AlphaScreen beads (μg/mL)	5	5	5	7	7	7
TRIM24 (nM)	50	10	10	-	-	-
TRIM33 (nM)	-	-	-	500	250	50
PHD: H3 ₁₋₂₇ K9Me ₃ (nM)	15	-	-	15	-	-
bromo: H3 ₁₋₂₇ K18Ac (nM)	-	8	-	-	30	-
dual: H3 ₁₋₂₇ K9Me ₃ K18Ac (nM)	-	-	4	-	-	15

From the results of the *in silico* screening described previously, the final selection of Maybridge compounds were tested *in vitro* for affinity towards TRIM24 or TRIM33. Based on availability of the 400 fragments selected, 201 compounds were ordered as solubilised stock (30 mM in DMSO) on LabCyte 384-LDV source plates which were compatible with an Echo acoustic liquid handler (Labcyte) to dispense volumes of 2.5-500 nL. Samples of 20 or 100 nL were transferred to assay plates in duplicate to test at 30 or 150 μ M FAC. All samples were backfilled with DMSO to a volume of 100 nL to maintain 0.1% final assay content across compound concentrations.

AlphaScreen 384-well ProxiPlate plus format																						volume transferred		final assay concentration	
																						100 nL	50 µM		
																						20 nL	10 µM		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
A	Buffer	c1	c1	c2	c2	c3	c3	c4	c4	c5	c5	c6	c6	c7	c7	c8	c8	c9	c9	c10	c10	High control			
B		c1	c1	c2	c2	c3	c3	c4	c4	c5	c5	c6	c6	c7	c7	c8	c8	c9	c9	c10	c10				
C		c11	c11	c12	c12	c13	c13	c14	c14	c15	c15	c16	c16	c17	c17	c18	c18	c19	c19	c20	c20				
D		c11	c11	c12	c12	c13	c13	c14	c14	c15	c15	c16	c16	c17	c17	c18	c18	c19	c19	c20	c20				
E		c21	c21	c22	c22	c23	c23	c24	c24	c25	c25	c26	c26	c27	c27	c28	c28	c29	c29	c30	c30				
F	Mid control	c21	c21	c22	c22	c23	c23	c24	c24	c25	c25	c26	c26	c27	c27	c28	c28	c29	c29	c30	c30	Mid control			
G		c31	c31	c32	c32	c33	c33	c34	c34	c35	c35	c36	c36	c37	c37	c38	c38	c39	c39	c40	c40				
H		c31	c31	c32	c32	c33	c33	c34	c34	c35	c35	c36	c36	c37	c37	c38	c38	c39	c39	c40	c40				
I		c41	c41	c42	c42	c43	c43	c44	c44	c45	c45	c46	c46	c47	c47	c48	c48	c49	c49	c50	c50				
J		c41	c41	c42	c42	c43	c43	c44	c44	c45	c45	c46	c46	c47	c47	c48	c48	c49	c49	c50	c50				
K	High control	c51	c51	c52	c52	c53	c53	c54	c54	c55	c55	c56	c56	c57	c57	c58	c58	c59	c59	c60	c60	Buffer			
L		c51	c51	c52	c52	c53	c53	c54	c54	c55	c55	c56	c56	c57	c57	c58	c58	c59	c59	c60	c60				
M		c61	c61	c62	c62	c63	c63	c64	c64	c65	c65	c66	c66	c67	c67	c68	c68	c69	c69	c70	c70				
N		c61	c61	c62	c62	c63	c63	c64	c64	c65	c65	c66	c66	c67	c67	c68	c68	c69	c69	c70	c70				
O		c71	c71	c72	c72	c73	c73	c74	c74	c75	c75	c76	c76	c77	c77	c78	c78	c79	c79	c80	c80				
P		c71	c71	c72	c72	c73	c73	c74	c74	c75	c75	c76	c76	c77	c77	c78	c78	c79	c79	c80	c80				

Compounds were dry loaded onto ProxiPlate assay plates *via* inversion (Figure 3.9). Protocols were adapted to this lower volume of sample (100 nL c.f. 5 μ L for peptide samples) by a two-step addition of protein-peptide mix, followed by AlphaScreen beads. Screens were conducted against

both TRIM24 and TRIM33 using peptides which can be displaced from the PHD finger, bromodomain, or spanning both domains (Table 3.2).

Results showed good differentiation between compounds, proteins, and regions. Several compounds stood out as having high levels of inhibition, with many of these demonstrating apparent selectivity. 24 compounds resulted in inhibition of >50% in any of the 6 assays. Inhibition was >50% at 150 μ M with all three competitor peptides for eight compounds against TRIM24 and eight compounds against TRIM33. Compounds c007, c089, and c152 resulted in higher inhibition of TRIM33 than TRIM24, with an average of 3.5, 10.7, and 4.1 fold-increase, respectively, at 150 μ M. Conversely compounds c201 seemed to show a preferential inhibition of TRIM24, with >50% inhibition at 150 μ M.

Several compounds seemed to show considerable increase in AlphaScreen emission: 3 compounds resulted in >50% increase in signal normalised to buffer controls. Increased emission was often observed across several competitor peptides, and was considered to be due to compound emission in the AlphaScreen detection range of 520-620 nm, rather than ligand interactions with proteins.

Table 3.3 Inhibition of Maybridge Screening compounds in AlphaScreen assays. Percentage inhibition values for 201 compounds from Maybridge Screening collection, reported relative to buffer blank wells on the same assay plate, as an average of duplicates. Compounds were coded against a blinded screen based on their position on source plates (c001-c235), and screened against TRIM24 and TRIM33, with three different competitor peptides designed to highlight selectivity for each domain: PHD = H3₁₋₂₇K9Me₃; bromo = H3₁₋₂₇K18Ac; dual = H3₁₋₂₇K9Me₃K18Ac. n.d. = not disclosed. Shading on cells represents inhibition levels from low (green) to high (red), inhibition of >50% is highlighted in bold.

Plate	Molecular formulae	Compound ID	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	30	150	30	150	30	150	30	150	30
1	C ₁₈ H ₁₃ F ₂ NO	c001	26.7	8.1	-37.4	-17.5	-47.4	-6.7	-1.1	17.1	-35.9	-3.3	-40.9	-9.1
1	C ₁₇ H ₁₅ ClN ₂ O ₂ S	c002	8.8	25.5	-27.0	-13.9	-17.4	-1.4	11.9	14.5	-0.5	4.0	-13.8	-8.6
1	C ₁₆ H ₁₀ N ₄ OS	c003	-9.3	17.2	-23.0	-14.5	-19.7	-4.5	15.8	15.4	-5.0	4.5	-2.6	-3.8
1	C ₁₉ H ₂₀ N ₄ O	c004	28.8	28.3	3.7	0.3	8.1	9.5	22.6	18.4	21.5	17.4	14.1	13.8
1	C ₁₄ H ₁₁ F ₃ N ₄ O	c005	-6.4	19.0	-36.1	-18.0	-26.8	-3.8	-1.2	12.5	-0.3	2.1	-10.1	-5.2
1	C ₁₉ H ₂₁ N ₅ O_dup1	c006	1.2	23.1	-29.5	-17.5	-19.8	12.0	4.4	14.8	5.5	6.2	-0.5	2.7
1	C ₁₈ H ₁₄ ClN ₃ O	c007	20.9	33.6	-13.3	-11.5	10.4	10.3	46.7	25.7	44.6	13.1	53.0	14.5
1	C ₂₀ H ₁₇ NO ₂ S	c008	-4.2	25.2	-31.7	-12.6	-25.7	-1.3	2.0	14.2	-5.6	3.6	-15.1	1.6
1	C ₁₄ H ₉ F ₂ N ₃ O ₃ S	c009	-10.1	19.4	-38.2	-23.6	-31.0	-5.1	-17.6	4.5	0.2	0.7	-8.9	-2.9
1	C ₁₅ H ₁₄ N ₂ O ₂ S ₂	c010	7.7	30.2	-27.0	-9.6	-18.1	3.1	-1.1	9.6	0.8	2.8	-4.9	4.6
1	C ₂₁ H ₂₉ N ₃	c011	-5.0	11.3	-12.0	-14.1	-9.4	2.5	13.3	28.3	-15.7	0.7	-18.0	-2.7
1	C ₁₈ H ₁₁ F ₂ N ₃ O	c012	23.9	20.6	-7.8	-13.0	7.6	4.6	26.5	21.2	12.7	6.6	4.4	-1.6
1	C ₁₉ H ₂₁ N ₅ O	c013	18.4	18.7	-4.4	-12.0	8.8	16.5	20.0	17.0	14.4	7.2	5.8	2.4
1	C ₁₉ H ₂₁ N ₃ O ₃	c014	23.4	21.1	-1.7	-10.7	15.4	4.4	9.6	19.8	9.3	12.0	7.6	7.9
1	C ₂₁ H ₁₅ N	c015	-17.8	17.6	-43.4	-23.6	-24.4	-7.6	-17.2	13.7	2.7	3.4	-4.5	-7.8
1	C ₁₉ H ₂₂ N ₆	c016	2.4	18.6	-36.2	-20.8	-16.1	5.4	-5.0	16.7	-1.9	4.2	-15.2	2.0
1	C ₁₅ H ₈ F ₆ N ₂ O	c017	26.0	32.6	-0.2	-10.5	17.2	13.6	32.7	22.4	25.8	13.5	22.5	16.7
1	C ₂₀ H ₁₇ N ₃ O ₃	c018	36.5	38.8	13.9	-6.6	27.2	10.0	28.3	22.4	42.2	15.6	39.2	13.7
1	C ₁₅ H ₁₀ F ₆ N ₂ O	c019	11.0	22.0	-13.6	-18.5	-7.7	-5.2	14.5	17.0	5.5	2.6	9.8	-0.9
1	C ₂₂ H ₂₅ N ₃ O_dup1	c020	48.6	37.5	36.4	11.1	37.2	21.3	26.7	19.1	34.4	15.8	44.4	19.8
1	C ₁₇ H ₁₅ N ₃ O ₂	c021	41.3	26.8	23.1	-6.6	31.9	8.4	33.1	21.4	42.1	13.0	32.6	6.8
1	C ₁₃ H ₉ FN ₄ O	c022	39.1	31.4	14.2	-9.3	28.8	13.2	20.6	15.6	32.2	11.1	26.8	7.5
1	n.d.	c023	-0.5	16.3	-28.7	-23.6	-18.6	-9.2	4.4	11.0	5.0	0.5	-4.7	-5.9

Plate	Molecular formulae	Compound ID	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	150	30	150	150	30	150	150	30	150
1	C ₈ H ₈ N ₄ O	c024	-9.7	5.1	-43.1	-27.9	-17.0	-17.0	-17.0	0.1	-6.3	-13.2	7.3	-15.6
1	C ₁₇ H ₁₃ ClFN ₃ O ₂	c025	23.1	29.1	-4.8	-9.1	12.9	5.1	22.7	19.6	12.3	11.3	1.8	3.5
1	C ₁₇ H ₁₂ Cl ₂ N ₂ O ₂ _dup ₁	c026	34.9	34.3	-4.5	-13.7	16.1	10.3	-30.5	4.0	19.8	1.0	20.6	12.0
1	C ₁₆ H ₁₃ ClN ₄ OS	c027	27.4	25.3	-4.4	-12.5	14.9	12.1	22.4	19.0	23.5	19.8	16.7	3.5
1	C ₁₄ H ₇ BrClF ₃ N ₂ O ₂	c028	13.3	32.1	-21.0	-27.0	-7.9	-10.7	12.3	13.4	8.7	5.9	5.1	5.0
1	C ₁₆ H ₁₁ ClN ₆	c029	24.9	33.1	-3.4	-8.8	8.1	12.2	25.0	30.1	19.7	23.8	14.8	11.6
1	C ₁₅ H ₉ N ₃ O ₂ S	c030	22.8	23.2	3.6	-4.5	7.4	12.0	25.0	19.9	16.1	11.2	13.2	11.0
1	C ₁₇ H ₁₂ ClN ₃ O ₃	c031	23.4	18.5	3.7	-10.9	15.4	3.3	21.0	18.7	36.8	11.4	26.1	5.5
1	C ₂₃ H ₂₄ N ₂ O	c032	24.3	23.6	7.0	-5.9	16.1	12.0	29.8	24.7	22.6	11.7	15.2	5.8
1	C ₁₈ H ₁₃ NO ₄ S_dup ₁	c033	58.4	31.0	51.5	27.9	52.9	19.8	64.7	39.1	58.1	31.6	49.5	27.6
1	C ₁₉ H ₁₄ Na ₄ O ₃	c034	15.4	21.8	-5.4	-10.0	15.1	14.5	22.4	21.5	2.1	7.6	-1.4	4.9
1	C ₁₇ H ₁₃ N ₅ O ₂	c035	24.3	32.8	-4.1	-9.7	13.6	6.9	27.6	27.5	20.9	18.3	14.7	8.4
1	C ₁₉ H ₁₆ Na ₄ S	c036	31.8	29.2	-4.9	-11.7	8.9	3.8	27.8	23.0	25.9	15.4	23.7	16.1
1	C ₁₆ H ₁₂ N ₄ O ₂ S	c037	11.0	23.0	-14.7	-13.3	10.0	18.5	24.6	22.0	20.9	17.6	18.4	17.1
1	C ₁₇ H ₁₂ Cl ₂ N ₂ O ₂	c038	32.3	35.3	1.7	-8.3	20.4	12.3	26.8	28.3	34.5	20.0	34.2	15.9
1	n.d.	c039	15.3	16.2	-25.9	-29.4	2.0	1.3	6.9	8.4	8.0	4.3	8.1	5.6
1	C ₁₁ H ₅ F ₃ N ₄ O ₂	c040	18.4	21.4	-1.2	-6.4	8.6	11.8	46.0	22.3	15.7	10.5	14.2	22.2
1	C ₁₇ H ₁₃ ClN ₆	c041	12.5	15.1	-6.5	-12.5	8.7	1.1	20.4	16.2	15.8	10.8	7.5	2.8
1	C ₁₇ H ₁₄ N ₂ O ₃ S	c042	8.8	11.6	-15.4	-19.3	6.9	5.7	16.4	45.3	10.0	-0.9	-0.4	-5.2
1	C ₁₆ H ₁₄ Na ₄ O ₄ _dup ₃	c043	11.3	14.5	-10.3	-14.2	3.3	1.5	20.3	16.0	13.4	10.0	2.4	3.2
1	C ₁₉ H ₂₄ Na ₄ O ₂	c044	16.4	19.6	-5.4	-13.5	12.7	7.8	24.9	23.6	20.7	11.5	9.8	5.4
1	C ₁₅ H ₁₀ ClNO ₃	c045	26.9	23.1	1.0	-8.6	22.8	8.3	27.2	20.6	33.5	15.2	25.9	9.1
1	C ₁₈ H ₁₆ N ₆	c046	23.8	18.4	-2.2	-9.4	7.6	4.8	25.0	22.4	19.4	19.2	21.9	46.6
1	C ₁₇ H ₁₁ N ₅	c047	-5.3	8.2	0.8	-28.8	15.4	6.6	45.2	17.6	24.4	3.9	23.7	1.8
1	C ₁₁ H ₁₂ Na ₄ O ₃ S	c048	-0.8	27.3	-30.6	-20.6	1.8	1.6	3.5	21.5	46.2	11.1	1.2	11.9
1	C ₁₆ H ₁₄ ClNO ₂	c049	53.8	45.1	37.6	23.4	45.5	37.5	53.6	48.5	54.9	43.3	55.4	42.5

Plate	Molecular formulae	Compound ID Concentration (µM)	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	150	30	150	150	30	150	150	30	150
1	C ₁₇ H ₁₆ ClN ₃ O	c050	21.4	23.2	4.0	-1.1	13.8	27.5	23.0	26.9	18.7	25.9	17.2	
1	C ₁₅ H ₁₀ ClN ₃ O ₂	c051	32.4	18.2	21.4	-3.8	5.5	21.0	22.6	21.0	9.4	26.3	9.7	
1	C ₁₂ H ₉ ClFN ₃	c052	21.2	17.6	-3.9	-9.7	9.8	22.0	16.9	22.2	14.4	15.6	11.4	
1	C ₁₈ H ₁₅ N ₇ _dup ₁	c053	20.8	15.6	-2.1	-10.6	8.7	25.1	18.7	22.7	13.7	15.9	10.8	
1	C ₁₅ H ₁₀ N ₄ OS	c054	26.3	19.4	7.2	-7.6	15.3	-7.8	2.7	15.7	6.5	19.7	7.4	
1	C ₂₀ H ₂₀ N ₄ O	c055	23.0	30.0	4.3	-2.6	13.5	37.9	33.7	28.6	20.5	20.6	14.3	
1	C ₂₁ H ₁₇ N ₃ O ₂ _dup ₁	c056	29.6	29.9	-0.5	-11.2	5.0	33.4	24.5	27.0	21.0	31.1	19.6	
1	C ₂₁ H ₁₃ N ₃ O	c057	10.8	19.0	0.9	-12.7	7.4	18.1	25.2	43.9	28.0	28.3	18.6	
1	C ₁₄ H ₈ F ₄ N ₄	c058	35.0	36.8	18.3	12.1	33.4	42.2	41.3	37.0	33.7	31.8	35.5	
1	C ₁₆ H ₁₁ ClFN ₃ OS	c059	-0.8	17.3	-22.0	-21.9	6.3	9.4	15.2	6.4	13.0	11.0	9.0	
1	C ₁₇ H ₁₈ N ₂ O ₃	c060	24.9	27.7	3.7	-0.6	10.9	27.2	26.9	24.4	20.1	25.1	22.3	
1	C ₁₇ H ₁₂ N ₄ O ₂	c061	16.0	13.5	-4.7	-11.3	0.1	21.1	21.4	19.7	10.1	14.6	4.1	
1	C ₁₈ H ₁₃ N ₃ O ₂	c062	17.3	12.1	-2.4	-9.9	3.4	17.1	17.8	19.1	11.7	16.7	8.8	
1	C ₂₅ H ₂₀ N ₂	c063	-11.9	-9.9	-94.0	-94.7	-62.4	-23.2	-45.4	-18.5	-24.7	-43.5	-60.0	
1	C ₁₈ H ₁₇ FN ₂ O ₂ S	c064	29.8	16.4	2.9	-9.6	12.6	31.1	22.4	28.2	16.6	30.2	18.2	
1	C ₁₂ H ₁₁ NO ₃	c065	22.2	25.3	-2.5	-9.9	3.3	30.9	26.7	21.3	15.2	23.2	17.9	
1	C ₁₃ H ₉ ClFN ₂ O ₂	c066	29.6	18.5	11.1	-5.5	4.2	33.7	25.1	32.8	21.0	33.3	22.7	
1	C ₁₅ H ₁₄ N ₂ O	c067	23.1	21.8	3.6	-9.5	14.7	41.7	26.3	32.1	20.0	30.4	24.2	
1	C ₁₃ H ₈ O ₄ S	c068	18.6	21.8	3.1	-7.1	13.9	31.2	31.1	25.6	19.2	24.5	22.1	
1	C ₁₆ H ₁₀ N ₂ O ₂	c069	21.3	22.9	8.7	-6.9	15.6	32.1	28.5	23.3	21.5	32.6	21.0	
1	C ₁₇ H ₁₅ N ₃ S	c070	23.6	22.8	5.7	0.5	22.1	35.0	29.7	29.6	21.9	26.2	20.5	
1	C ₁₈ H ₁₅ N ₃ O ₃	c071	14.7	10.6	-1.5	-12.2	0.0	26.6	19.5	20.1	11.1	15.7	8.1	
1	C ₁₇ H ₁₃ N ₃ O ₂ S	c072	86.8	52.0	77.6	25.8	44.9	73.6	31.9	98.7	81.3	98.6	79.1	
1	C ₁₈ H ₁₂ N ₂ O ₂ _dup ₁	c073	36.4	20.6	20.0	-3.6	12.6	50.9	31.3	49.9	26.7	39.8	22.5	
1	C ₂₀ H ₁₉ BrFN ₃ O	c074	53.6	39.0	45.0	22.0	36.0	62.7	48.8	58.7	45.6	56.7	42.5	
1	C ₁₅ H ₁₆ N ₆ O ₂ S	c075	-11.7	3.2	-31.2	-17.3	2.1	-1.2	20.3	-2.3	13.2	0.7	13.0	

Plate	Molecular formulae	Compound ID	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	150	30	150	150	30	150	150	30	150
1	C ₁₅ H ₁₃ NOS	c076	24.0	14.0	6.1	-9.6	2.9	34.7	27.2	36.3	24.0	45.8	33.5	33.5
1	C ₁₄ H ₁₀ F ₃ N ₃ OS	c077	20.3	10.4	-0.1	-15.0	8.3	32.3	24.8	29.1	20.1	29.5	26.3	26.3
1	C ₂₁ H ₁₅ NO	c078	25.3	19.2	-0.5	-9.6	11.1	34.6	32.1	30.6	21.4	28.7	26.0	26.0
1	n.d.	c079	19.4	24.1	0.7	-10.7	10.3	28.2	24.4	25.6	21.4	32.1	22.1	22.1
1	C ₁₇ H ₁₁ F ₂ N ₅ O	c080	-2.9	-10.0	-24.4	-34.7	-7.9	15.3	6.6	7.4	2.9	13.7	9.2	9.2
2	C ₁₃ H ₁₃ N ₃ O ₄	c081	-12.2	8.5	-11.5	-2.1	1.5	-21.5	-10.0	-9.0	-11.4	-8.6	-2.0	-2.0
2	C ₁₆ H ₁₂ Cl ₂ N ₄ O	c082	-19.8	9.9	-13.6	1.3	4.0	-18.5	-6.3	-28.2	-12.8	-13.7	-2.2	-2.2
2	C ₁₁ H ₉ N ₅ O	c083	-13.6	15.0	-11.8	3.6	10.5	-24.4	-15.5	-17.9	-11.7	-14.7	-0.6	-0.6
2	C ₁₄ H ₁₀ F ₂ N ₂ O ₃ S	c084	-52.8	-17.1	-61.3	-31.2	-23.3	-73.9	-38.4	-69.0	-42.6	-66.1	-34.5	-34.5
2	C ₁₇ H ₁₂ N ₂ O	c085	-10.2	13.4	-14.4	4.3	11.8	-22.5	-1.5	-27.8	-5.5	-18.9	0.7	0.7
2	C ₁₆ H ₉ ClFN ₅ O	c086	-30.2	17.7	-31.1	0.2	4.9	-42.2	-4.8	-41.3	-8.8	-36.0	-1.2	-1.2
2	C ₂₀ H ₂₁ NOS	c087	-2.0	17.4	-14.1	3.6	10.9	-4.3	-2.5	-9.3	-6.1	6.5	2.3	2.3
2	C ₁₇ H ₁₂ ClNO ₄	c088	-5.2	18.7	-7.6	6.0	12.0	-6.8	-5.0	-10.5	-2.0	-2.2	2.9	2.9
2	C ₁₉ H ₁₅ NO ₄	c089	3.4	21.4	-5.0	5.5	10.0	51.8	25.1	58.0	16.8	45.3	10.4	10.4
2	C ₁₃ H ₈ F ₃ N ₃ O ₂	c090	-8.2	20.4	-11.3	2.4	4.7	-25.9	-7.6	-13.1	-7.1	-8.0	1.4	1.4
2	C ₁₄ H ₇ ClF ₂ N ₂ O	c091	8.7	15.3	9.0	5.2	8.4	-5.6	-4.3	-0.7	-7.5	3.9	2.9	2.9
2	C ₁₈ H ₁₃ FN ₂ OS	c092	12.9	15.1	11.2	8.1	12.5	0.2	7.8	0.3	-4.8	4.7	4.5	4.5
2	C ₁₆ H ₁₅ N ₅ S	c093	20.1	19.8	18.3	11.6	14.1	13.4	5.6	12.3	1.4	15.1	7.4	7.4
2	C ₁₃ H ₉ NO ₃ S	c094	27.0	23.3	11.8	7.3	18.3	3.0	5.5	2.8	-1.3	10.5	8.6	8.6
2	C ₂₅ H ₂₁ N	c095	-5.9	18.8	9.5	9.4	12.0	8.8	7.9	-3.2	-1.5	48.1	6.5	6.5
2	C ₁₇ H ₁₄ N ₂ O ₂	c096	19.2	25.6	12.4	9.6	16.8	-0.2	6.4	0.9	3.6	3.0	8.2	8.2
2	C ₁₉ H ₁₆ N ₂ O ₄	c097	24.2	28.3	12.0	12.0	33.6	6.1	1.2	4.7	7.1	13.4	10.0	10.0
2	C ₂₄ H ₁₈ FN	c098	33.4	30.2	14.8	10.1	22.1	15.7	4.1	25.1	7.1	35.2	15.3	15.3
2	C ₁₄ H ₉ N ₃ O ₅ S	c099	-10.9	25.7	7.8	10.6	17.5	9.7	10.2	-22.0	-2.6	-13.9	6.4	6.4
2	C ₁₂ H ₁₃ N ₃ O ₄ S	c100	58.9	39.5	47.2	15.9	24.9	41.4	15.1	51.9	11.6	51.0	19.1	19.1
2	C ₁₅ H ₁₁ N ₃ O ₂	c101	25.6	28.7	23.2	25.8	22.1	10.2	8.2	13.6	7.4	22.3	29.1	29.1

			Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction																	
Plate	Molecular formulae	Compound ID	TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD			TRIM33-bromo			TRIM33-dual		
			150	30	150	30	150	30	150	30	150	30	150	30	150	30	150	30		
2	C ₁₆ H ₁₁ N ₅ O	c102	22.7	21.3	19.1	8.2	28.1	15.3	6.8	1.0	30.6	7.3	19.8	8.1						
2	C ₁₉ H ₁₄ N ₂ S	c103	25.5	21.9	18.8	11.4	25.9	16.7	12.3	6.6	14.4	4.6	17.3	9.3						
2	C ₁₇ H ₁₂ F ₂ N ₆	c104	32.0	26.0	20.7	11.1	31.6	18.5	12.1	4.6	21.6	6.2	26.5	13.0						
2	C ₁₇ H ₁₄ N ₆	c105	44.7	38.8	16.2	18.7	23.9	26.8	12.0	11.9	11.6	22.1	13.3	25.6						
2	C ₁₉ H ₁₁ FN ₂ O ₃	c106	33.7	34.5	21.8	11.0	27.8	19.5	-22.3	3.1	49.3	17.7	40.7	18.6						
2	C ₂₁ H ₁₄ N ₂ O ₂	c107	29.5	31.5	15.3	12.3	27.4	19.9	3.9	4.1	12.6	12.8	17.9	16.4						
2	C ₁₅ H ₁₁ F ₃ N ₆ O	c108	28.0	25.9	12.8	10.6	22.0	22.2	9.1	6.6	10.5	7.8	21.1	12.1						
2	C ₁₉ H ₁₅ N ₃ O ₂	c109	39.4	38.1	14.7	13.7	26.9	20.0	10.5	9.1	9.7	9.7	13.1	14.0						
2	C ₁₄ H ₉ F ₃ N ₄ S	c110	45.8	46.5	34.4	29.8	30.6	31.9	24.4	10.7	25.3	17.1	18.0	15.5						
2	C ₁₆ H ₁₁ N ₃ O ₂	c111	76.2	26.7	78.3	9.0	88.8	14.7	74.7	10.0	93.5	24.2	96.6	35.4						
2	C ₁₄ H ₁₀ F ₃ NO	c112	17.4	18.0	14.4	8.1	20.4	18.2	3.6	5.5	6.5	3.4	12.5	10.1						
2	C ₁₆ H ₁₄ F ₂ N ₂ O ₃	c113	19.7	20.5	17.2	7.6	28.6	17.1	2.5	6.6	18.8	6.0	16.1	7.1						
2	C ₂₀ H ₁₈ N ₄ S	c114	22.2	21.9	11.1	7.6	22.1	16.1	9.9	5.8	6.3	2.3	13.2	12.1						
2	C ₂₀ H ₁₅ N ₃ O ₂	c115	-33.4	-1.6	-90.3	-20.5	-62.1	-11.7	-79.6	-29.0	-71.9	-27.4	-56.8	-18.3						
2	C ₁₂ H ₈ BrClFN ₃ O ₃	c116	19.3	20.8	14.2	6.8	18.6	15.5	4.5	8.9	6.7	4.0	9.7	9.1						
2	C ₁₆ H ₁₁ N ₃ O ₂ S	c117	27.5	29.7	17.0	9.9	36.1	26.4	13.2	17.2	20.6	22.4	15.6	18.1						
2	C ₁₅ H ₁₄ N ₂ O ₂ S	c118	31.6	25.7	17.8	12.1	27.7	22.7	17.2	6.4	20.6	9.7	27.7	15.1						
2	C ₁₄ H ₉ F ₃ N ₄ O	c119	32.7	42.2	12.0	11.9	27.9	24.8	19.7	7.4	11.4	9.2	16.8	13.7						
2	C ₁₇ H ₁₃ N ₅ S	c120	7.6	20.5	-8.0	3.0	0.7	18.3	-20.7	-2.1	-6.7	7.9	3.6	8.6						
2	C ₁₅ H ₁₀ F ₃ N ₃ OS	c121	14.6	19.7	9.9	3.0	12.1	10.2	3.1	7.4	4.6	0.7	12.8	8.9						
2	C ₁₇ H ₁₂ FN ₅ O	c122	-4.6	12.4	-8.6	0.9	4.9	6.8	-16.0	-0.5	-12.1	-1.4	-2.2	6.3						
2	C ₈ H ₈ N ₄ O_dup1	c123	32.2	23.6	27.8	13.5	34.0	20.4	19.5	7.9	59.6	19.6	52.9	20.1						
2	C ₁₇ H ₁₄ F ₂ N ₂ O ₂ S	c124	16.1	12.7	5.7	1.8	19.1	11.4	-2.5	2.7	6.5	-2.3	17.4	7.8						
2	C ₁₂ H ₁₀ N ₆ O	c125	37.1	24.5	24.5	11.8	46.2	21.2	40.1	14.7	49.8	14.8	52.6	21.6						
2	C ₁₈ H ₁₁ NO ₃	c126	18.3	20.6	10.4	8.5	19.2	17.1	5.5	9.8	12.2	9.7	16.5	12.8						
2	C ₉ H ₈ N ₄ O	c127	27.3	33.3	17.9	11.4	29.2	21.9	22.6	8.3	29.8	26.9	21.9	16.7						

Plate	Molecular formulae	Compound ID Concentration (µM)	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	150	30	150	150	30	150	150	30	150
2	C ₁₇ H ₁₆ N ₄ O	c128	25.8	18.9	5.1	5.0	5.1	32.1	19.8	16.5	8.0	25.6	8.9	40.1
2	C ₁₆ H ₁₄ N ₄ O ₄	c129	57.4	39.0	39.3	18.2	18.2	42.7	22.9	8.1	16.4	27.4	21.5	20.4
2	C ₁₆ H ₁₀ F ₃ NO ₂	c130	23.8	22.4	18.7	13.2	13.2	23.0	23.9	5.8	9.0	12.4	8.1	23.8
2	C ₁₇ H ₁₇ N ₃ O ₃	c131	12.6	17.1	12.3	3.9	3.9	12.5	13.2	2.5	0.0	4.1	2.8	12.3
2	C ₁₆ H ₁₅ ClN ₂ O ₂ S	c132	19.3	17.2	18.5	7.2	7.2	27.7	14.9	2.0	2.3	25.2	11.7	29.8
2	C ₁₉ H ₁₇ N ₅	c133	23.9	20.7	26.2	13.4	13.4	32.1	18.5	23.8	15.4	27.3	15.3	32.7
2	C ₁₅ H ₁₃ ClFNO	c134	15.4	14.3	12.8	8.3	8.3	24.2	15.3	13.2	17.2	17.3	7.5	24.5
2	C ₂₀ H ₁₅ N ₃ OS_dup ₁	c135	29.2	23.9	12.8	7.9	7.9	20.4	17.2	12.4	11.1	14.2	11.1	21.7
2	C ₁₄ H ₁₁ ClN ₂ O ₃ S	c136	13.9	14.8	6.1	6.7	6.7	29.6	22.6	6.3	7.1	8.3	9.4	30.5
2	C ₁₇ H ₁₁ BrN ₂ O ₃	c137	22.6	27.6	15.7	13.3	13.3	30.7	24.1	21.9	10.0	30.8	26.2	30.9
2	C ₁₄ H ₁₁ ClF ₃ N ₃ O	c138	16.0	17.5	9.0	9.4	9.4	11.5	19.1	12.4	4.8	-1.9	29.8	-0.3
2	C ₁₁ H ₁₀ N ₂ O ₂	c139	18.3	20.8	13.2	9.5	9.5	27.9	16.7	7.1	3.1	13.5	8.6	19.0
2	C ₁₉ H ₁₂ F ₂ O ₃	c140	24.5	26.3	11.1	10.8	10.8	17.1	20.6	2.1	6.9	12.0	11.4	18.6
2	C ₁₂ H ₈ F ₃ N ₃ OS	c141	-5.9	7.4	-8.4	-2.3	-2.3	-5.2	0.6	-19.2	-8.5	-12.8	-3.4	-1.4
2	n.d.	c142	8.8	13.6	10.8	7.8	7.8	20.1	14.3	9.9	6.8	10.5	8.4	23.8
2	C ₁₈ H ₁₂ N ₂ O ₃	c143	11.2	14.0	12.0	9.4	9.4	20.2	15.3	7.8	5.4	9.8	8.1	18.8
2	C ₁₆ H ₁₀ ClN ₅ O	c144	22.1	16.4	16.1	8.8	8.8	25.9	21.3	36.9	8.9	23.7	11.5	34.8
2	C ₂₂ H ₁₈ N ₂ O ₂	c145	38.9	29.8	32.7	20.0	20.0	43.2	26.4	29.2	22.0	34.9	26.4	39.5
2	C ₁₆ H ₁₅ ClN ₄ O ₃	c146	31.3	19.1	24.5	11.7	11.7	40.5	19.2	39.5	18.4	41.5	20.4	55.3
2	C ₁₄ H ₁₃ NO ₃ S	c147	61.7	41.1	63.0	28.1	28.1	67.7	33.6	56.1	25.4	72.4	37.8	69.7
2	C ₁₇ H ₁₃ N ₅ O	c148	11.1	15.2	9.3	7.3	7.3	25.1	18.3	10.4	4.4	15.5	10.9	27.6
2	C ₁₈ H ₁₁ ClN ₂ O ₃	c149	2.8	18.3	2.6	9.2	9.2	17.8	17.5	-1.8	6.5	14.3	13.3	12.3
2	C ₁₄ H ₁₀ F ₃ N ₃ O ₂ _dup ₁	c150	18.2	19.2	2.6	-5.5	-5.5	5.5	5.4	-9.8	-1.6	-2.6	-7.3	10.0
2	C ₁₇ H ₁₂ ClN ₅ O	c151	6.8	5.3	8.2	-2.8	-2.8	13.7	5.4	4.9	0.4	5.1	0.4	12.5
2	C ₁₆ H ₁₀ N ₄ O ₃	c152	10.1	6.0	10.9	-0.4	-0.4	26.0	6.5	-42.1	-6.7	65.0	19.4	55.3
2	C ₁₆ H ₁₁ F ₃ N ₄ O ₂	c153	5.9	5.8	5.2	1.9	1.9	14.0	11.5	4.2	6.9	12.5	9.4	19.3

Plate	Molecular formulae	Compound ID Concentration (µM)	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	150	30	150	30	150	30	150	30	150
2	C ₁₇ H ₁₃ FN ₆	c154	24.6	14.0	20.6	6.7	35.3	14.3	13.7	20.0	34.2	17.2	43.5	25.1
2	C ₁₆ H ₁₀ ClN ₃ O	c155	14.2	13.2	9.2	7.2	22.9	15.2	15.1	8.8	26.4	18.5	31.9	21.6
2	C ₁₆ H ₁₁ FN ₆	c156	20.5	19.3	15.9	6.3	25.5	11.0	13.3	12.5	24.0	20.1	35.1	24.0
2	C ₁₇ H ₁₁ ClF ₂ N ₂ O ₂	c157	94.5	81.1	96.4	79.9	97.0	78.7	78.3	97.4	98.2	82.3	96.0	82.1
2	C ₁₄ H ₁₇ N ₃ O ₂	c158	26.3	22.4	27.1	16.0	33.5	24.2	11.9	14.2	30.6	21.4	36.9	29.1
2	C ₂₀ H ₁₅ N ₃	c159	41.2	20.8	35.1	15.2	49.1	24.5	24.3	44.2	54.3	25.8	52.1	31.2
2	C ₂₄ H ₁₈ FN ₄ dup ₁	c160	8.1	13.6	-0.9	5.9	7.9	22.0	-0.7	11.7	12.7	10.8	17.9	18.8
3	C ₁₃ H ₁₁ N ₅ O	c161	-12.1	1.2	-18.6	-0.2	-7.8	12.2	-13.8	-32.7	-32.8	-19.1	-30.9	-16.2
3	C ₁₄ H ₆ F ₅ N ₃ O ₂	c162	-18.4	3.2	-13.7	3.3	-9.9	12.4	-2.6	-20.7	-36.5	-16.4	-31.8	-17.1
3	C ₁₉ H ₁₃ FN ₂ OS	c163	-7.8	9.1	-1.5	4.6	13.4	22.1	2.9	3.6	17.0	2.0	9.4	1.0
3	C ₁₉ H ₁₅ N ₃ O ₂ S	c164	87.2	41.2	72.9	23.6	82.8	30.6	18.7	69.0	82.2	13.9	92.3	31.1
3	C ₁₆ H ₁₂ N ₄ S	c165	19.2	43.1	-2.4	17.2	9.9	26.7	-6.0	0.1	6.3	17.1	1.2	20.0
3	C ₁₆ H ₁₆ N ₄ O	c166	-6.9	20.4	-1.6	11.0	4.0	23.4	-7.2	-7.2	-18.3	-5.8	-15.2	-8.4
3	C ₁₅ H ₉ Cl ₂ NO	c171	12.5	12.7	9.9	8.8	21.0	20.7	-4.7	-4.7	-6.1	2.2	-1.5	-8.5
3	C ₁₅ H ₁₀ F ₃ N ₃ O ₃ dup ₂	c172	-1.3	5.0	-2.2	1.7	9.2	14.8	-12.3	-5.7	-18.5	-14.1	-16.0	-10.6
3	C ₁₇ H ₁₂ FN ₇	c173	17.3	17.5	13.3	8.9	30.8	24.7	1.4	3.6	-3.5	-6.8	5.2	-4.4
3	C ₁₄ H ₁₀ F ₃ N ₃ O	c174	12.2	18.4	5.8	9.5	15.3	21.0	-0.2	2.6	-7.7	-7.0	-7.4	-3.2
3	C ₂₁ H ₁₄ ClN ₃	c175	49.5	38.9	52.7	33.2	57.8	48.6	4.5	11.0	37.0	50.1	26.8	38.2
3	C ₁₆ H ₁₁ FN ₆ dup ₁	c176	31.2	25.5	36.4	18.7	38.9	26.6	-3.5	-1.4	1.3	-0.1	8.8	-0.5
3	C ₁₇ H ₁₃ N ₅ O ₃	c181	14.7	14.7	11.8	6.7	21.7	19.2	-2.9	-2.9	-4.3	-9.8	-1.7	-7.7
3	C ₁₅ H ₁₀ F ₃ N ₃ O ₃ dup ₁	c182	-3.2	14.1	9.5	8.8	12.1	16.3	-2.5	0.9	-29.9	-12.7	-29.5	-10.4
3	C ₁₄ H ₉ ClFN ₃ S	c183	61.7	36.2	61.8	27.2	68.0	39.5	26.8	66.8	72.2	19.1	75.9	18.5
3	C ₁₃ H ₁₂ ClNO ₂ S ₂	c184	0.6	16.9	-0.6	4.5	10.7	13.6	-3.8	-7.9	-11.1	-12.2	-13.3	-5.7
3	C ₁₇ H ₁₄ N ₄ O ₂ S	c185	27.5	24.7	19.8	12.2	35.2	27.7	12.2	14.6	8.1	3.1	14.9	2.3
3	C ₁₇ H ₁₅ N ₃ O ₂ dup ₁	c186	45.0	32.5	41.7	25.6	50.2	40.1	16.1	33.2	30.8	12.3	29.3	10.5
3	C ₁₆ H ₉ BrF ₃ NO ₂ dup ₁	c191	15.4	14.7	4.7	5.2	41.2	24.9	5.2	20.8	19.9	9.6	15.8	-1.2

Plate	Molecular formulae	Compound ID Concentration (µM)	Percentage inhibition of AlphaScreen signal generated from protein-peptide interaction											
			TRIM24-PHD			TRIM24-bromo			TRIM24-dual			TRIM33-PHD		
			150	30	150	150	30	150	150	30	150	150	30	150
3	C ₁₉ H ₁₅ N ₅ O ₂	c192	16.7	18.8	13.4	10.0	10.0	23.3	19.7	2.2	5.5	0.8	-1.5	-2.2
3	C ₁₉ H ₁₈ N ₄ O	c193	18.2	21.0	17.8	9.1	9.1	29.7	22.5	12.5	7.7	11.2	-3.4	18.4
3	C ₁₇ H ₁₂ N ₄ OS	c194	50.4	24.1	32.6	2.4	2.4	61.0	30.7	42.5	14.4	38.8	5.5	63.3
3	C ₁₆ H ₁₂ CIN ₃ OS	c195	32.5	21.3	27.8	14.0	14.0	42.5	27.2	29.0	12.7	31.7	11.6	28.7
3	C ₁₇ H ₉ CIN ₂ O ₂ S	c201	57.8	40.7	50.5	28.9	28.9	59.4	42.8	40.3	21.1	44.8	20.2	47.1
3	C ₁₂ H ₆ F ₃ N ₃ O ₂	c202	22.9	19.4	26.0	12.4	12.4	36.5	20.8	18.8	8.8	22.7	2.6	24.9
3	C ₁₉ H ₁₉ NO ₄	c203	17.8	18.5	15.4	9.1	9.1	23.4	21.9	15.3	10.6	4.4	-3.0	9.7
3	C ₁₃ H ₁₀ CINO	c204	21.2	22.4	15.5	10.5	10.5	26.8	27.8	9.7	7.2	4.3	2.2	5.8
3	C ₁₀ H ₅ F ₂ N ₃ O	c205	10.0	18.6	12.8	8.8	8.8	24.6	24.3	12.0	-0.3	-2.7	-1.5	-1.7
3	C ₁₆ H ₁₅ N ₃	c211	18.9	8.9	13.9	6.0	6.0	30.3	17.0	11.3	-5.7	16.9	-5.1	34.0
3	C ₁₅ H ₁₀ F ₃ N ₃ O ₃	c212	12.5	18.2	10.8	6.0	6.0	23.3	19.9	3.7	3.4	0.7	-1.1	7.8
3	C ₁₉ H ₁₂ CIN ₃ O	c213	15.1	20.0	13.9	7.1	7.1	31.8	24.7	0.2	1.6	19.0	9.8	24.0
3	C ₁₈ H ₁₇ CIN ₄	c214	15.9	18.6	13.6	9.4	9.4	26.0	21.6	8.3	5.9	6.3	1.5	9.0
3	C ₁₈ H ₁₂ CIN ₅ O	c215	18.3	20.9	12.8	4.6	4.6	28.3	15.3	7.0	17.2	10.3	3.6	13.4
3	C ₁₉ H ₁₃ N ₃ O ₂	c222	12.8	13.7	15.1	9.8	9.8	27.9	16.2	7.6	4.2	12.5	3.4	14.6
3	C ₁₄ H ₁₄ N ₄ O ₂ S ₂	c223	-5.0	6.6	1.9	-1.6	-1.6	17.4	10.1	6.5	-3.4	-3.4	-7.1	8.0
3	C ₁₇ H ₁₄ N ₆ O	c224	-4.2	11.6	-1.1	6.8	6.8	13.5	20.2	-3.6	9.1	-10.7	-1.6	-2.7
3	C ₁₇ H ₁₂ BrN ₇	c225	5.5	12.5	0.9	-6.0	-6.0	19.0	12.4	-1.8	-6.8	-5.0	-9.7	28.3
3	C ₁₆ H ₉ BrF ₃ NO ₂	c232	-5.6	-1.7	1.4	3.6	3.6	16.0	13.4	-17.1	0.2	2.7	-1.5	8.1
3	C ₁₆ H ₁₇ N ₅ O ₃	c233	11.9	11.6	15.8	4.7	4.7	27.0	15.4	12.1	13.9	11.5	6.9	11.7
3	C ₂₂ H ₂₅ N ₃ O	c234	3.4	-0.3	-2.7	-3.8	-3.8	18.5	20.6	23.2	14.3	17.4	1.2	9.8
3	C ₁₄ H ₁₀ F ₃ N ₃ O ₃	c235	-9.6	-5.2	14.0	-8.2	-8.2	33.1	10.8	16.0	3.3	17.6	7.4	24.4

3.3.2 PPI-net collection screening

The PPI-net screening collection previously introduced (see page 67), was supplied as 10 mM DMSO stocks in 96-well V-bottomed microplates (Greiner). These compounds were first transferred to 20 LabCyte 384-LDV plates for use with the Echo acoustic liquid handler (Labcyte) using a Bravo Automated Liquid-Handling Platform (Agilent). In a similar fashion to Maybridge screening (Figure 3.9), samples of 20 or 100 nL were transferred to assay plates in duplicate to test at 10 or 50 μ M FAC against TRIM24 and TRIM33 with a complementary set of competitor peptides in an AlphaScreen displacement assay: H3₁₋₂₇K9Me₃ (PHD) and H3₁₋₂₇K9Me₃K18Ac (dual). Average inhibition was calculated as a percentage of the maximum signal obtained with buffer reference wells (0.5% FAC DMSO), and appended to the PPI-net screening collection SDF (CD appendix: AlphaScreen\CompoundScreening\PPInetPrimaryScreen.icb). 40 ligands resulted in >40% inhibition were collated to review (Table 3.4).

Table 3.4 Inhibition levels observed in AlphaScreen assay for 40 compounds of interest from the PPI-net collection. Inhibition was normalised to buffer reference wells (0.5% FAC DMSO). Compounds were identified based on their positions on assay plates, and screened against TRIM24 and TRIM33, with two different competitor peptides: PHD = H3₁₋₂₇K9Me₃; dual = H3₁₋₂₇K9Me₃K18Ac. n.d. = not determined. From 1534 compounds, these 40 were selected based on inhibition and selectivity across the assay. Shading on cells represents inhibition levels from low (green) to high (red).

PPI-net Ligand	Plate	Position	MW	Assay format/ Ligand concentration (μM)							
				TRIM24-PHD		TRIM24-dual		TRIM33-PHD		TRIM33-dual	
				50	10	50	10	50	10	50	10
C ₂₂ H ₃₀ N ₆ O ₃	1	B06	426.52	6.9	7.1	70.1	13.5	17.2	8.2	21.2	23.8
C ₂₇ H ₃₀ N ₄ O ₂	1	F10	442.56	13.7	14.5	99.7	18.7	23.5	20.9	28.6	25.7
C ₂₁ H ₂₉ N ₅ O ₄ S	2	B10	447.56	n.d.	n.d.	50.8	47.1	6.8	-6.7	30.2	39.2
C ₂₄ H ₃₁ FN ₄ O ₃	2	G08	442.53	n.d.	n.d.	74.9	42.6	71.0	32.3	82.3	44.7
C ₂₄ H ₂₈ N ₆ O ₂	4	A06	432.53	36.4	-1.0	31.8	32.4	44.3	28.6	46.1	22.6
C ₂₄ H ₃₁ FN ₄ O ₃	5	C03	442.53	48.3	24.0	59.2	38.0	47.9	22.7	46.2	32.7
C ₂₅ H ₃₁ FN ₂ O ₂	5	C11	410.53	12.8	15.9	53.8	60.0	23.1	20.9	83.7	83.7
C ₂₃ H ₃₂ FN ₃ O ₃	5	D11	417.52	19.2	21.1	43.6	50.9	26.7	22.2	85.9	86.4
C ₂₅ H ₃₁ NO ₃	7	A11	393.52	94.8	86.9	92.8	91.8	91.3	38.4	96.4	94.3
C ₂₇ H ₃₆ N ₄ O ₂	7	D03	448.61	60.0	33.3	60.9	48.9	48.3	30.4	55.0	42.2
C ₂₅ H ₂₉ N ₅ O ₂	7	E05	431.54	99.5	75.8	98.6	79.7	97.7	73.1	99.0	80.3
C ₂₂ H ₂₂ FN ₅ O ₂	7	F06	488.36	50.8	15.9	63.3	45.1	61.8	39.6	70.8	42.4
C ₂₃ H ₂₆ N ₆ O ₃	8	A11	434.50	53.6	6.3	61.3	41.0	-7.0	-2.5	79.5	54.5
C ₂₅ H ₃₃ N ₅ O ₂	8	F06	435.57	65.6	31.8	69.7	48.0	-22.1	3.5	76.0	53.3
C ₂₆ H ₃₁ N ₃ O ₃	8	H04	514.46	45.6	16.9	60.8	30.7	18.6	30.7	77.7	47.1
C ₂₇ H ₃₃ N ₃ O ₃	9	E08	447.58	35.9	22.4	45.6	35.1	59.0	41.7	60.6	59.2
C ₂₂ H ₂₄ F ₃ N ₃ O	10	F05	403.45	99.2	57.1	97.3	64.5	96.5	49.3	98.6	74.7
C ₂₅ H ₂₇ N ₅ O	10	F09	413.52	99.4	80.1	98.3	82.8	98.1	75.3	99.0	89.8
C ₂₄ H ₃₁ N ₃ O ₂	10	F11	393.53	40.8	34.5	60.0	39.2	15.9	28.2	47.3	45.0
C ₂₆ H ₃₄ N ₄ O ₂	10	B09	434.58	29.6	17.9	78.4	32.0	-33.8	-19.2	13.2	16.3
C ₂₆ H ₃₅ N ₃ O ₃	11	C10	437.58	20.5	16.4	44.5	37.6	-49.7	-23.8	18.9	14.6
C ₂₄ H ₃₀ N ₄ O ₃	13	B09	422.53	17.0	12.5	99.2	42.9	-34.5	-15.4	9.8	4.9
C ₂₈ H ₃₆ N ₂ O ₄	13	D06	464.60	14.6	10.2	37.1	99.1	11.9	6.6	7.2	2.4
C ₂₅ H ₂₆ FN ₃ O ₃	13	E09	407.48	13.8	13.6	46.2	43.0	-44.0	-27.8	6.3	6.1
C ₂₅ H ₃₃ N ₃ O ₂	14	A08	407.55	-11.5	6.9	19.6	26.0	-92.7	-34.0	-11.5	6.9
C ₂₅ H ₃₁ FN ₂ O ₄	14	B06	442.53	10.2	11.7	78.6	43.1	5.0	5.8	10.2	11.7
C ₂₆ H ₂₉ N ₅ O ₃	14	B11	459.55	10.8	8.4	83.1	27.2	-1.8	8.7	10.8	8.4
C ₂₇ H ₃₅ NO ₄	14	H07	437.58	47.9	34.2	56.4	43.3	48.8	33.2	47.9	34.2
C ₂₆ H ₃₃ N ₃ O ₂	14	E03	419.57	55.7	29.2	66.0	40.1	20.3	7.5	55.7	29.2
C ₂₆ H ₂₈ ClNO ₃	15	G03	437.96	99.1	65.0	98.9	85.4	98.1	64.1	99.1	65.0
C ₂₀ H ₂₇ N ₇ O ₂ S	16	D03	429.55	69.5	20.9	70.9	52.7	45.9	57.2	69.5	20.9
C ₁₈ H ₂₂ N ₄ O ₃ S	16	D09	410.93	17.4	5.7	57.0	43.8	73.4	53.4	17.4	5.7
C ₁₉ H ₂₆ F ₂ N ₂ O ₂	17	B04	352.43	87.6	54.9	83.1	62.6	95.8	55.7	87.6	54.9
C ₁₈ H ₂₅ ClN ₂ O ₅	17	B11	384.86	61.1	21.0	69.3	50.1	61.3	14.8	61.1	21.0
C ₁₉ H ₂₁ ClN ₄	17	D02	340.86	56.2	34.2	69.4	58.6	25.6	1.6	56.2	34.2
C ₁₅ H ₁₂ ClN ₃ O ₃ S	17	H08	349.80	33.4	6.2	55.0	30.7	68.9	15.8	33.4	6.2
C ₁₉ H ₂₇ N ₇ OS	19	C08	401.54	52.9	13.9	65.5	49.7	35.2	16.0	52.9	13.9
C ₁₈ H ₁₈ N ₄ O ₄ S	19	F07	386.43	40.4	13.6	66.2	45.2	39.6	37.4	40.4	13.6
C ₁₅ H ₁₄ FN ₅ O ₂ S	20	A03	347.37	-11.7	-47.5	79.4	53.0	6.9	8.3	-11.7	-47.5
C ₁₇ H ₂₀ N ₄ OS	20	C09	328.44	-5.7	-31.3	69.6	58.5	32.2	20.2	-5.7	-31.3

3.4 Hit validation

The primary disadvantage of the AlphaScreen platform for high-throughput screening is the high false positive rate obtained, due to compound interference with chemiluminescence rather than inhibition of binding. A commercial TruHits Kit (GE Healthcare) was used to validate specific compounds following primary screening.

After analysis of assay interference, and removal of compounds which were likely to be false positives, we sought a method of validating the remaining compounds as true binding partners, through orthogonal, label-free binding studies. Based on the compound concentrations available, DSF was used as a secondary screen for PPI-net hits, and the NMR technique WaterLOGSY was used for Maybridge compound hits.

3.4.1 AlphaScreen TruHits kit

After primary screening, 16 Maybridge compounds and 40 PPI-net compounds were tested for non-inhibitory interference in the AlphaScreen format using TruHits beads (Perkin Elmer) as a control. This includes biotinylated acceptor beads (b-A) and streptavidin coated donor beads (SA-D) to generate a non-specific high emission signal, to observe the compound effects on emission in the absence of protein.

The main mechanisms of compound interference in AlphaScreen are: i) quenching the singlet oxygen, ii) competing against the interaction between biotin and streptavidin, and iii) acting as an inner filter. Strong singlet oxygen quenchers include antioxidants such as azide and ascorbate, compounds with structures based on thiophene (when at high concentrations of 100 μ M to 10 mM), and other groups which were not included in our screening library (transition metals used in the micromolar range, and transition metal-containing hemes, such as hemoglobin). Competitors of the interaction between biotin and streptavidin are biotin-like structures which may contain thiophenes and ureas. Compounds which absorb light at 520-620 nm or 675-685 nm, the respective emission and excitation wavelengths for AlphaScreen, are known as 'inner filters' which would also quench

signal in a concentration dependent manner. Two formats can identify false positives: pre-incubation of b-A and SA-D beads prior to addition to hit compounds in assay wells, which determines quenchers or filters; or pre-incubation of compounds with SA-D beads before addition of b-A beads, to additionally identify those which interfere with the streptavidin-biotin interaction.

16 Maybridge compounds were assessed the TruHits assay (Table 3.5). Compounds were obtained from dry stores (Pfizer Neusentis), and solubilised in fresh DMSO to 50 mM. Compounds c07 and c14 were limited by solubility, and lower concentration stocks of 40 and 30 mM respectively were prepared in the same manner. The assay was performed in duplicate for both pre-incubation formats, at comparable concentrations to in the primary screen: 31 μ M with five two-fold increases in concentration to 500 μ M (final DMSO content 1%). High concentrations were used to sensitively determine any non-specific effects of the compound in the assay. A signal deviation from the average of reference blank wells by more than two standard deviations indicates an interfering compound. Three compounds, which resulted in average standard deviations of 1.92 ± 0.34 , 4.97 ± 0.65 and 3.88 ± 0.34 , were shown to be interfering in the fluorescence readout. Data for four compounds were inconclusive, as they resulted in an observable reduction in signal (1.2-1.5 standard deviations), which was none-the-less lower than inhibition observed in the screening assay. Eight compounds were taken forwards for further validation.

Of the 40 PPI-net compounds subjected to the TruHits assay, the majority showed very strong assay interference (Table 3.6). Four two-fold concentration dilutions from 200 to 25 μ M (2% final assay DMSO concentration) were assessed in the preincubated bead format. 22 of these compounds resulted in a normalised standard deviation of >4 , which is twice the recommended limit. For eight compounds reductions in signal during primary screen did not seem to be caused by compound effects, and were likely to be inhibitory. These compounds were taken forward for further validation.

Table 3.5 AlphaScreen emission of 16 Maybridge fragments (structures left) subjected to TruHits assay. Duplicates of values obtained from preincubation of TruHits beads (A) or preincubation of compounds with streptavidin coated donor beads (B). STD – standard deviation normalised to the standard deviation of reference buffer wells; red background – normalised compound standard deviations >1.5; quenching – compounds determined to be assay artifacts; block arrow – compounds taken forward for validation.

Compound		Final assay concentration (μM)					Overall STD (normalised)
		500	250	125	62.5	31.2	
	A	77084	101540	110363	113876	116112	2.28
		90670	99598	107664	112826	116286	1.46
	B	74044	81018	102829	107194	104850	2.04
		67377	90354	100680	98066	99081	1.91
	A	84902	90295	94411	95504	102745	1.05
		88391	94472	89281	95585	101455	0.84
	B	82270	83540	80867	84701	90503	0.55
		89437	80043	80091	73845	83127	0.87
	A	25096	28523	41332	45049	62007	5.40
		25656	41942	40468	47387	70122	5.31
	B	27038	27196	34276	45929	66474	5.18
		28379	23841	29989	41660	51186	4.00
	A	96427	101420	102377	107155	108232	0.69
		98579	93706	91779	102111	105930	0.88
	B	93516	93603	98616	98710	105502	0.63
		97978	86743	94953	93281	87601	0.66
	A	85624	88565	94090	100336	96727	0.95
		83772	93193	93271	88625	94959	0.75
	B	81231	84003	87746	92112	96806	0.88
		86162	76142	82027	75868	77636	0.70
	A	95519	91586	82030	87287	89825	0.84
		91089	86402	79849	83396	87785	0.74
	B	72749	71588	70922	77500	83381	0.87
		71621	64333	65183	70073	66869	0.58
	A	68658	74480	75098	80444	85808	1.26
		69014	80052	77563	85311	83526	1.20
	B	57076	66616	62596	72803	81320	1.73
		61051	62416	65231	66841	70578	0.72
	A	52983	49218	52919	61165	70093	2.18
		53063	70276	62111	73680	72569	1.95
	B	47575	43302	45670	50788	52504	0.98
		46473	44986	47900	49381	52091	0.71
	A	80053	88540	88847	91558	95986	0.97
		83271	85951	87981	92972	97476	0.94
	B	76224	77857	74616	88480	88036	1.03
		65777	60009	79624	77809	79780	1.58
	A	80101	89426	93426	98700	100372	1.31
		84976	91058	92303	96635	102205	1.02
	B	74237	83759	84636	93810	97084	1.31
		71698	79422	91630	85742	79761	1.15
	A	43923	57609	73730	83496	89149	3.99
		42442	61384	73289	85030	93721	4.21
	B	40018	52643	63750	82698	89840	3.94
		35640	51300	69618	73185	71340	3.40
	A	64959	71601	80209	91679	95488	2.38
		65022	76482	80943	87041	94213	2.03
	B	65515	68847	70339	89712	94059	2.12
		59118	64452	78189	73723	73283	1.39
	A	167220	177865	173254	187343	129650	1.97
		171937	168513	146078	127944	131451	2.03
	B	143959	142501	135858	153502	150302	0.60
		141972	140430	151004	139818	125679	0.81
	A	89811	95193	95777	98113	99528	0.58
		89871	90623	89867	93225	97157	0.50
	B	94837	90301	82221	94926	94567	0.75
		82478	78699	85053	78236	90145	0.74
	A	82404	70451	84022	92107	96719	1.76
		65744	70781	79863	83431	93346	2.04
	B	73796	79604	79942	87105	92881	1.12
		77188	83341	76173	74204	82295	0.63
	A	97667	101911	100417	102009	103437	0.32
		97690	95974	94999	97781	101323	0.37
	B	99100	93017	92644	97756	104645	0.63
		98212	91159	88053	87396	91984	0.59

AlphaScreen emission (a.u.)

Table 3.6 AlphaScreen emission of 40 PPI-net fragments subjected to TruHits assay. STD – standard deviation normalised to the standard deviation of reference buffer wells. Compounds were determined to be assay artifacts if normalised compound standard deviation >1.9 (red background). Remaining nine compounds were taken forward for validation.

Position	Final assay concentration (μM)				STD
	200	100	50	25	
1 C11	49901	55608	53633	56863	1.64
1 K19	50479	52471	49538	54212	1.18
2 C19	19968	23904	28078	33349	6.36
2 M15	40069	45850	45633	53354	3.44
4 A11	32114	40967	44353	51799	5.63
5 E05	18317	28593	35511	47542	11.03
5 E21	50077	52819	52941	56710	1.50
5 G21	49409	53684	53269	59908	2.35
7 A21	39343	45138	46905	50563	3.00
7 G05	19569	23030	30171	35783	7.81
7 I09	27754	35134	39378	42562	5.16
7 K11	49902	50531	51560	54687	1.20
8 A21	36060	45508	48165	52133	4.39
8 K11	7383	14211	23086	34165	17.16
8 O07	35297	46020	49459	55723	5.36
9 I15	32340	40300	43947	52863	5.87
10 K09	27209	38189	44286	53110	7.81
10 K17	52760	53404	54918	56447	0.88
10 K21	25345	29105	32430	37731	4.92
10 C17	31354	36663	40753	46106	4.71
11 E19	33374	41674	44545	48129	4.38
13 C17	38862	45845	48174	52135	3.52
13 G11	46750	53197	53270	51613	1.75
13 I17	29321	33216	39293	43746	5.13
14 A15	37704	43251	45374	51302	3.69
14 C11	41824	47475	49099	53450	2.93
14 C21	48046	51767	50000	52502	1.15
14 O13	25037	24988	26982	33314	4.16
14 I05	19114	26934	32450	38931	8.36
15 M05	44420	45750	46868	51647	1.95
16 G05	38240	45786	48103	50826	3.45
16 G17	52380	55689	51486	52311	1.03
17 C07	8555	23973	37379	53017	18.01
17 C21	15981	21319	29571	40557	11.64
17 G03	12579	19518	25335	33109	11.24
17 O15	29680	39897	42722	49330	5.90
19 E15	33933	43435	46484	49864	4.61
19 K13	28259	40129	47736	50106	6.91
20 A05	38793	48111	48929	50526	3.32
20 E17	44774	48121	50569	51564	1.81

AlphaScreen emission (a.u.)

3.4.2 Differential scanning fluorimetry

In Chapter 2, a differential scanning fluorimetry (DSF) assay was described, that was able to detect concentration dependant stabilisation of TRIM24 when exposed to a probe compound. Eight compounds, which had been identified as possible binding partners for TRIM24 or TRIM33 from the PPI-net compound collection, were subjected to the same assay in an attempt to validate their interactions.

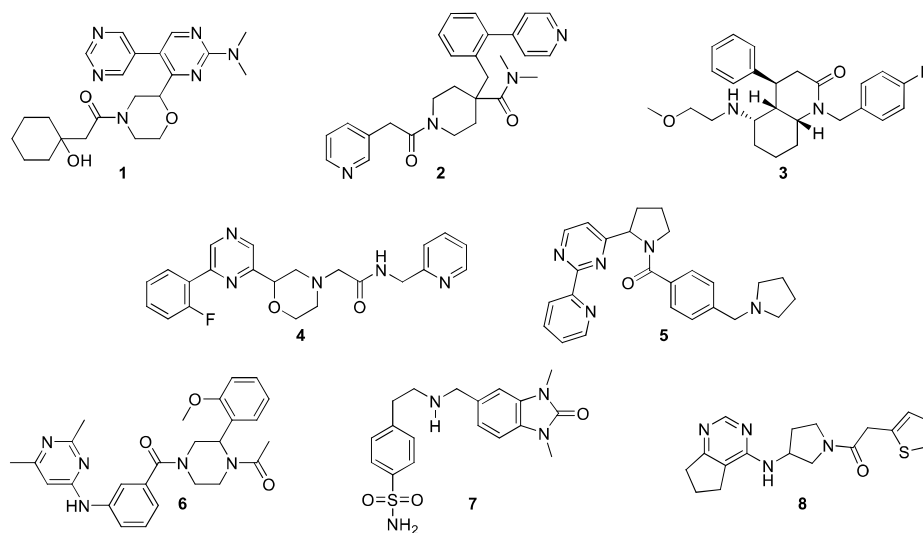


Figure 3.10 Lead molecules from PPI-Net compound library screening assessed in differential scanning fluorimetry. Compounds were tested from 10 mM DMSO stocks, stored in 96-well plates at -78°C .

Compounds were tested at a final assay concentration of 20 or 10 μM , against TRIM24 and TRIM33 in the WT, and inactive domain forms (TRIM24: W828A, N980F. TRIM33 W889A, N1039F). Disappointingly, none of these demonstrated any significant shift in melting temperature which could be attributed to thermal stabilisation from compound binding (Figure 3.11). Proteins were incubated with compound concentrations at the limit of the DMSO tolerance (20 μM compound 0.5% DMSO). Failure to positively confirm binding under these conditions was not surprising, since these potential starting points were identified with low affinities to the target proteins. Resynthesis of hits identified through AlphaScreen, followed by solubilisation as a more concentrated DMSO stock (e.g. 50 mM) would allow the assay to be conducted at much higher compound concentrations, which could help confirm any positive binding partners for TRIM24 and TRIM33 through DSF.

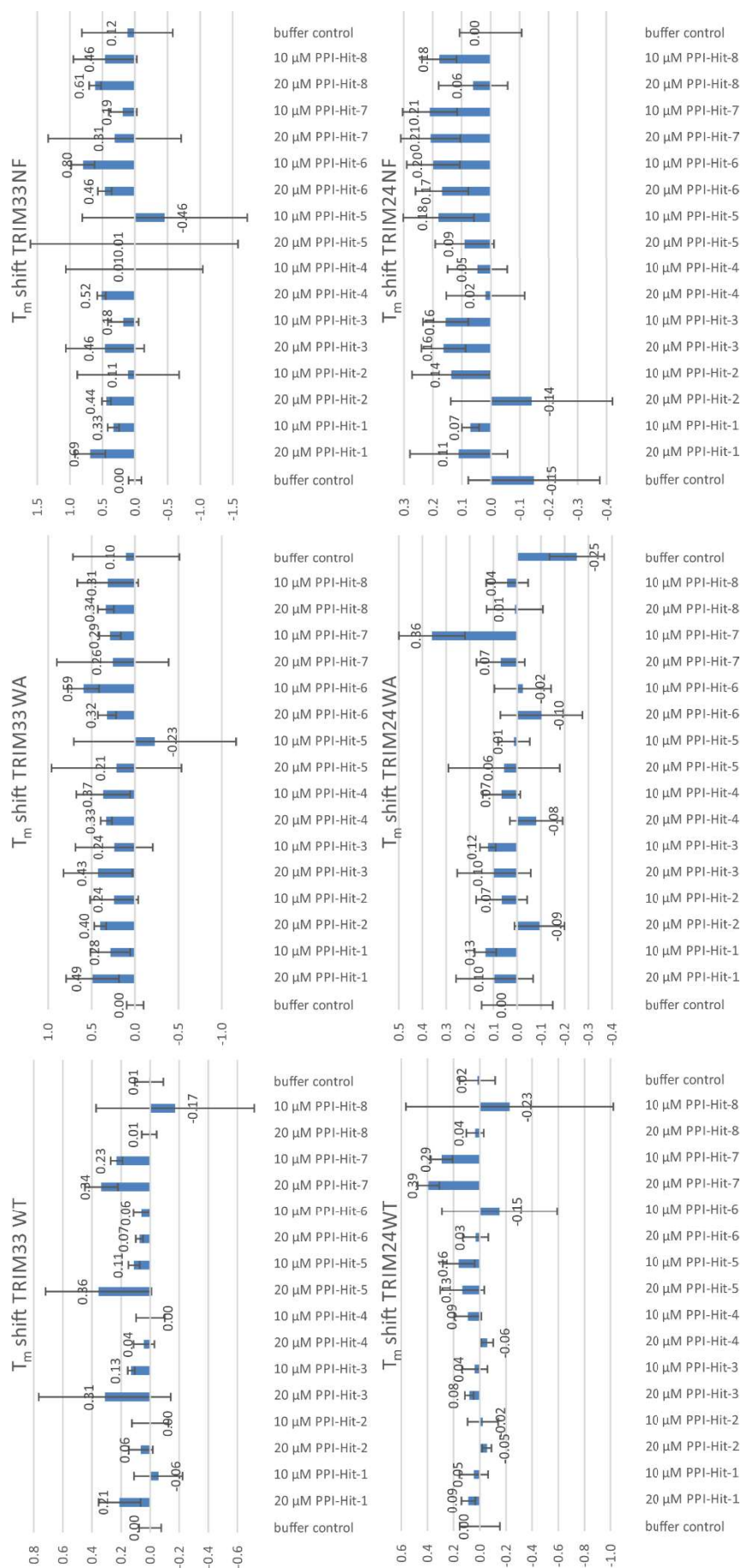


Figure 3.11 T_m shift results from PPI-net 'Hit' compounds against TRIM24 and TRIM33. Compounds were included at two concentrations (20 μM and 10 μM), measured from 20-80 °C in 1 °C intervals with a hold of 1.5 min. Domain inactive mutants (PHD mutant WA; bromodomain mutant NF), were included to attribute any stabilisation to each domain. Error bars are shown as standard deviations of results in triplicate. Overall the results profile was fairly flat, with no compounds demonstrating stabilisation beyond background noise.

3.4.3 Maybridge fragment characterisation

Dose-response curves were obtained for seven Maybridge compounds, identified in primary screening of ligands for TRIM24 or TRIM33, whose AlphaScreen signal was determined not to have been caused by compound interference in the assay (Table 3.7). Data could not be obtained for compound c02 due to solubility limitations. Compounds were tested at the highest

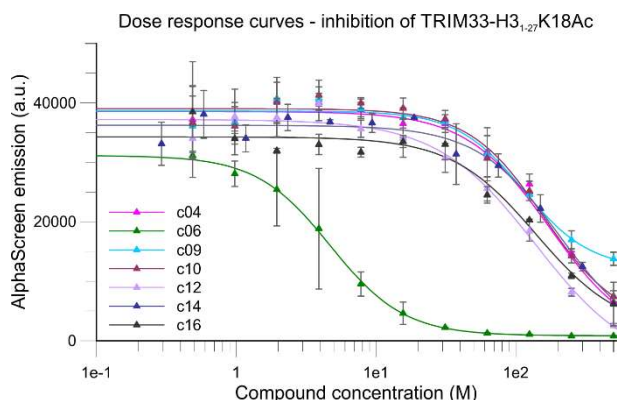


Figure 3.12 Representative dose response curves obtained for seven Maybridge compounds in AlphaScreen assay detecting TRIM33-H3₁₋₂₇K18Ac interaction. Each titration was performed in duplicate, however, these did not span the full range of inhibition. IC₅₀ values calculated from fitting to four-parameter fit using GraFit.

concentrations possible within the known DMSO tolerance of the assay (500 μM, 1% DMSO), however, apart from compound c06 (IC₅₀ values 2.10-9.18 μM), these titrations did not span a full range of inhibition, and errors in curve fitting (to a four-parameter fit) in order to determine an IC₅₀ value were large (Figure 3.12). Despite this, these results do provide an indication of the relative binding affinity of each compound to the target proteins.

Table 3.7 IC₅₀ value approximations for inhibition of TRIM24 and TRIM33 histone peptide recognition. Values reported to 3 s.f. Error reported as standard error of IC₅₀ for four-parameter fitting.; n.d. – values for which error of fitting was larger than IC₅₀ value. Geometric mean calculated where suitable. H3₁₋₂₇K9Me₃, H3₁₋₂₇K18Ac, and H3₁₋₂₇K9Me₃K18Ac, competitor peptides report inhibition at PHD finger, bromodomain, and dual construct, respectively.

Compound	Protein	Peptide	IC ₅₀ (μM)	Geometric mean – where suitable (μM)
c04	TRIM24	H3 ₁₋₂₇ K9Me ₃	312±197	-
		H3 ₁₋₂₇ K18Ac	>1000	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	>1000	
	TRIM33	H3 ₁₋₂₇ K9Me ₃	n.d.	221±143
		H3 ₁₋₂₇ K18Ac	n.d.	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	231±45.7	
c06	TRIM24	H3 ₁₋₂₇ K9Me ₃	3.22±0.18	2.64±0.27
		H3 ₁₋₂₇ K18Ac	2.73±0.25	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	2.10±0.50	
	TRIM33	H3 ₁₋₂₇ K9Me ₃	9.18±1.13	4.64±0.40
		H3 ₁₋₂₇ K18Ac	4.71±0.21	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	2.31±0.28	
c09	TRIM24	H3 ₁₋₂₇ K9Me ₃	n.d.	-
		H3 ₁₋₂₇ K18Ac	n.d.	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	679±538	

Compound	Protein	Peptide	IC ₅₀ (μM)	Geometric mean – where suitable (μM)
	TRIM33	H3 ₁₋₂₇ K9Me ₃	n.d.	128±41.9
		H3 ₁₋₂₇ K18Ac	113±20.2	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	144±86.8	
c10	TRIM24	H3 ₁₋₂₇ K9Me ₃	3.69±7.07	-
		H3 ₁₋₂₇ K18Ac	228±76.4	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	n.d.	
	TRIM33	H3 ₁₋₂₇ K9Me ₃	n.d.	-
		H3 ₁₋₂₇ K18Ac	158±45.4	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	176±24.2	
c12	TRIM24	H3 ₁₋₂₇ K9Me ₃	>1000	
		H3 ₁₋₂₇ K18Ac	207±41.3	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	1000±615	
	TRIM33	H3 ₁₋₂₇ K9Me ₃	n.d.	161±53.0
		H3 ₁₋₂₇ K18Ac	139±45.4	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	186±62.0	
c14	TRIM24	H3 ₁₋₂₇ K9Me ₃	198±101	-
		H3 ₁₋₂₇ K18Ac	>1000	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	n.d.	
	TRIM33	H3 ₁₋₂₇ K9Me ₃	260±256	228±192
		H3 ₁₋₂₇ K18Ac	n.d.	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	220±133	
c16	TRIM24	H3 ₁₋₂₇ K9Me ₃	163±28.2	195±42
		H3 ₁₋₂₇ K18Ac	176±30.6	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	257±84.7	
	TRIM33	H3 ₁₋₂₇ K9Me ₃	190±36.3	146±48
		H3 ₁₋₂₇ K18Ac	140±60.5	
		H3 ₁₋₂₇ K9Me ₃ K18Ac	118±49.5	

In addition to the apparent high affinity of compound c06, of significant interest is the behaviour of compound c12, which demonstrated a higher level of inhibition of histone recognition for the bromodomain than the PHD finger, for each protein. Conversely, compound c04 inhibited the TRIM24-H3₁₋₂₇K9Me₃ interaction with an IC₅₀ value of 312±197 μM, whereas inhibition of the bromodomain was beyond detection in this assay.

Attempts were made to expand the range of inhibition of TRIM33 peptide recognition detected. The possibility that higher DMSO concentrations required to achieve this would contribute to inhibition of the bromodomain were mitigated by performing dilutions at several DMSO percentages. These limits dictated maximum compound concentrations: 1 mM at 2.0% DMSO, 500 μ M at 1.0% DMSO, and 250 μ M at 0.5% DMSO. Reassuringly, normalised dose response curves overlaid from each of these titrations suggested DMSO concentration did not affect inhibition (Figure 3.13). Titrations were performed against each TRIM33 competitor peptide, and collated as geometric means (Table 3.8). IC_{50} values obtained

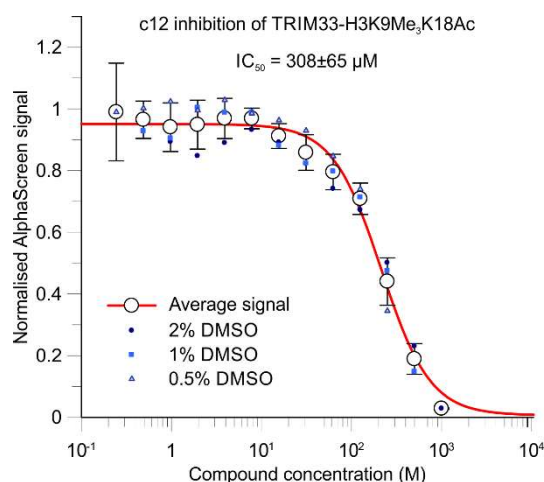


Figure 3.13 Representative dose response curve obtained with different assay DMSO percentages: compound c12 in AlphaScreen assay detecting TRIM33-H3₁₋₂₇K9Me₃K18Ac interaction. Each titration was performed in duplicate, and normalised against blank wells with the corresponding DMSO percentage. IC_{50} values calculated from fitting of average signal to four-parameter fit using GraFit.

in this method were approximately two-fold lower than previously calculated. This is likely to be a more accurate representation of true inhibition levels, since fitting was performed over a greater range of inhibition, resulting in considerably lower errors.

Table 3.8 IC_{50} values of selected Maybridge compounds inhibiting TRIM33 recognition of H3 peptides. Titrations performed in duplicate at three DMSO percentages (2.0%, 1.0%, 0.5%). Data was fitted in GraFit using four-parameter fit. Error reported as standard error of IC_{50} for four-parameter fitting.; n.d. – values for which error of fitting was larger than IC_{50} value. Values for c09 were deemed unreliable due to high error and incomplete range of inhibition observed in titration. $IC_{50} > 1000$ – fitting not possible across assay concentration range.

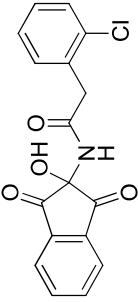
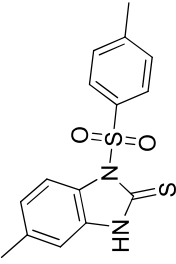
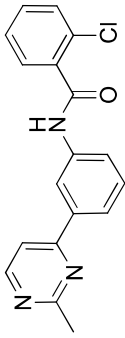
Compound	Peptide	IC_{50} (μ M)	Geometric mean (μ M)
c04	H3 ₁₋₂₇ K9Me ₃	449 $\pm$ 49	473 $\pm$ 86
	H3 ₁₋₂₇ K18Ac	453 $\pm$ 147	
	H3 ₁₋₂₇ K9Me ₃ K18Ac	521 $\pm$ 88	
c09	H3 ₁₋₂₇ K9Me ₃	>1000	incomplete inhibition curve
	H3 ₁₋₂₇ K18Ac	n.d.	
	H3 ₁₋₂₇ K9Me ₃ K18Ac	227 $\pm$ 104	
c10	H3 ₁₋₂₇ K9Me ₃	>1000	262 $\pm$ 40
	H3 ₁₋₂₇ K18Ac	259 $\pm$ 52	
	H3 ₁₋₂₇ K9Me ₃ K18Ac	266 $\pm$ 30	
c12	H3 ₁₋₂₇ K9Me ₃	529 $\pm$ 291	373 $\pm$ 166
	H3 ₁₋₂₇ K18Ac	319 $\pm$ 241	
	H3 ₁₋₂₇ K9Me ₃ K18Ac	308 $\pm$ 65	
c16	H3 ₁₋₂₇ K9Me ₃	284 $\pm$ 27	259 $\pm$ 50
	H3 ₁₋₂₇ K18Ac	291 $\pm$ 133	
	H3 ₁₋₂₇ K9Me ₃ K18Ac	212 $\pm$ 35	

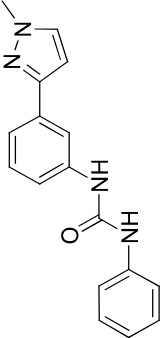
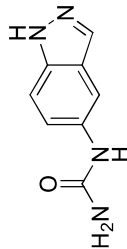
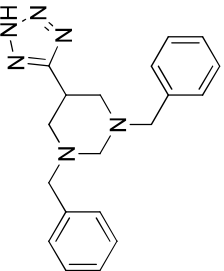
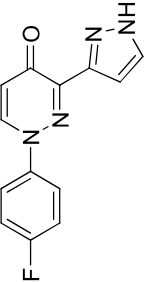
To provide further information on the suitability of ligands identified as lead compounds for probe development, the same seven compounds (c04, c06, c09, c10, c12, c14, c16) were analysed to generate computational predictions of molecular properties and pharmacodynamics testing (Table 3.9). All compounds possessed a MW <350 Da as a condition of their inclusion in the Maybridge Fragment collection. Measured distribution-coefficients at physiological pH (LogD) correlated well with those predicted (cLogD) which were within 5-31%, with the exception of c6 (LogD not determined) and c14 (LogD 63% higher than predicted). The difference between LogP and LogD for c14 reflects the acidity of the basic tertiary amine which would be protonated at pH 7.4. The topological polar surface area (TPSA) is a parameter for prediction of drug transport properties, including bioavailability.¹⁹ The spread for the Maybridge fragments was 55-84 Å² which suggests they may be sufficiently non-polar to cross the CNS barrier.

Ligand efficiency ($LE = pIC_{50} \div \text{number of heavy atoms}$) compares molecules according to their average binding energy per atom,²⁰ while lipophilic ligand efficiency ($LLE = pIC_{50} - cLogP$) extends this by defining the specificity of a molecule for binding to a target relative to partitioning.²¹ Drug like molecules have been shown to commonly possess values of $LE > \sim 0.3$ and $LLE > \sim 5$. The range of values for the Maybridge compounds, calculated from results from IC_{50} curves, were consistent with small ligands prior to any optimisation: four molecules showed particular potential with $LE > 0.2$ and $LLE > 2.5$.

Most compounds had a reasonable solubility of 300-500 µM, which should allow for effective dosage of ligands after optimisation. Intrinsic clearance, the ability of hepatic enzymes to metabolize the molecule, could not be determined for two compounds, but in all others was <26 µL/min/mg which would not restrict the duration of action (in humans, >47 µL/min/mg is considered to be high clearance).

Table 3.9 Calculated and measured pharmacodynamics data for seven Maybridge fragments. These provide information on suitability for probe development. MW – molecular weight; cLogP – calculated partition coefficient; (c)LogD – (calculated) distribution coefficient; TPSA – topological polar surface area; (L)LE – (lipophilic) ligand efficiency; cHLM – calculated human liver microsomal stability; MDCK/MDR1 – Madin-Darby canine kidney cells transfected with the human MDR1 gene.

Compound	Permeability P_{app} (10^{-6} cm/sec)			
	Passive Permeability	MDCK/MDR1: AB	MDCK/MDR1: BA	MDCK/MDR1: BA/AB
	Intrinsic Clearance (μ L/min/mg)	n.d.	<0.100	n.d.
	cHLM	21.00	<0.100	n.d.
Solubility (μ M) pH 6.5	458	21.00	<0.100	n.d.
LLE	2.80	21.00	<0.100	n.d.
LE	0.23	21.00	<0.100	n.d.
No. Heavy Atoms	23	21.00	<0.100	n.d.
Estimated IC ₅₀ (μ M)	230	21.00	<0.100	n.d.
BASIC pK _a	n.a.	21.00	<0.100	n.d.
ACID pK _a	9.10	21.00	<0.100	n.d.
TPSA ($\text{\AA}^2$)	83.47	21.00	<0.100	n.d.
LogD pH 7.4 (cLogD)	1.758 (2.31)	21.00	<0.100	n.d.
cLOGP	1.14	21.00	<0.100	n.d.
MW (Da)	330	21.00	<0.100	n.d.
c04 		21.00	<0.100	n.d.
c06 		75.30	n.d.	n.d.
c09 		46.60	20.890	1.116

Compound	Permeability P_{app} (10^{-6} cm/sec)			
	MDCK/MDR1: BA/AB	MDCK/MDR1: BA	MDCK/MDR1: AB	Passive Permeability
Intrinsic Clearance (μ L/min/mg)	16.874	<8.000	12.099	25.766
cHLM	22.10	13.30	38.50	7.95
Solubility (μ M) pH 6.5	164	302	330	75
LLE	0.68	3.81	0.62	3.38
LE	0.25	0.51	0.20	0.31
No. Heavy Atoms	22	13	25	19
Estimated IC ₅₀ (μ M)	200	30	500	100
BASIC pK _a	1.66	1.36	5.64	1.16
ACID pK _a	n.a.	n.a.	4.50	11.86
TPSA ($\text{\AA}^2$)	58.95	83.80	60.94	63.57
LogD pH 7.4 (cLogD)	3.189 (3.34)	0.565 (0.70)	0.960 (0.35)	1.637 (1.25)
cLOGP	3.32	1.02	2.98	0.93
MW (Da)	292	176	334	256
Compound	c10 	c12 	c14 	c16 

Permeability was measured as the ratio of apical to basolateral and basolateral to apical in either Madin-Darby canine kidney (MDCK) cells (passive) or MDCK cells transfected with the human *MDR1* gene (gene encoding for the efflux protein, P-glycoprotein: P-gp). P_{appAB} indicates cell permeability in the gastrointestinal tract (P-gp expressing cells) and permeability of c4 and c12 was prohibitively low ($<0.1 \times 10^{-6}$ cm/sec). All other compounds demonstrated a $P_{appAB} > 3.5$ which may suggest central nervous system (CNS) permeability. An efflux ratio ($P_{appBA}/P_{appAB} \geq 2$) would indicate drug efflux is occurring, and was not observed for any compounds, hence none are substrates of P-gp

3.4.4 WaterLOGSY

Eight Maybridge fragments (c02, c04, c06, c09, c10, c12, c14, c16) which were not considered artefacts after the TruHit assay were further validated in a water-ligand observed *via* gradient spectroscopy (WaterLOGSY) assay. Nuclear magnetic resonance (NMR) -based ligand interaction studies often involve the nuclear Overhauser effect (NOE) to study transfer of spin polarisation between complexes. WaterLOGSY uses magnetization transfer from bulk water.²² A difference spectrum is generated from a spectrum recorded for a compound in the presence of a protein, and a blank spectrum of the compound alone involving magnetisation transfer to the ligand in solution only. Opposite signs for ligand signals obtained in protein-bound (where Overhauser effect has negative sign since NOE transfer is in the slow tumbling regime) versus free (where Overhauser effect has the opposite sign since NOE transfer is in fast tumbling regime) form are observed in a WaterLOGSY spectrum, enabling easy discrimination between binders and non-binders.

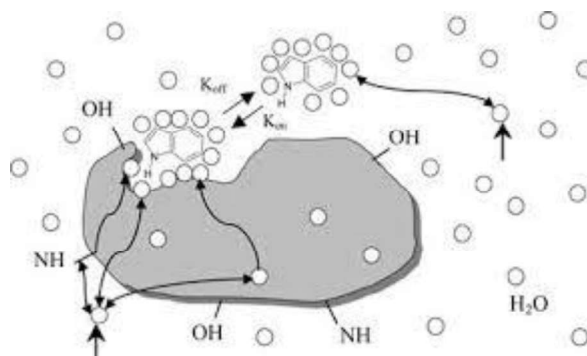


Figure 3.14 WaterLOGSY principle. Ligand is shown in the free and bound state, excitation of bulk water (circles) is followed by magnetisation transfer pathways shown as arrows.

To avoid obscuring regions of the spectra with organic buffer signals, protein samples were exchanged into sodium phosphate buffer (10 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, 150 mM NaCl, pH 7.5) prior to addition to samples.

Maybridge compounds had been stored as

50 mM solutions in DMSO, which was likely to dominate the ^1H NMR spectra, so a 1:100 dilution of compound into buffer was used. This would provide a sufficient concentration to detect binding (500 μM) but with a lower DMSO concentration (1%) to minimise solvent peaks in the spectra, and any competitive inhibition of DMSO in the KAc binding pocket. A concentration ratio of 25-fold excess of ligand was maintained by using protein concentrations of 20 μM .

Of the eight samples tested, four resulted in strong evidence of a binding interaction (Figure 3.15). Three demonstrated positive WaterLOGSY spectrum with TRIM24 (c04, c06, c09) and three with TRIM33 (c06, c09, c12). Although this technique can only provide evidence for a binding interaction, without quantification, compounds c04 and c12 show significantly more effect for TRIM24 and TRIM33 respectively, suggesting a level of selectivity. The spectrum for c06 suggested a mixture of species, with each one demonstrating distinct binding with TRIM24 or TRIM33; resynthesis of this compound would confirm the identity of the interacting species.

Data sets from several compounds were insufficient for conclusive validation through this technique (Figure 3.16). Compound c02 was insoluble in the NMR solution at 500 μM and spectra could not be obtained, however, low solubility suggests that the readout from the AlphaScreen assay may be primarily due to signal diffraction from precipitation rather than binding. Compounds which are prone to aggregation are often observed as positive WaterLOGSY signals in the absence of a protein, and are often deemed unsuitable for this technique. This was the case for c10 and c16, which resulted in increased WaterLOGSY signals or inversion of sign, respectively, in the presence of TRIM proteins. These would require validation *via* an alternative technique. WaterLOGSY spectra of compound c14 with TRIM24 or TRIM33 showed a reduction in signal from compound peaks, compared to the WaterLOGSY of the compound only control. This may be due to weak binding, since the spectrum observed is a population average of all states, but is inconclusive for validation.

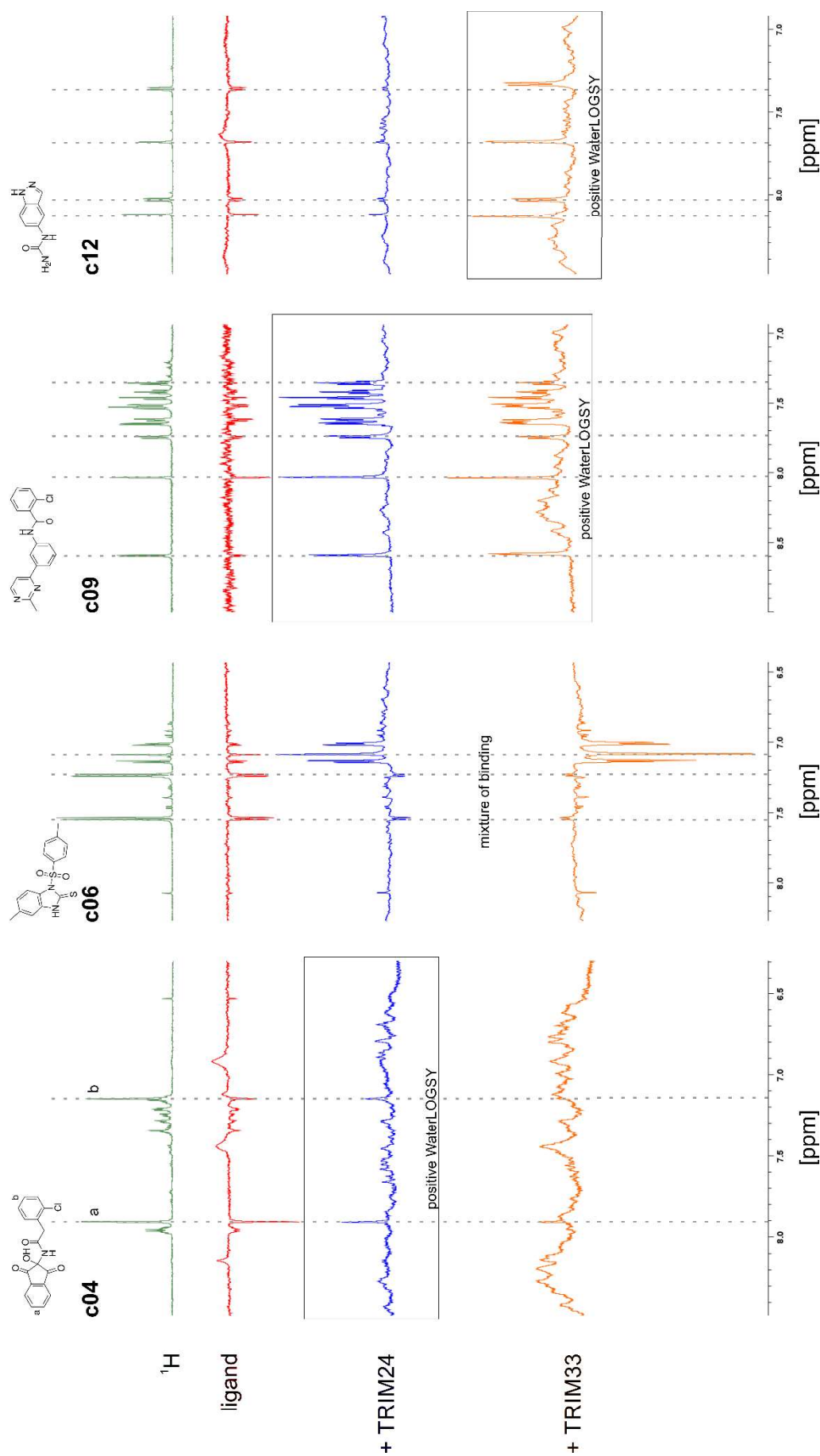


Figure 3.15 Partial waterLOGSY spectra from compounds demonstrating binding interactions with TRIM24 or TRIM33 (black box). ^1H spectra for ligand only shown for comparison (green), with WaterLOGSY of ligand blank (red), with TRIM24 (blue), or with TRIM33 (orange). Transfer of magnetisation in the slow tumbling regime results in positive NOE signal when ligand is bound to the protein. Inversion of peaks from the ligand only WaterLOGSY demonstrates binding interactions.

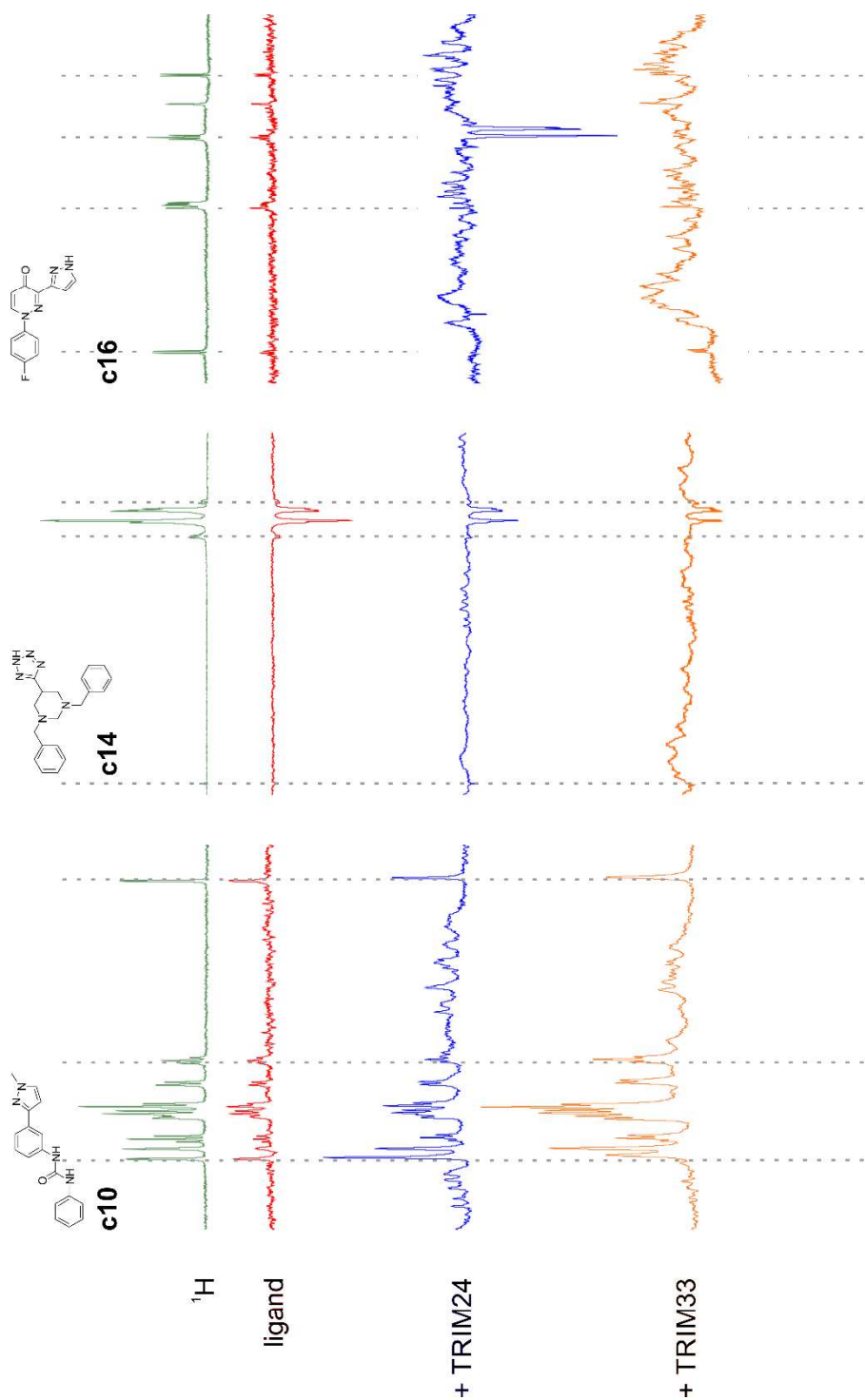


Figure 3.16 Inconclusive WaterLOGSY spectra from compounds which could not be validated using this technique. ^1H spectra for ligand only shown for comparison (green), with WaterLOGSY of ligand blank (red), or with TRIM33 (orange). Inversion of peaks from the ligand only WaterLOGSY demonstrates binding interactions. Compounds **c10** and **c16** demonstrated positive NOE in the ligand only spectra, complicating interpretation. Compound **c14** demonstrated a reduction in negative NOE, which could be due to weak binding, since the spectra observed is a population average of the bound and unbound states. These compounds require further validation *via* alternative techniques.

3.5 Conclusions

On commencement of this work there were no known binding partners for either TRIM24 or TRIM33. Through generation of an enriched computational library, and primary screening in the AlphaScreen format, we identified four small molecules within the Maybridge Diversity collection, which have demonstrated inhibitory activity in multiple orthogonal assays, including selectivity between TRIM24 and TRIM33, and are promising leads for probe development.

Fragment lead molecules for TRIM24 and TRIM33 probes

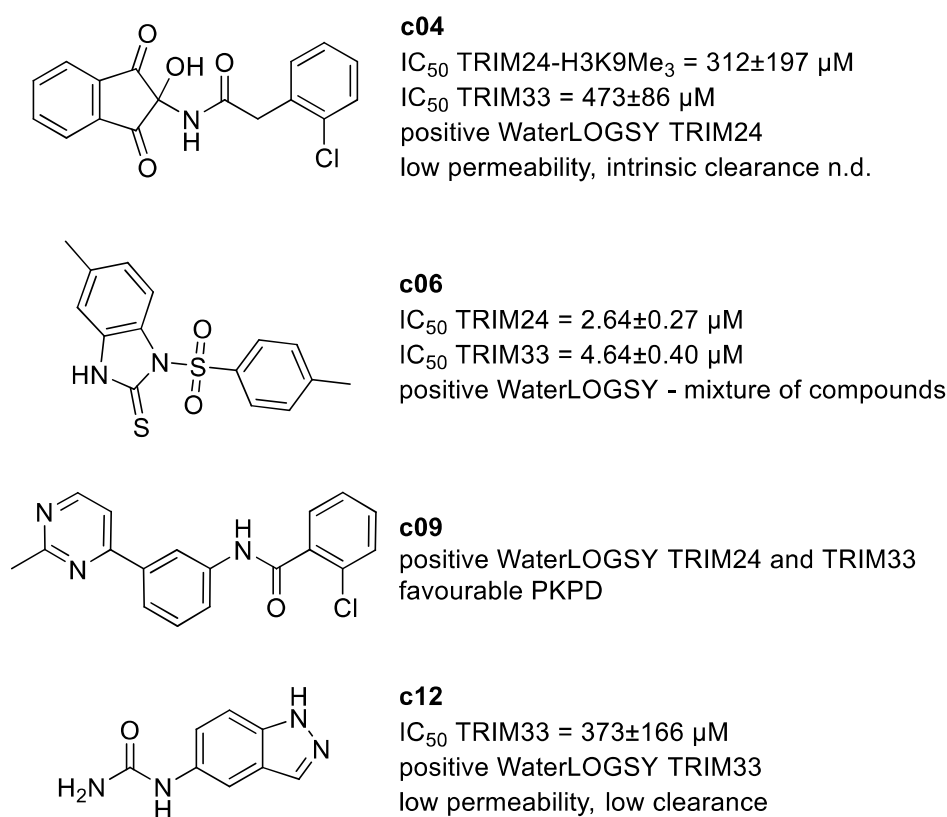


Figure 3.17 Summary of lead molecules identified in this study, for the development of probes inhibiting histone peptide recognition for TRIM24 or TRIM33. Unless competitor peptide partner is stated, IC₅₀ values are given as the geometric mean for H3₁₋₂₇K9Me₃, H3₁₋₂₇K18Ac, and H3₁₋₂₇K9Me₃K18Ac, competitor peptides. WaterLOGSY signals observed at 500 µM

The low molecular weight of these compounds allows scope for a 'fragment growth' approach, and diversifying SAR could capitalise on selectivity already demonstrated, resulting in orthogonal probes for both TRIM24 and TRIM33. Compound c04 inhibited histone peptide recognition for the PHD finger rather than the bromodomain of TRIM24, and also resulted in a strong positive WaterLOGSY signal for this protein. The highly polar nature of the molecule, containing six

heteroatoms in close proximity and a TPSA of 83.47 Å, could be responsible for low permeability. Assessment of the contribution to affinity from these ketones could lead to improvements in pharmacodynamics. Considering the physical properties evaluated, the most promising of these lead compounds seems to be c09, for which SAR studies could focus on inclusion of affinity or selectivity generating interactions, without requirements to improve physical properties for probe utility. Compound c12 demonstrated a surprising level of selectivity for inhibition of TRIM33 histone recognition, given its small size and minimal distinguishing features: elucidation of the mode of this selectivity could provide insight which could translate to other fragments.

Additionally, eight compounds were identified from the PPI-net collection as potential binding partners in the AlphaScreen assay which were not caused by assay interference. Resynthesis or purchase of these compounds would enable further binding studies to validate these as inhibitors of the tandem PHD and bromodomain of TRIM24 and TRIM33.

Orthogonality of the AlphaScreen assay for the bromodomain and PHD finger may only be possible for higher affinity binding partners. Contributions of each domain to the binding affinity may exceed the selectivity window, and relative inhibition levels would not represent binding regions proportionally. Use of the inactive mutant constructs, designed and produced in Chapter 2, would provide further insight into the location of inhibition, across either the PHD finger or the bromodomain.

In order to continue to investigate these ligands as inhibitors of peptide recognition of TRIM24 and TRIM33, structural information would be of great value. Obtaining a protein bound crystal structure would not only validate the compounds as binding, but also indicate which surface of the tandem domain they bind to, and provide information for structure guided optimisation. In the absence of further structural information, optimisation of lead compounds towards probes could be designed around docking predictions for these molecules.

3.6 References

- (1) Palmer, W. S., Poncet-Montange, G., Liu, G., Petrocchi, A., Reyna, N., Subramanian, G., Theroff, J., Yau, A., Kost-Alimova, M., Bardenhagen, J. P., Leo, E., Shepard, H. E., Tieu, T. N., Shi, X., Zhan, Y., Zhao, S., Draetta, G., Toniatti, C., Jones, P., Geck Do, M., and Andersen, J. N. (2015) Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J. Med. Chem.* 10.1021/acs.jmedchem.5b00405.
- (2) Bennett, J. M., Fedorov, O., Tallant, C., Monteiro, O. P., Meier, J., Gamble, V., Savitski, P., Nunez-Alonso, G. A., Haendler, B., Rogers, C., Brennan, P. E., Müller, S., and Knapp, S. (2015) Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *J. Med. Chem.* 10.1021/acs.jmedchem.5b00458.
- (3) Santiago, C., Nguyen, K., and Schapira, M. (2011) Druggability of methyl-lysine binding sites. *J. Comput. Aided. Mol. Des.* 25, 1171–1178.
- (4) Herold, J. M., Wigle, T. J., Norris, J. L., Lam, R., Korboukh, V. K., Gao, C., Ingeman, L. A., Kireev, D. B., Senisterra, G., Vedadi, M., Tripathy, A., Brown, P. J., Arrowsmith, C. H., Jin, J., Janzen, W. P., and Frye, S. V. (2011) Small-Molecule Ligands of Methyl-Lysine Binding Proteins. *J. Med. Chem.* 54, 2504–2511.
- (5) Vidler, L. R., Brown, N., Knapp, S., and Hoelder, S. (2012) Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J. Med. Chem.* 55, 7346–59.
- (6) Vidler, L. R., Filippakopoulos, P., Fedorov, O., Picaud, S., Martin, S., Tomsett, M., Woodward, H., Brown, N., Knapp, S., and Hoelder, S. (2013) Discovery of novel small-molecule inhibitors of BRD4 using structure-based virtual screening. *J. Med. Chem.* 56, 8073–88.
- (7) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26.
- (8) Trott, O., and Olson, A. J. (2010) AutoDock Vina : Improving the Speed and Accuracy of Docking with a New Scoring Function , Efficient Optimization , and Multithreading. *J Comput Chem* 31, 455–461.
- (9) Zhao, H., Gartenmann, L., Dong, J., Spiliotopoulos, D., and Caflisch, A. (2014) Discovery of BRD4 bromodomain inhibitors by fragment-based high-throughput docking. *Bioorg. Med. Chem. Lett.* 24, 2493–6.
- (10) Xu, M., Unzue, A., Dong, J., Spiliotopoulos, D., Nevado, C., and Caflisch, A. (2015) Discovery of CREBBP Bromodomain Inhibitors by High-Throughput Docking and Hit Optimization Guided by Molecular Dynamics. *J. Med. Chem.* 150715115225002.
- (11) Lewis, A., and Overton, M. (2009) Nonsmooth optimization via BFGS. *Submitt. to SIAM J. Optim.* 1–35.
- (12) Ross, G. A., Morris, G. M., and Biggin, P. C. (2012) Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* 7, e32036.

- (13) Abagyan, R., Totrov, M., and Kuznetsov, D. (1993) ICM - A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J. Comput. Chem.* 15, 488–506.
- (14) Laskowski, R. A., and Swindells, M. B. (2011) LigPlot + : Multiple Ligand Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* 2778–2786.
- (15) Eglen, R. M., Reisine, T., Roby, P., Rouleau, N., Illy, C., Bossé, R., and Bielefeld, M. (2008) The use of AlphaScreen technology in HTS: current status. *Curr. Chem. Genomics* 1, 2–10.
- (16) Fish, P. V., Filippakopoulos, P., Bish, G., Brennan, P. E., Bunnage, M. E., Cook, A. S., Federov, O., Gerstenberger, B. S., Jones, H., Knapp, S., Marsden, B., Nocka, K., Owen, D. R., Philpott, M., Picaud, S., Primiano, M. J., Ralph, M. J., Sciammetta, N., and Trzupek, J. D. (2012) Identification of a chemical probe for bromo and extra C-terminal bromodomain inhibition through optimization of a fragment-derived hit. *J. Med. Chem.* 55, 9831–7.
- (17) Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B., and Denu, J. M. (2012) Identification and Characterization of Small Molecule Inhibitors of a Plant Homeodomain Finger. *Biochemistry* 51, 8293–8306.
- (18) Lolli, G., and Battistutta, R. (2013) Different orientations of low-molecular-weight fragments in the binding pocket of a BRD4 bromodomain. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 69, 2161–2164.
- (19) Ertl, P., Rohde, B., and Selzer, P. (2000) Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* 43, 3714–3717.
- (20) Hopkins, A. L., Groom, C. R., and Alex, A. (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431.
- (21) Leeson, P. D., and Springthorpe, B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6, 881–890.
- (22) Dalvit, C., Pevarello, P., Tato, M., Veronesi, M., Vulpetti, a., and Sundstrom, M. (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* 18, 65–68.

Chapter 4: Structural Studies

4.1 Introduction

In the work conducted towards small molecule probes for the tandem domains of TRIM24 and TRIM33 in Chapter 3, several assays were used for hit identification and validation. A combination of biophysical techniques are helpful in providing comprehensive binding information about a lead candidate. However, structural information is of great value in guiding the design of optimised ligands.¹ Within the epigenetic field, probe development from initial lead molecules has been expedited towards several targets, through the use of structure-guided optimisation.²⁻⁴

4.1.1 Structural coverage of bromodomain family

Historically, drug discovery efforts have used protein crystallography to aid compound development. Recent technological advances have greatly increased access to structural data for target proteins, with higher throughput of screening,⁵ and a greater knowledge of crystallisation protocols. Visualisation of how ligands bind to a target protein allows rational design of optimised compounds, rather than relying on unguided SAR studies. Ligand-protein co-structures also contribute to validation of target engagement for compounds which proceed to clinic.

In humans, epigenetic reader modules occur as part of complex protein architectures, often containing several other domains with known functions. For example, in addition to the well characterised bromodomain, CREBBP also contains a histone acetyltransferase (HAT) domain and a PHD finger motif vital for activity,⁶ as well as several modules involved in transcriptional activation: nuclear receptor interaction domain (RID), kinase-inducible domain interacting (KIX), cysteine/histidine regions (TAZ1/TAZ2), and an interferon response binding domain (IBiD). Despite there being good structural coverage of the bromodomain family,⁷ the majority of these are structures of the isolated bromodomain, and few structures of multidomain epigenetic readers exist in the public sphere (Table 4.1). Most small-molecule ligands bound in these structures do so in the KAc pocket of the bromodomain; in some cases, an adjacent PHD finger has been included in

expression constructs to shield the hydrophobic surface exposed on the bromodomain when a PHD finger is excluded. The structures of isolated PHD fingers from bromodomain containing proteins are also available for BAZ2A, BAZ2B,⁸ and BRD1.⁹

Table 4.1 Structures deposited in the PDB of bromodomain containing proteins including additional epigenetic reader domains. Although structural information exists for a large proportion of the bromodomain family, few of these contain adjacent domains, which often also have epigenetic functions.

PDB ID(s)	Bromodomain containing protein	Domains	References
2F6J, 2F6N, 2FSA, 2FUI	bromodomain PHD finger transcription factor (BPTF)	Bromodomain PHD finger	Li <i>et al.</i> 2006 ¹⁰
3O33, 3O34, 3O35, 3O36, 3O37, 4YBM, 4YAB, 4YAD, 4YAT, 4YAX, 4YC9, 4YBS, 4YBT, 4ZQL	tripartite motif-containing 24 (TRIM24)	Bromodomain PHD finger	Tsai <i>et al.</i> 2010 ¹¹ Palmer <i>et al.</i> 2015 ³ Bennett <i>et al.</i> 2015 ¹²
3U5M, 3U5N, 3U5O, 3U5P	tripartite motif-containing 33 (TRIM33)	Bromodomain PHD finger	Xi <i>et al.</i> 2011 ¹³
4BHW	histone acetyltransferase p300 (EP300)	Bromodomain CH2 region (PHD domain interrupted by RING domain) HAT domain	Delvecchio <i>et al.</i> 2013 ¹⁴
3UV5	TATA box binding protein-associated factor (TAF1)	Tandem bromodomains	Filippakopoulos <i>et al.</i> 2012 ⁷
3LQH, 3LQI, 3LQJ	mixed lineage leukemia 1 (MLL1)	Bromodomain PHD finger (third)	Wang <i>et al.</i> 2010 ¹⁵
2RO1	tripartite motif-containing 28 (TRIM28)	Bromodomain PHD finger	Zeng <i>et al.</i> 2008 ¹⁶

At the start of this work, crystal structures were available of TRIM24 and TRIM33 in *apo*- form, or with peptides bound (see Table 4.1 for PDB codes). These peptides possess a range of lysine modifications, but in the case of TRIM24 no single peptide spanned both domains. With an overall RMSD of 2.70 Å between the *apo*- structures of TRIM24 and TRIM33 (Figure 4.1), there is much more structural similarity between their individual PHD fingers (RMSDs of 0.66 Å over 56 residues) than their bromodomains (RMSD of 2.88 Å over 104 residues). Differences in apparent distances between recognition sites may result in substrate binding specificity, with a spacing of 26.4 Å between KAc and PHD binding sites in TRIM24 compared to 23.6 Å in TRIM33.

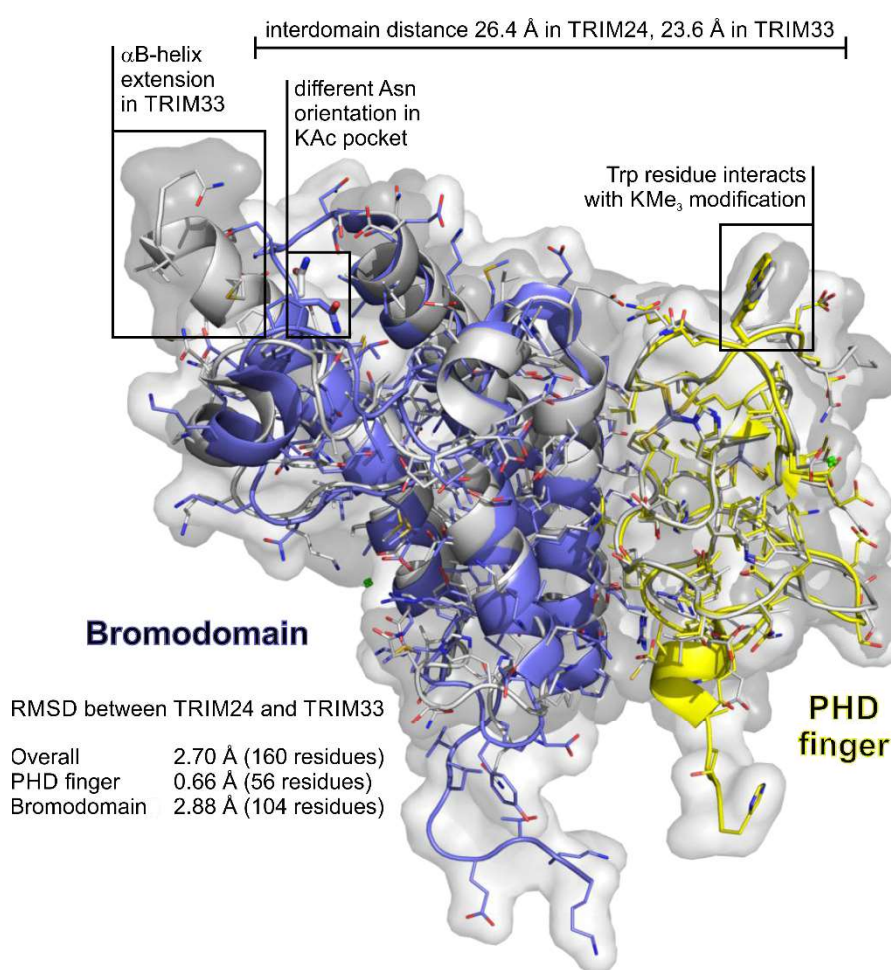


Figure 4.1 Overlay of TRIM24 (PDB: 3O33) bromodomain (blue) and PHD finger (yellow) with TRIM33 (grey – PDB: 3U5M). Structures aligned in PyMOL with cealign function. Notable differences include the interdomain spacing (26.4 Å - TRIM24; 23.6 Å - TRIM33), extension of the αB helix (top left helix), and positioning of the KAc binding asparagine (top left stick representation). The tryptophan which interacts with KMe₃ is present in both, with very high similarity within the PHD finger (RMSD 0.66 Å). Surfaces shown in grey (light - TRIM24; dark - TRIM33).

4.1.2 Aims

The aims of this study were to:

- Express and purify unlabelled TRIM24 and TRIM33 for crystallisation.
- Collect and resolve high resolution *apo*-structures.
- Co-crystallise TRIM proteins with peptides spanning both domains, to further elucidate binding modes and selectivity.
- Obtain small molecule crystal structures from previously identified hits for structure-guided optimisation.

4.2 Expression and purification

In order to obtain high quality crystal structures, unlabelled TRIM proteins were required for crystallisation. Successful crystallisation would depend on protein isolation with very high levels of purity, and structural homogeneity.

4.2.1 Cleavage from N-terminal His₆ vector

TRIM24 and TRIM33 proteins had been expressed from a pNIC28-Bsa4 vector (Appendix 8.2), with an *N*-terminal polyhistidine tag (His₆) for purification, which was left intact for assays. These terminal residues made it possible to immobilise the protein for use in labelled assays, for example in AlphaScreen using a nickel chelating acceptor bead. Proteins expressed from this vector also contain a sequence recognised by the Tobacco Etch Virus (TEV) cysteine protease (ENLYFQ|S) which allows cleavage of this portion of the protein,¹⁷ providing isolated tandem domains for use in label free studies, such as DSF and tryptophan fluorescence studies, or crystallisation.

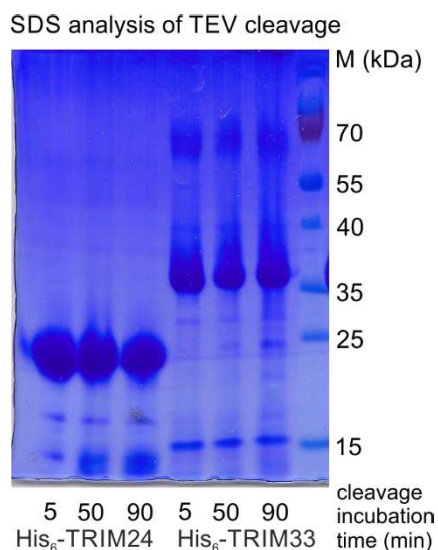


Figure 4.2 SDS-PAGE gels from TEV cleavage of TRIM24 and TRIM33 in pNIC28-Bsa4 expression vector. Poor cleavage was observed, longer time course (not shown) resulted in incomplete cleavage after incubation overnight.

In order to isolate unlabelled proteins, cleavage was attempted using TEV under a range of conditions, however, cleavage never reached completion, even after overnight incubation at 4 °C (Figure 4.2). Additionally, isolation of cleaved TRIM24 and TRIM33 revealed inherent instability in the tandem PHD-bromodomain proteins, which suggested the PHD finger was unstable towards cleavage in this manner. Analysis of the melting temperature (T_m) under a range of buffer conditions showed little variation, which implicates internal stability as the limiting factor for protein stability, rather than the buffer conditions that the protein was subjected to (Figure 4.3).

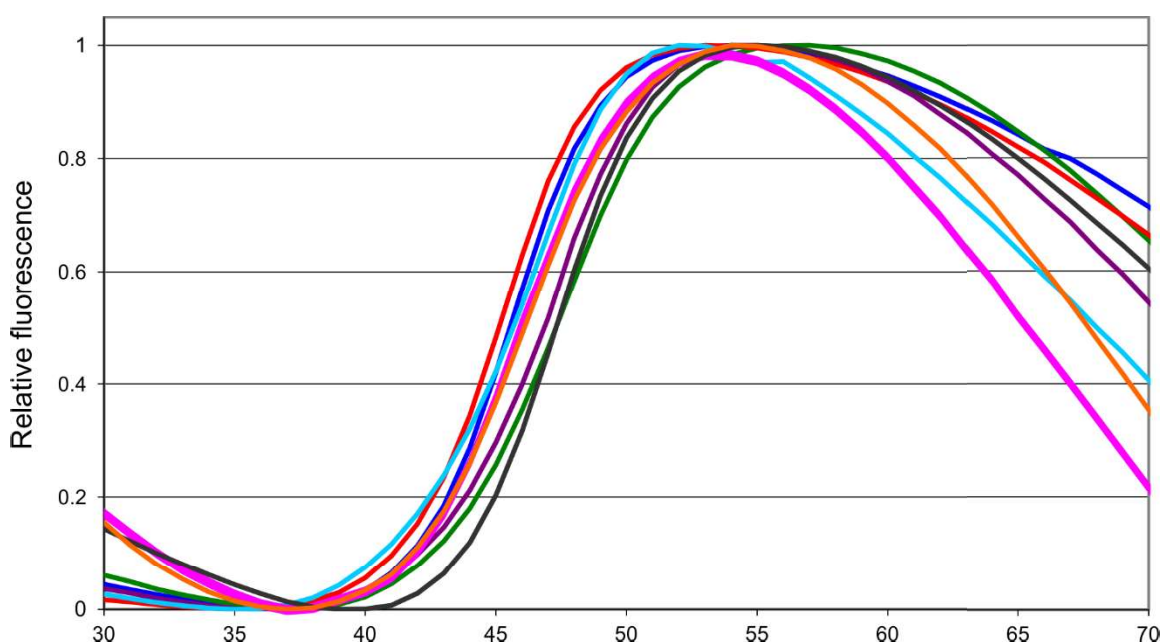


Figure 4.3 Normalised melting curves from TRIM24 isolated after pNIC28-Bsa4 expression (TEV cleavage of His₆-tag) in various buffer conditions. The buffer conditions ranged from pH 5.5-10.0, little variation in melting curves suggests that internal stability of protein is limiting factor, rather than conditions.

4.2.2 N-terminal GST vector

Attempts were made to clone TRIM genes into a complementary vector that would facilitate cleavage in a stable form for crystallisation. It was hypothesised that stabilisation of the minimal structure in the PHD finger with a larger expression label would allow subsequent cleavage without disruption of function. The extra domain could act as a chaperone to facilitate folding of the PHD finger during expression, stabilise the protein during purification, and increase protein stability after cleavage. pGTvL1-SGC (SGC, Oxford) is a pGEX expression vector containing an N-terminal glutathione *S*-transferase (GST) tag and a sequence recognised by TEV protease. The large 26 kDa GST is

frequently used to increase solubility of expressed proteins,¹⁸ and cleavage of affinity labels from these constructs would provide isolated tandem PHD-bromodomains for use in crystallisation.

TRIM24 and *TRIM33* genes were cloned into the pGTvL1-SGC vector (Appendix 8.2) using LIC cloning. Amplification of gene fragments, with upper and lower LIC sequences specified for the vector, was carried out by PCR (Figure 4.4), and the linear vector was obtained by treatment with BseRI. Cohesive end generation using T4 DNA polymerase chew-back action in presence of dGTP, for both the fragments and the linear vector, was followed by direct transformation into XL10 (Gold) competent cells.

High levels of overexpression of the protein of interest were achieved using the same conditions for growth and expression as with the previous pNIC28-Bsa4 vector. Correct folding of the GST tag was confirmed by purification with glutathione sepharose chromatography (GE Healthcare), and further purification by size exclusion chromatography (Sephadex gel S75, GE Healthcare) resulted in ~3 mg of pure protein per litre of cell culture (Figure 4.5).

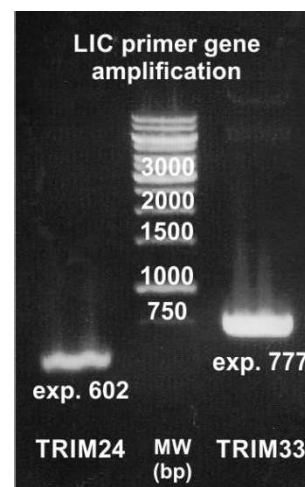


Figure 4.4 Agarose gel purification from gene amplification of *TRIM24* and *TRIM33* using LIC primers.

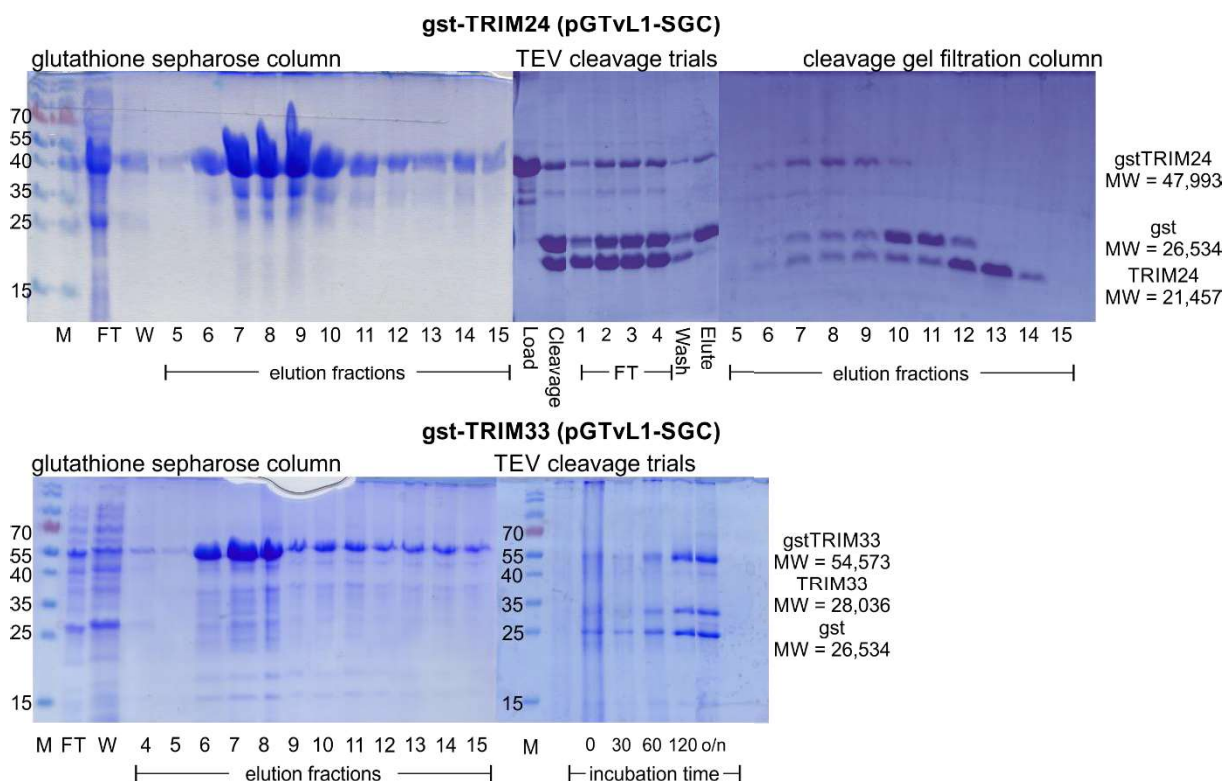


Figure 4.5 SDS-PAGE gels from purification of TRIM24 (top) and TRIM33 (bottom) after expression using pGTvL1-SGC vectors. Glutathione sepharose column purification was followed by TEV cleavage of GST- tag, and further IMAC column (to remove His₆- tagged TEV) and glutathione column (to remove uncleaved gst-TRIM and GST fragment). Cleaved TRIM24 was isolated after size exclusion chromatography (fractions 13-14); yields of cleaved TRIM33 were too low to isolate.

Initial trials of cleavage from GST labelled TRIM24 and TRIM33, with β -mercaptoethanol reducing conditions suggested for use with zinc finger proteins, resulted predominantly in precipitation and aggregation. Adjustment to milder conditions at 4 °C were insufficient for complete cleavage of His₆ labelled proteins, and yields of cleaved TRIM33 were prohibitively low for purification. Complications in cleavage were thought to be due to the conformational shock from loss of such a large chaperoning group. However, protein isolated after cleavage did demonstrate varying stability across a range of buffer conditions, suggesting inherent stability of protein fold was no longer a limiting factor (Figure 4.6).

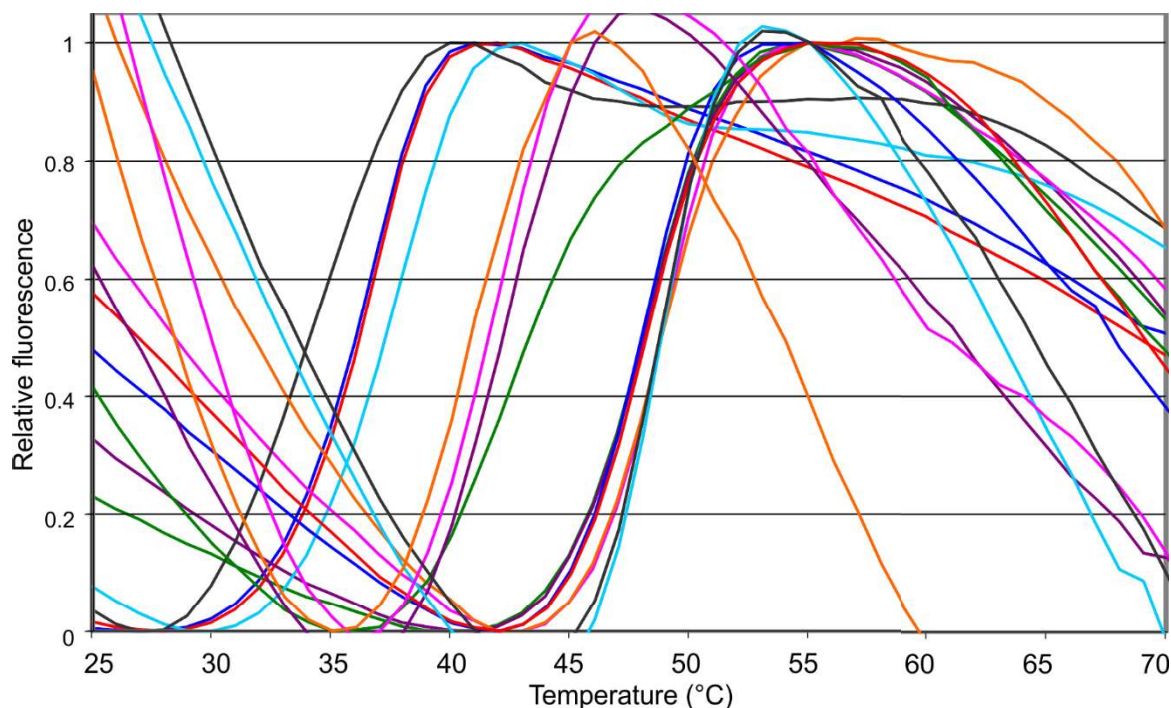


Figure 4.6 Normalised melting curves from TRIM24 isolated after pGTVL1-SGC expression (TEV cleavage of GST- tag) in various buffer conditions. Buffers ranged from pH 5.5-10.0: strong effect of conditions on melting temperature suggests internal stability of protein is not a limiting factor. Protein resulting from GST- cleavage appears more stable than from pNIC28-Bsa4 vector.

4.2.3 Cloning into N-terminal SUMO vector

An alternative strategy was pursued, namely to express proteins in combination with a Small Ubiquitin-like Modifier (SUMO) group. Ubiquitin-based tags have demonstrated utility in aiding protein folding¹⁸ and could provide a better compromise between sufficient chaperoning ability to facilitate folding, without causing disruption to structure upon cleavage.

Similar to previous vectors used, the pETM11-SUMO3GFP vector (EMBL, Heidelberg Germany) is also under the control of an inducible T7 promoter based system, and includes an N-terminal His₆ and SUMO label. Attempts to clone *TRIM24* and *TRIM33* into this vector using both LIC and sequence and ligation independent cloning (SLIC) were unsuccessful.

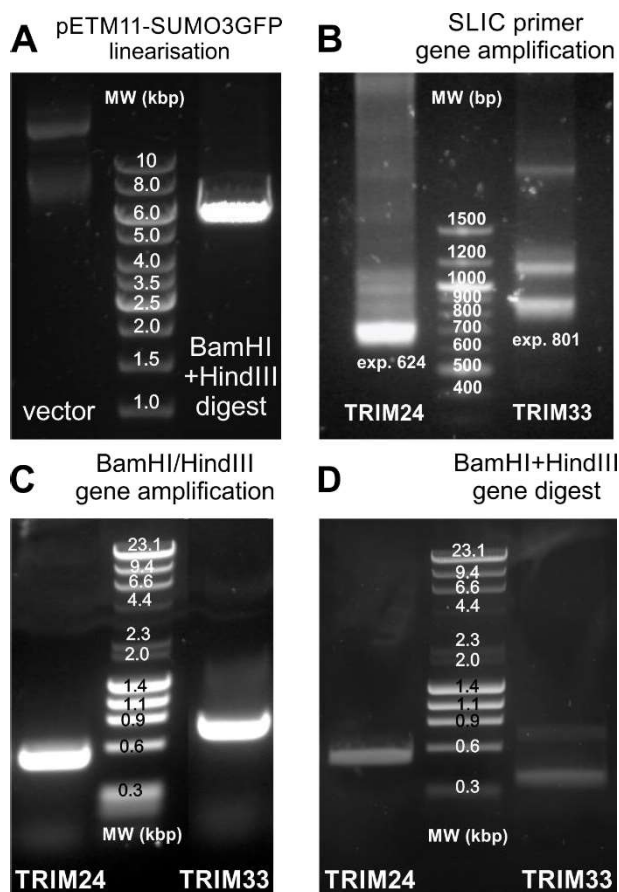


Figure 4.7 Agarose gel purifications. **A:** restriction enzyme digest of pETM11-SUMO3GFP vector using BamHI and HindIII (37 °C, 4 h). **B:** gene amplification of *TRIM24* and *TRIM33* using extended SLIC primers. **C:** gene amplification of *TRIM24* and *TRIM33* using restriction complementary restriction enzyme sites. **D:** isolation of *TRIM24* and *TRIM33* after digest using BamHI and HindIII demonstrating internal cut in *TRIM33* gene

Linearisation of the pETM11-SUMO3GFP using BamHI and HindIII excised the 720 bp GFP fragment from the vector (Figure 4.7:A). After cohesive end generation, transformation of plasmids formed from incubation with gene fragments amplified from LIC and SLIC primers designed according to standard guidelines were unsuccessful (see Materials and Methods for primer details). PCR amplification of genes using extended SLIC primer sequences with 30 bp homology to the vector, in order to increase the likelihood correct annealing, gave a mixture of products (Figure 4.7:B). Expected gene products were 624 bp and 801 bp for *TRIM24* and *TRIM33*, respectively, however excision in this region

and purification from agarose did not lead to successful cloning. Attempts to ligate the genes into the linear pETM11-SUMO3GFP vector using complementary restriction enzymes were also unsuccessful: *TRIM33* contained an internal HindIII recognition site, and ligation with *TRIM24* after treatment with BamHI and HindIII did not produce any colonies after transformation (Figure 4.7:D).

Alternatively, cloning of the *SUMO* gene into the pNIC28-Bsa4 vector was attempted, with insertion between the His₆ label and the *TRIM* gene (Figure 4.8). Linearisation of the vector with NdeI and NcoI, and gene amplification *via* PCR with corresponding primers was successful. However, preparation of genes for ligation resulted in additional cuts within the gene, due to the

existence of an NcoI recognition site after the His₆ label, in addition to the multiple cloning site (MCS).

Successful cloning was achieved using a Gibson Assembly protocol consisting of a one-step isothermal *in vitro* recombination.¹⁹ The pETM11-SUMO3GFP vector (Appendix 8.2) was linearised with BamHI and HindIII, and genes amplified by PCR with a 30 bp homology region, before incubation with a Gibson Assembly Master Mix (NEB) including optimised proportions of T5 exonuclease, Phusion DNA polymerase and Taq DNA

ligase, for chew-back, annealing, and ligation respectively. Transformation efficiency of this reaction mixture was much higher with the use of commercial NEB 5-alpha (#C2987) competent cells, and isolated plasmids were subsequently used to transform BL21 (Gold) DE3 for expression.

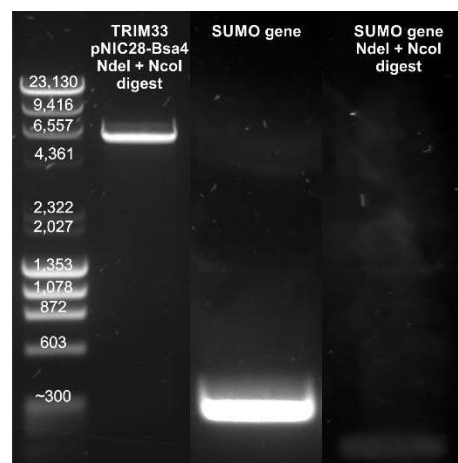


Figure 4.8 Agarose gel purification from SUMO cloning using NdeI and NcoI restriction enzyme sites. Successful linearisation of vector, and PCR amplification of gene from pETM11-SUMO3GFP vector. Internal NcoI recognition site in SUMO gene exposed during NdeI and NcoI digest.

4.2.4 Purification and cleavage for crystallisation

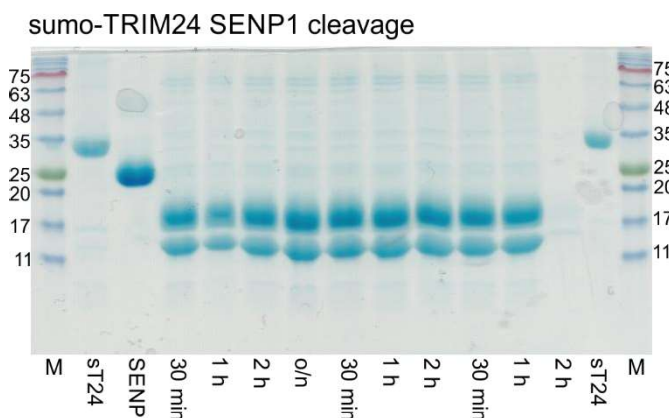


Figure 4.9 SDS-PAGE gel from SENP1 cleavage trials of sumo-TRIM24 (sT24) expressed from pETM11-SUMO3 vector. M – molecular ladder (kDa). Expected MW: sT24, 33.8 kDa; SENP1, 29.2 kDa; TRIM24, 21.5 kDa; His₆SUMO, 12.4 kDa.

Sentrin-specific protease 1 (SENP1) is a human protease which recognises the C-terminal section of a SUMO modification and causes hydrolysis of a peptide bond after a conserved di-glycine motif.²⁰

Initial trials of SENP1 cleavage of the SUMO label were conducted using

protein supplied by the SGC (Figure 4.9). Compared to previous use of TEV protease, a great improvement was observed, with cleavage appearing to reach completion within 30 min at 4 °C. SENP1 also demonstrated high selectivity, with no detectable off-site cleavage after 2 h at 25 °C,

or even at 37 °C. Very low protein concentrations were observed in samples from 2 h at 37 °C, which was thought to be due to protein instability and precipitation at these elevated temperatures.

In addition to small quantities of purified protein, the SGC supplied SENP1 plasmids, to facilitate continued production and scale up of cleavage. Plasmids containing this gene were extracted from a small scale growth of transformed *E. coli* cells, sequenced to confirm identity, and used to transform BL21 (Gold) DE3 cells for expression. Purification by metal affinity chromatography and size exclusion chromatography produced high yields of 5.2 mg purified SENP1 per gram of whole cells. Cleavage efficiency of in-house prepared SENP1, was comparable with that provided by the SGC (Figure 4.10), and cleavage was complete within 60 minutes at 4 °C. Additionally, the cleavage process may assist in separating TRIM33 oligomers observed previously, since bands visible at the interface between the stacking and resolving gel (>80 kDa), decreased in intensity as cleavage progressed.

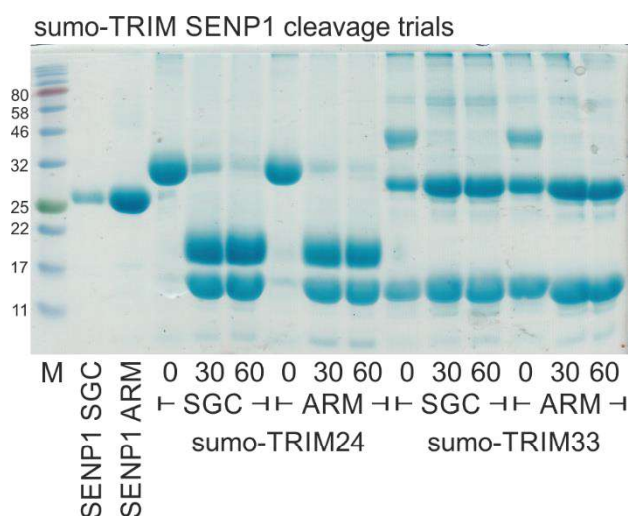


Figure 4.10 SDS-PAGE gel analysis of cleavage efficiency of SENP1 supplied by the structural genomics consortium (SGC) or expressed and purified (ARM). Incubation time shown in minutes. Complete cleavage of both TRIM24 and TRIM33 was achieved by both proteases in 60 minutes at 4 °C. Partial cleavage of TRIM33 at t=0 due to residual SENP1 on IMAC column during purification; oligomeric forms of TRIM33 seem reduced during cleavage.

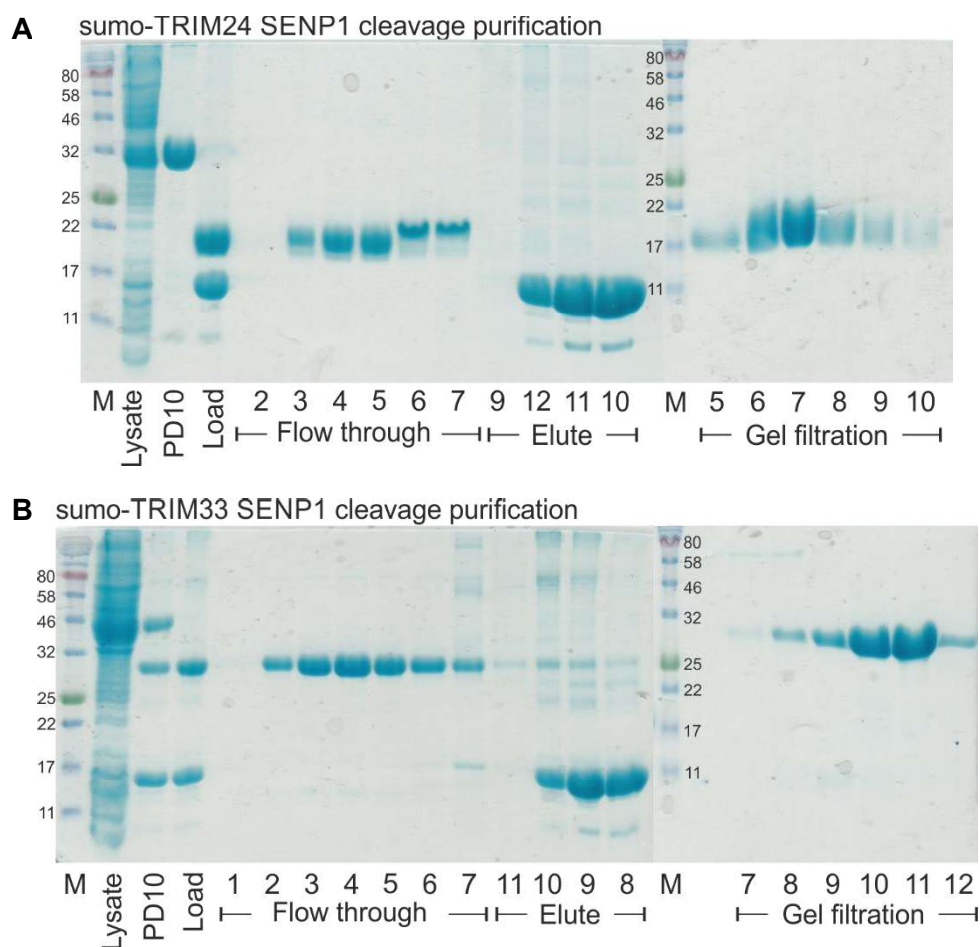


Figure 4.11 SDS-PAGE analysis of purification of unlabelled TRIM proteins from pETM11-SUMO3 vector. A – TRIM24; B – TRIM33. M – molecular ladder, weight shown in kDa. PD10 – eluent from buffer exchange using PD10 column (GE Healthcare) of fractions containing protein from IMAC purification (1 mL HP HisTrap, GE Healthcare). Load – sample after SENP1 cleavage for 1 h at 4 °C. Numbers indicate separately collected fractions. Expected molecular weights (kDa): sumo-TRIM24, 33.8; TRIM24, 21.5; sumo-TRIM33, 40.4; TRIM33, 28.0; His₆SUMO, 12.4; SENP1, 29.2.

Both the His₆-labelled SENP1 protein, and the cleaved His₆SUMO fragment from the TRIM proteins, could be removed through a further IMAC purification: unlabelled TRIM24 or TRIM33 protein would not interact with the column, and could be isolated in the flow through (Figure 4.11). Separation of any oligomeric forms was then performed using size exclusion chromatography (Sephadex gel S75, GE Healthcare) to ensure homogeneous protein for crystallisation. Melting temperatures of the cleaved TRIM24 and TRIM33 varied depending on buffer conditions, suggesting improved internal stability compared to cleavage from the pNIC28-Bsa4 vector (Figure 4.12). Overall, this procedure consistently resulted in >1 mg of purified TRIM protein per gram of whole cells from expression.

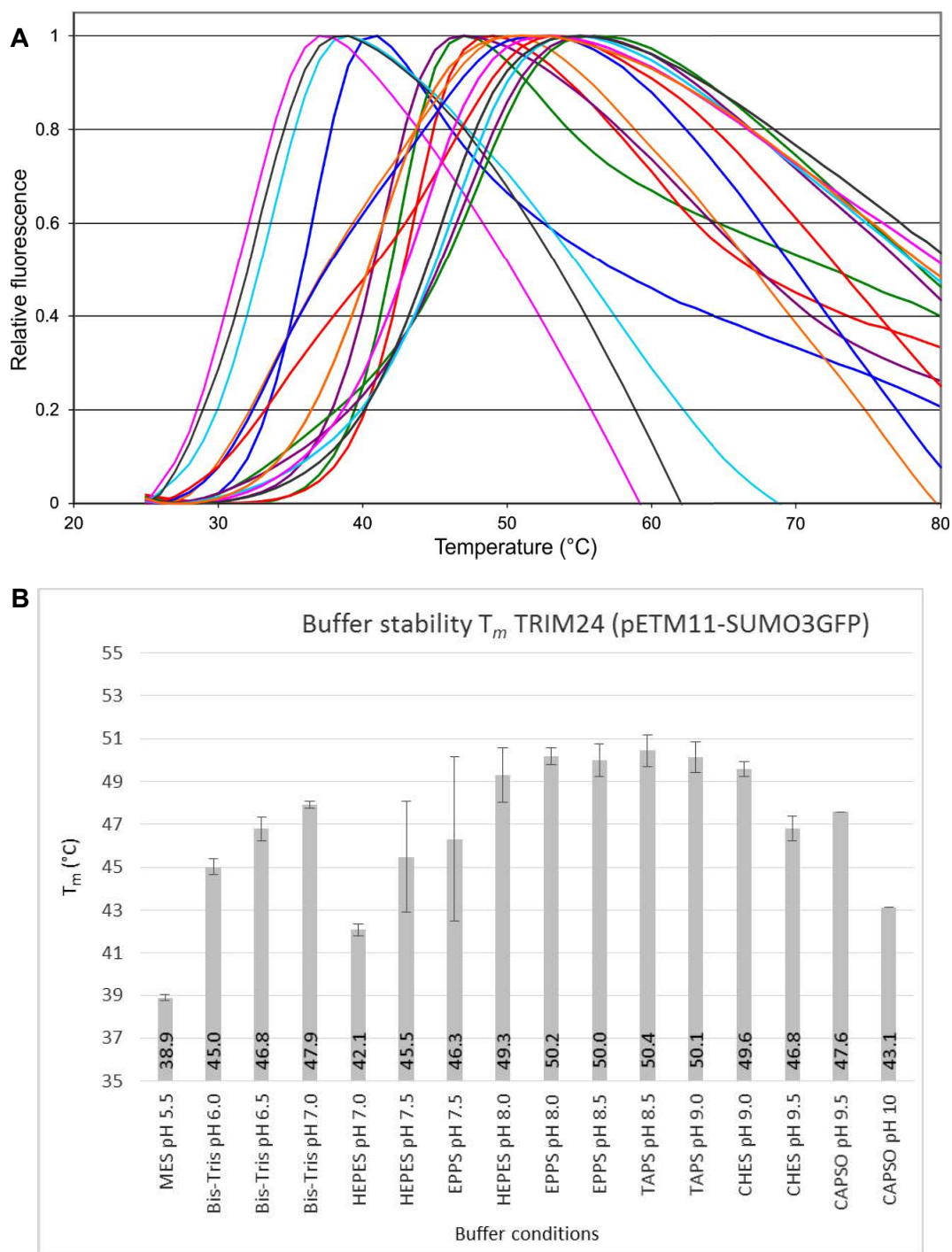


Figure 4.12 A: Normalised melting curves from TRIM24 isolated after pETM11-SUMO3GFP expression (SENPI cleavage of His₆SUMO- tag) in various buffer conditions. Buffers ranged from pH 5.5-10.0: clear effect of conditions on melting temperature suggests internal stability of protein is not a limiting factor. **B:** T_m values from Boltzmann fitting of melting curves of TRIM24 PHD-bromodomain after pETM11-SUMO3GFP expression (SENPI cleavage of His₆SUMO- tag) in various buffer conditions.

The secondary structure of proteins cleaved from various constructs was investigated by circular dichroism (CD) spectroscopy, using His₆-labelled proteins for comparison, assuming the extra C-terminal histidine residues do not greatly affect secondary structure (Figure 4.14). All constructs showed a high proportion of α -helical structure, as would be expected from the bromodomain, but of interest were discrepancies at lower wavelengths. Although for both TRIM24 and TRIM33, cleavage from the SUMO label (green) seemed to result in a lower helical content, positive dichroism at lower wavelengths was better maintained, which

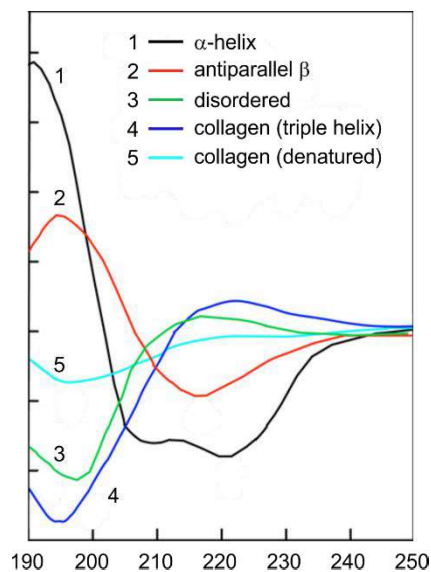


Figure 4.13 Circular dichroism (CD) spectra of polypeptides and proteins with representative secondary structures. Adapted from Greenfield 2006.²³

may be related to beta sheet formation (Figure 4.13). Since the sequences of these proteins are identical, once cleaved from their expression label systems, these data highlight the differences in structure resulting from cleavage, demonstrating the importance of chaperoned folding and stabilisation of the PHD finger.

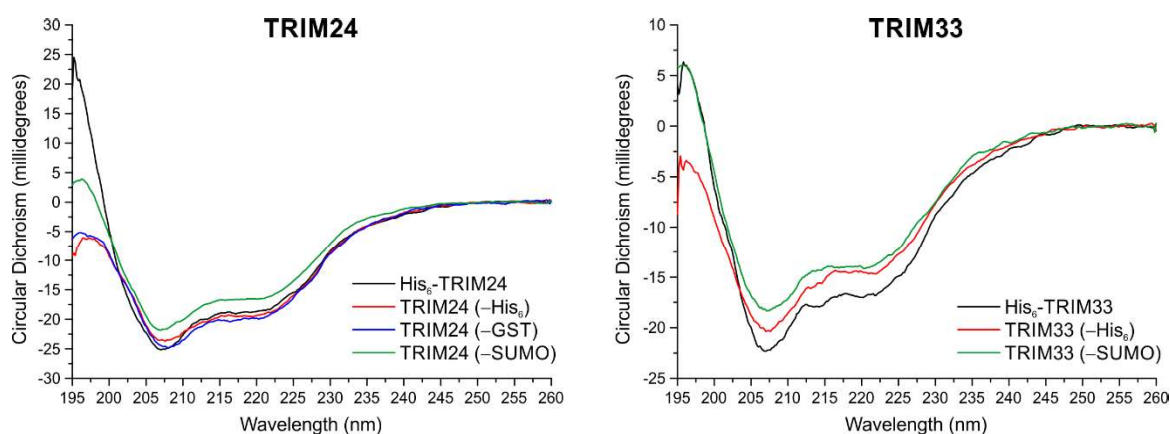


Figure 4.14 Circular Dichroism spectra from TRIM24 (left) and TRIM33 (right) cleaved from various constructs, in comparison to polyhistidine labelled protein (His₆-). Smoothed average of results in triplicate with 0.2 nm step size, 0.5 sec/step, 10 mm path length, 0.2 mg/mL protein in NaF buffered phosphate buffer (pH 7.8). -His₆: TEV cleavage from pNIC28-Bsa4. -GST: TEV cleavage from pGTvL1-SGC. -SUMO: SENP1 cleavage from pETM11-SUMO3.

4.3 Crystallisation of TRIM24

With a reliable method for producing high yields of structurally homogeneous, unlabelled TRIM24, methods for crystallisation were then considered. Crystallisation buffers were manually compiled into screening selections based around literature conditions,¹¹ or in the JCSG+ format.²¹ Protein was concentrated to ~20 mg/mL immediately after purification, and plated using a Phoenix Liquid Handling System (Art Robbins Instruments) onto Intelli-Plate 96-3 low-profile crystallisation plates (Hampton Research) in three ratios with precipitation buffer.

Intelli-Plate 96-3 low-profile crystallisation plates							
[(NH ₄) ₂ SO ₄] (M)		1.7	1.8	1.9	2	2.1	2.2
		1	2	3	4	5	6
0.5	A	starting screening conditions					
1.0	B	100/200 mM HEPES					
1.5	C				narrow screening conditions 100 mM HEPES		
2.0	D						
2.5	E						
3.0	F						
3.5	G						
4.0	H						
% PEG400						crystals observed	

Figure 4.15 Crystallisation condition screening for TRIM24. Varying concentrations of ammonium sulfate and PEG400 were used as precipitants, with both 100 mM and 200 mM HEPES buffer (pH 7.8), and plated with TRIM24 protein (~20 mg/mL) in a 1:2, 1:1, or 2:1 ratio on Intelli-Plate 96-3 low-profile crystallisation plates (Hampton Research). A further narrow screen was conducted around conditions in which crystals were observed (grey shading).

The broad JCSG+ screen resulted in only precipitation, with little indication of crystalline formations. Initial crystals were obtained from a screen of 100/200 mM HEPES (pH 7.5) with 1.7-2.2 M ammonium sulfate and 0.5-4% PEG400 (Figure 4.15). Following this, a narrow screening plate of 100 mM HEPES (pH 7.5) with 2-2.2 M ammonium sulfate and 1.5-3% PEG400 was used, resulting in the formation of large plates of overlapping crystals within 2-3 days, which reached full size within a week (Figure 4.16). This yielded in a greater number of crystals suitable for mounting compared to plates consisting entirely of literature precipitation conditions (100 mM HEPES, 2.0 M ammonium sulfate, 2% PEG400, pH 7.5). For the most defined crystals there were few nucleation points per well, often occurring near the droplet boundary.

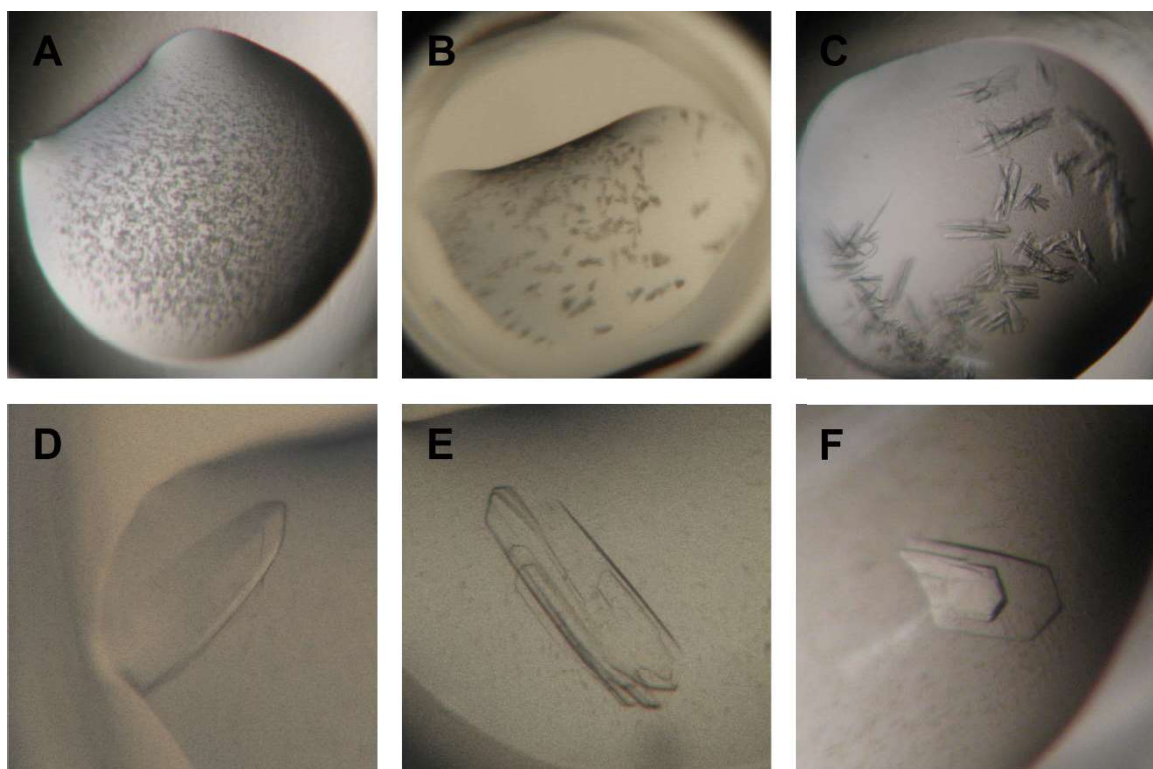


Figure 4.16 Crystal forms of TRIM24 observed under different precipitant conditions. **A:** Broad screen optimisation conditions resulted primarily in precipitation. **B:** Formation of bundles of poorly defined crystals around many nucleation sites were observed; a narrow optimisation plate was designed around these conditions. **C:** Improvement in crystallisation conditions resulted in spindle shaped wheels of crystals across many wells. **D:** In regions of higher PEG content, crystals had blade-like curved edges. **E:** The majority of crystals were in the form of clumps of overlapping, sheer, sheet-like crystals. **F:** Several crystals could be isolated as a single plate extending beyond the main growth, these were successfully separated to be mounted as a single crystal.

Single crystals were separated from sheet-like clusters using a needle, and cryoprotected using a solution of 20% glycerol in the precipitant from the same well reservoir. X-Ray data was obtained at the Diamond Light Source (Harwell, Oxfordshire) by Dr Michael McDonough (University of Oxford), and refined by David Zollman (University of Oxford). Crystals produced high-quality data (Figure 4.17) which was refined with a resolution of 2.0 Å, including well resolved waters of solvation. This structure corroborated well with the existing PDB structure (3O33), alignment with the function `cealign` afforded an RMSD of 1.5 Å over 168 residues. This work has demonstrated that cleavage from the His₆-SUMO affinity label with SENP1 produces structurally homogeneous protein of high purity for crystallisation.

Screening was performed using freshly purified TRIM24, however, crystallisation trials were also conducted with protein flash frozen at the desired concentration, in thin walled PCR tubes using liquid nitrogen.²² Aliquots of protein were thawed on ice and centrifuged (14,000 rpm, 4 °C,

20 min) prior to plating in the same manner. Crystallisation efficiency of freeze-thawed TRIM24 was noticeably lower than that of freshly purified, however, several wells still contained distinguishable crystals of the same form, with well-defined edges and no visible inclusions. The ability to form high quality crystals from stored protein greatly improves the utility of this procedure for obtaining structural information with bound compounds.

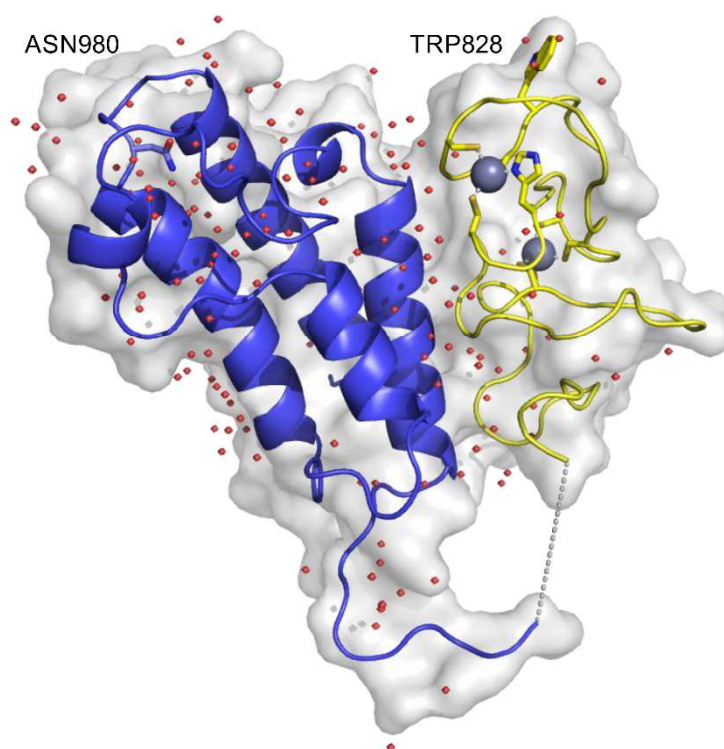


Figure 4.17 Crystal structure of *apo*-TRIM24 obtained in this study (2.0 Å resolution). Bromodomain (blue) and PHD finger (yellow) tandem module, cleaved from pMET11-SUMO3 vector. Highlighted residues: asparagine 980 forms hydrogen bonding interactions with KAc residues; tryptophan 828 forms cation- π interactions with KMe₃; cysteine and histidine residues co-ordinate two structural Zn²⁺ atoms to stabilize the cross-braced topology of PHD finger. $R_{\text{work}}/R_{\text{free}}$: 0.168/0.197.

4.4 Association with nucleosome

In order to connect the affinity and structural information obtained to their physiological relevance in the context of modifications to histone tails as a component of a nucleosome, we attempted to characterise the interactions of TRIM24 with a nucleosome bearing a modification analogous to K18Ac: an acetylated cysteine incorporated in the K18 position, denoted K_s18Ac. Native mass spectrometry could demonstrate the formation of a (nucleosome)(TRIM24) complex, by observing peaks corresponding to the combined masses in a non-denaturing environment. Modified nucleosomes, containing the K_s18Ac modification on both copies of the H3 peptide, were produced by David Zollman (Oxford). Samples of nucleosome and TRIM24 were preincubated at 4 °C, and native MS experiments were performed by Nisha Patel (Oxford).

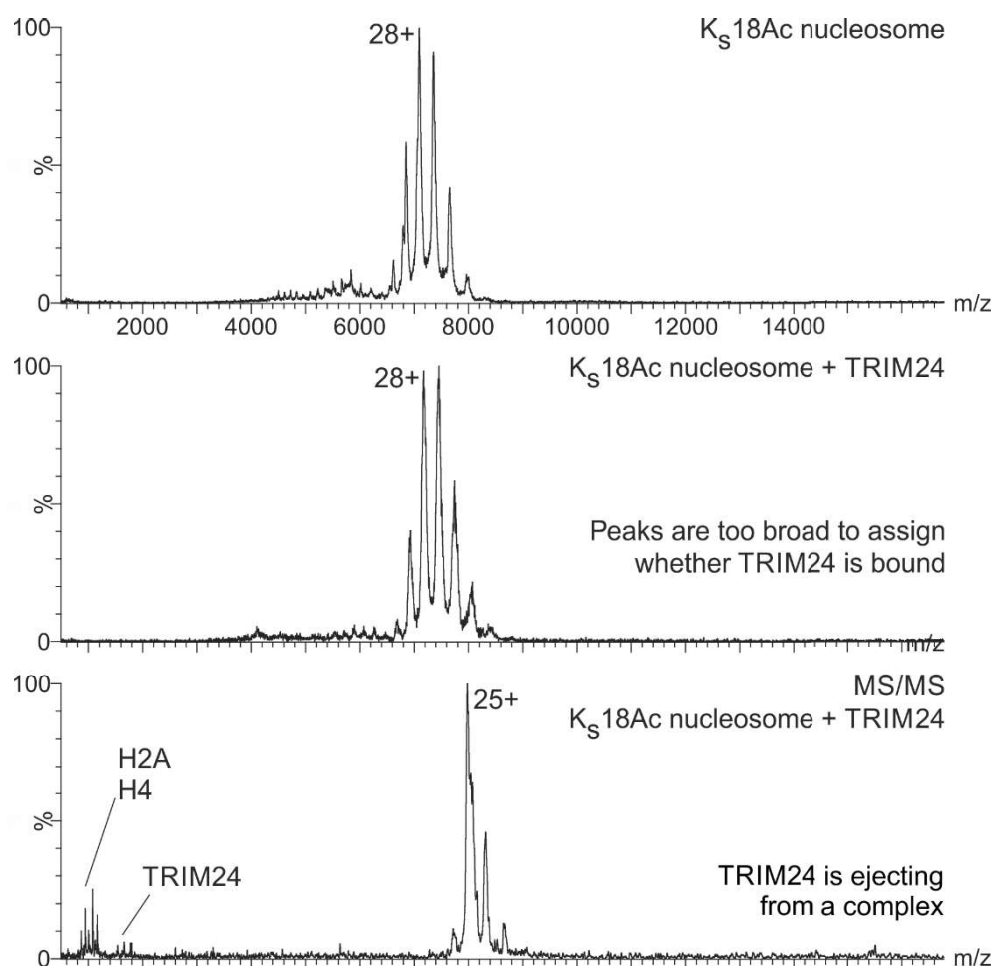


Figure 4.18 Native MS data from incubation of K_s18Ac nucleosome with TRIM24. Intact mass resolution was insufficient to assign a complex formed with TRIM24, however, MS/MS data demonstrated the ejection of a species corresponding to the mass of TRIM24 (as well as H2A and H4) from the complex. Data obtained and figure prepared by Nisha Patel (Oxford).

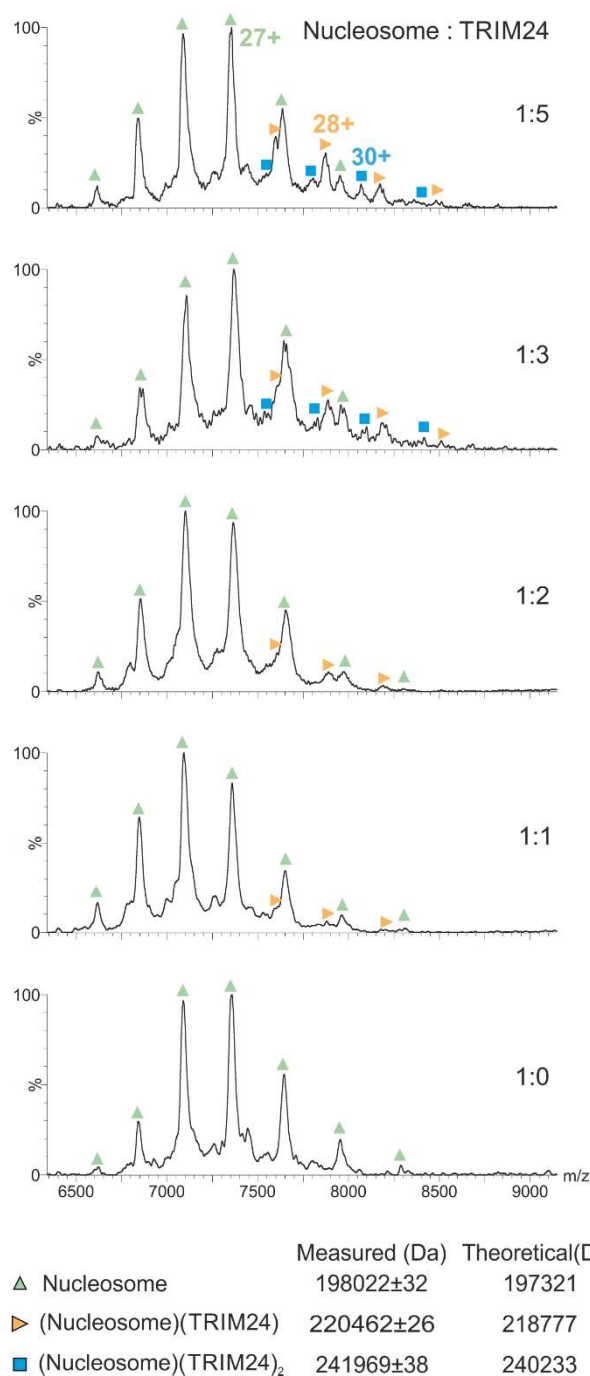


Figure 4.19 Native MS data from titrated incubations of K_s18Ac nucleosome with TRIM24. Peaks corresponding to nucleosome species in complex with TRIM24 indicated. With an excess of TRIM24, complexes containing TRIM24 protein bound in one or two equivalents were observed. Data obtained and figure prepared by Nisha Patel (Oxford).

Initial native MS data was of insufficient resolution to determine whether TRIM24 was bound from peaks corresponding to the whole complex, however, MS/MS data demonstrated TRIM24 was ejecting from a complex, along with H2A and H4 (Figure 4.18). Incubation of the K_s18Ac nucleosome with higher concentrations of TRIM24 (1:5 ratio), enabled peak fitting which matched the observation of both a single associated TRIM24 protein (nucleosome)(TRIM24) and the (nucleosome)(TRIM24)₂ complex, which would be expected from binding of TRIM24 to both H3 histone tails (Figure 4.20).

Titration incubations of varying proportions of TRIM24 with the K_s18Ac nucleosome suggest that although the affinity of the interaction observed in isolated peptides was very high, the kinetics of the association in the context of an entire nucleosome complex may have a high k_{off} , with an equilibrium in favour of a dissociated state (Figure 4.19). A

small proportion of complex was observed from a 1:1 ratio of nucleosome:TRIM24, from peaks which were not present in the nucleosome only negative control. With larger excesses of TRIM24 (>1:3 ratio) the (nucleosome)(TRIM24)₂ complex was also observed, however, even with a 1:5 ratio of nucleosome:TRIM24 the dominant species was nucleosome unit without any associated

TRIM24. Differences between measured and theoretical masses for the complexes were <0.8%, and MS/MS studies demonstrated ejection of TRIM24 from the complex with good resolution of mass identity (Figure 4.21).

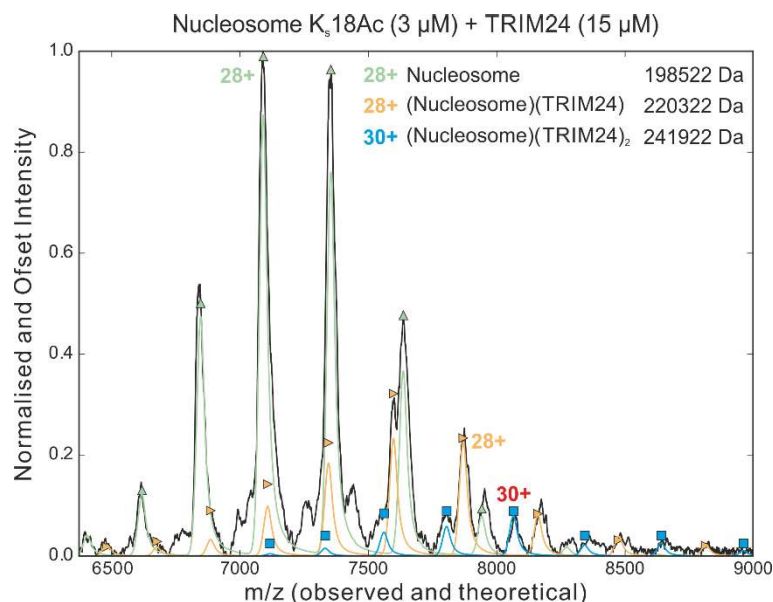


Figure 4.20 Intact mass MS data obtained from incubation of K_s18Ac nucleosome with TRIM24 in a 1:5 ratio. Overlaid theoretical peaks for (Nucleosome)(TRIM24) complex in orange, and (Nucleosome)(TRIM24)₂ complex in red. Data obtained and figure prepared by Nisha Patel (Oxford).

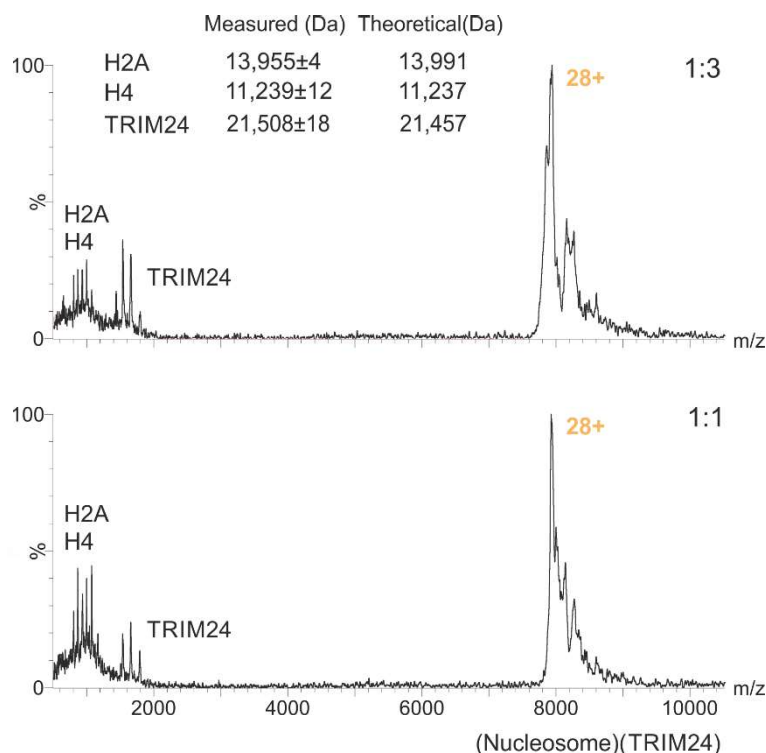


Figure 4.21 MSMS data obtained from incubation of K_s18Ac nucleosome with TRIM24. Demonstrates TRIM24 ejecting from complex, confirming identity of nucleosome-TRIM24 complexes observed in MS studies. Ratio of nucleosome:TRIM24 indicated top right. Data obtained by Nisha Patel (Oxford).

4.5 Conclusions

Structural integrity of TRIM24 and TRIM33 was shown to depend on the expression vector, and was thought to be linked to the low secondary structural content of the PHD finger. Cleavage from His₆ affinity labels with TEV was inefficient and resulted in protein with low internal stability. Successful LIC cloning into a pGTvL1-SGC vector yielded GST- labelled TRIM24 and TRIM33, however, cleavage of the affinity label resulted primarily in precipitation, and the later could not be isolated in a cleaved form. Further transfer *via* Gibson Assembly into a pETM11-SUMO3GFP vector allowed expression of the PHD-bromodomain with an N-terminal His₆-SUMO affinity label, cleavage of which went to completion and resulted in structural homogeneous protein for crystallisation.

After optimisation of conditions, a protocol was refined to reliably produce a several single crystals per plate, from one of which a high quality crystal structure was obtained. Crystal soaking, using the small molecule lead compounds identified in Chapter 4, would enable direct observation of (solid state) binding conformations, and structure guided optimisation of these compounds. Given the dominance of interaction with the KAc pocket observed in docking studies, any compounds which bind in an alternative region, especially the cleft present between domains, would be of particular interest.

Much structure-based design implicitly assumes that the structure is correct, precise and rigid. In the case of the tandem PHD-bromodomain, it is yet to be determined how well defined the cleft between the individual domains remains in solution. A strong association between domains has been demonstrated, through co-elution of a complex formed from individual domains in size exclusion chromatography, but the flexibility of this interdomain region is unknown. Further structural studies, including co-crystallisation with trans-domain spanning peptides, or solvent soaking with compounds predicted to bind in the proximity of the cleft, may provide further insight into structural variability.

Association with a nucleosome bearing a K_s18Ac modification was confirmed by native MS studies, although to a lower degree than would be predicted from affinities. The implications of this result in terms of access to histone modifications in the context of a nucleosome unit could be further investigated by co-crystallisation of the nucleosome-TRIM24 complex.

4.6 References

- (1) Chung, C., Dean, A. W., Woolven, J. M., and Bamborough, P. (2012) Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* 55, 576–586.
- (2) Zhang, G., Plotnikov, A. N., Rusinova, E., Shen, T., Morohashi, K., Joshua, J., Zeng, L., Mujtaba, S., Ohlmeyer, M., and Zhou, M.-M. (2013) Structure-guided design of potent diazobenzene inhibitors for the BET bromodomains. *J. Med. Chem.* 56, 9251–64.
- (3) Palmer, W. S., Poncet-Montange, G., Liu, G., Petrocchi, A., Reyna, N., Subramanian, G., Therooff, J., Yau, A., Kost-Alimova, M., Bardenhagen, J. P., Leo, E., Shepard, H. E., Tieu, T. N., Shi, X., Zhan, Y., Zhao, S., Draetta, G., Toniatti, C., Jones, P., Geck Do, M., and Andersen, J. N. (2015) Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J. Med. Chem.* 10.1021/acs.jmedchem.5b00405.
- (4) Rooney, T. P. C., Filippakopoulos, P., Fedorov, O., Picaud, S., Cortopassi, W. a, Hay, D. a, Martin, S., Tumber, A., Rogers, C. M., Philpott, M., Wang, M., Thompson, A. L., Heightman, T. D., Pryde, D. C., Cook, A., Paton, R. S., Müller, S., Knapp, S., Brennan, P. E., and Conway, S. J. (2014) A series of potent CREBBP bromodomain ligands reveals an induced-fit pocket stabilized by a cation- π interaction. *Angew. Chem. Int. Ed. Engl.* 53, 6126–6130.
- (5) Blundell, T. L., Jhoti, H., and Abell, C. (2002) High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* 1, 45–54.
- (6) Kalkhoven, E., Roelfsema, J. H., Teunissen, H., den Boer, A., Ariyurek, Y., Zantema, A., Breuning, M. H., Hennekam, R. C. M., and Peters, D. J. M. (2003) Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein-Taybi syndrome. *Hum. Mol. Genet.* 12, 441–450.
- (7) Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C. H., Knapp, S., and Mueller, S. (2012) Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* 149, 214–231.
- (8) Tallant, C., Valentini, E., Fedorov, O., Overvoorde, L., Ferguson, F. M., Filippakopoulos, P., Svergun, D. I., Knapp, S., and Ciulli, A. (2015) Molecular Basis of Histone Tail Recognition by Human TIP5 PHD Finger and Bromodomain of the Chromatin Remodeling Complex NoRC. *Structure* 23, 80–92.
- (9) Liu, L., Qin, S., Zhang, J., Ji, P., Shi, Y., and Wu, J. (2012) Solution structure of an atypical PHD finger in BRPF2 and its interaction with DNA. *J. Struct. Biol.* 180, 165–173.
- (10) Li, H. T., Ilin, S., Wang, W. K., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442, 91–95.

- (11) Tsai, W.-W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M.-C., Patel, D. J., and Barton, M. C. (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927–932.
- (12) Bennett, J. M., Fedorov, O., Tallant, C., Monteiro, O. P., Meier, J., Gamble, V., Savitski, P., Nunez-Alonso, G. A., Haendler, B., Rogers, C., Brennan, P. E., Müller, S., and Knapp, S. (2015) Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *J. Med. Chem.* 10.1021/acs.jmedchem.5b00458.
- (13) Xi, Q., Wang, Z., Zaromytidou, A.-I., Zhang, X. H.-F.-F., Chow-Tsang, L.-F., Liu, J. X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., Studer, L., Mark, W., Patel, D. J., Massagué, J., and Massague, J. (2011) A Poised Chromatin Platform for TGF-beta Access to Master Regulators. *Cell* 147, 1511–1524.
- (14) Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., and Panne, D. (2013) Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat. Struct. Mol. Biol.* 20, 1040–6.
- (15) Wang, Z., Song, J., Milne, T. a., Wang, G. G., Li, H., Allis, C. D., and Patel, D. J. (2010) Pro isomerization in MLL1 PHD3-Bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell* 141, 1183–1194.
- (16) Zeng, L., Yap, K. L., Ivanov, A. V., Wang, X., Mujtaba, S., Plotnikova, O., Rauscher, III, F. J., and Zhou, M.-M. (2008) Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat. Struct. Mol. Biol.* 15, 626–633.
- (17) Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal. Biochem.* 216, 413–417.
- (18) Arnau, J., Lauritzen, C., Petersen, G. E., and Pedersen, J. (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* 48, 1–13.
- (19) Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Iii, C. A. H., Smith, H. O., and America, N. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 12–16.
- (20) Shen, L., Tatham, M. H., Dong, C., Zagorska, A., Naismith, J. H., and Hay, R. T. (2006) SUMO protease SENP1 induces isomerization of the scissile peptide bond 13, 1069–1077.
- (21) Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Jelloul, M. Ben, Sussman, J. L., Stuart, D. I., and Perrakis, A. (2005) Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 61, 1426–1431.
- (22) Deng, J., Davies, D. R., Wisedchaisri, G., Wu, M., Hol, W. G. J., and Mehlin, C. (2004) An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallogr. D. Biol. Crystallogr.* 60, 203–204.
- (23) Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876–2890.

Chapter 5: Acetyl-lysine-mimicking Residue

5.1 Introduction

Chapter 2 described the development of screening methods for identifying binding partners for TRIM24 and TRIM33. The AlphaScreen assay developed depends on correct identification of appropriate cognate peptides which can be displaced by ligands. For many proteins such competition assays are not feasible, either because the endogenous binding partner is unknown, or the affinities in protein-protein interactions are prohibitively low. To resolve this for KAc-binding proteins, we attempted to utilise the understanding of bromodomain-peptide interactions derived from AlphaScreen assays to explore novel methods of increasing peptide affinity for binding to bromodomains. It was envisioned that chemical modification of a histone peptide, to incorporate a KAc-mimicking residue with increased affinity compared to a natural KAc modification, could enable competitor peptide design for assay development, with lower dependence on knowledge of cognate peptide sequence (Figure 5.1). This approach could then be expanded to other bromodomains without prior knowledge of binding partners or small-molecule ligands.

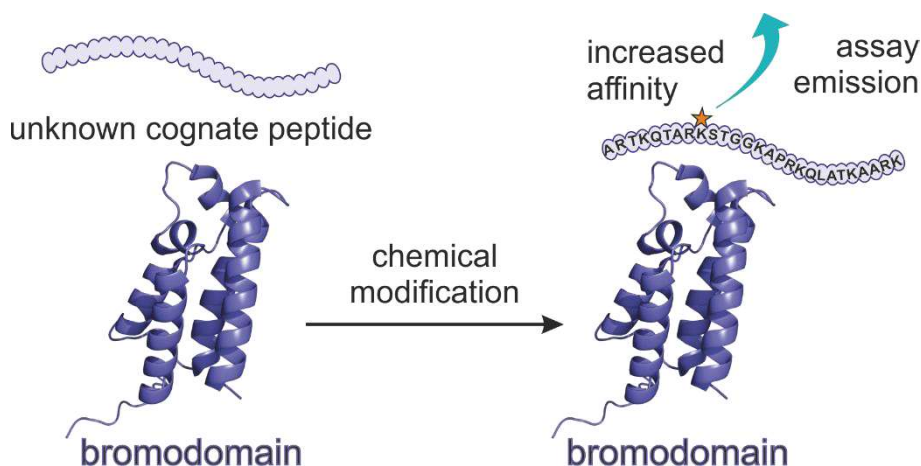


Figure 5.1 Competition based assays cannot be developed for bromodomains without prior knowledge of binding partners. Chemical modifications to incorporate a KAc-mimicking residue into a histone-like peptide, and hence increase the affinity of this peptide for a bromodomain, could result in sufficient levels of interaction for displacement assay development.

5.1.1 Non-natural acetyl-lysine mimicking amino acids

Homogeneously modified proteins are essential in understanding the role of particular post-translational modifications (PTMs). A number of chemical strategies have been developed to synthesise single homogeneous forms of proteins, either containing the native PTM or an analogue which functions in a similar manner. These include: i) native chemical ligation (NCL)¹/expressed protein ligation (EPL),² a semi-synthetic route involving the reaction between a cysteine residue and a thioester on an expressed protein or peptidic fragment containing a PTM;³ ii) Amber codon suppression, which allows an unnatural amino acid to be introduced co-translationally at a genetically-encoded site;^{4,5} iii) “tag-and-modify” approach, which uses chemical reaction of a translationally incorporated amino acid ‘handle’, often a cysteine residue, to form a mimic of the desired modified amino acid.^{6,7} These techniques have all been used to introduce modified lysine residues or mimics thereof, and their value in uncovering the mechanistic roles of histone modifications has been clearly demonstrated (Figure 5.2).^{8,9}

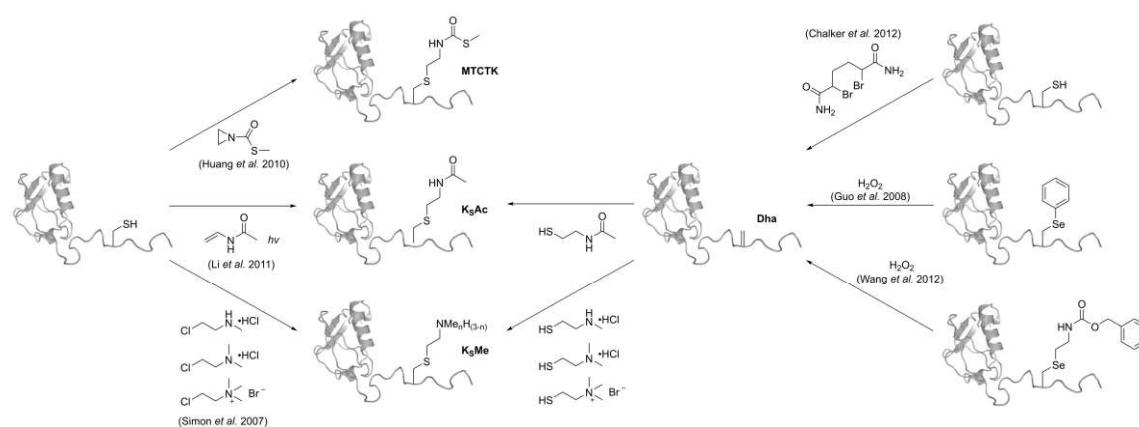


Figure 5.2 Strategies of chemical PTM modifications towards analogues of acetylated (MTCTK and K_εAc) and methylated (K_εMe) lysine. “Tag-and-modify” approaches using direct alkylation of cysteine residue, or addition to a dehydroalanine residue (Dha).

Nonetheless, current KAc mimics suffer from several disadvantages, including the harsh oxidising conditions required for their preparation, or formation of epimeric mixtures. Perhaps most significantly, there has only been one reported attempt to assess the binding efficiency of KAc mimics to bromodomains, in which a pull-down assay assessing the binding efficiency of histone H4 tails to BRDT showed that KAc mimics containing methylthiocarbonylthialysine (MTCTK), a thiocarbamate analog of KAc, interacted 2-4 fold less efficiently than KAc itself.⁷ Therefore,

information learnt in the development of BET bromodomain ligands, was employed to design peptides incorporating novel KAc-mimicking amino acids.

5.1.2 BRD4 bromodomains

Bromodomain containing protein 4 (BRD4) was chosen as the bromodomain containing protein (BCP) on which to conduct initial studies. This protein is a member of the BET (bromodomain and extra C-terminal domain) bromodomain family, members of which have been shown to be involved in apoptosis, metastasis and proliferation in cancer.¹⁰ There are specific implications of BRD4 in pathology: nuclear protein in testis (NUT) midline carcinoma, a rare but incurable cancer, is driven by the fusion of the *BRD4* (or less frequently bromodomain containing protein 3 – *BRD3*) gene and the *NUT* gene;¹¹ suppression of BRD4 in Myc-driven acute myeloid leukemia (AML) leads to cell-cycle arrest and apoptosis of leukemia cells.¹² Despite the emerging role of BCPs in human disease,¹³ the precise biological function of the bromodomains in pathogenesis has not yet been fully elucidated.

Like other BET bromodomain proteins, BRD4 contains two bromodomains, denoted BRD4(1) and BRD4(2), the former being the focus of this work. There is less similarity between BRD4(1) and BRD4(2) than between the first bromodomains of homologous proteins within the BET family i.e. between BRD2(1), BRD3(1) and BRD4(1).

BRD4(1) binds strongly to multiply-acetylated H4 tails both *in vitro* and in live cells: reported affinities to tetra-acetylated H4 include $4.8 \pm 4 \mu\text{M}$ (competition TR-FRET),¹⁴ and $2.8 \pm 0.2 \mu\text{M}$ (ITC),¹⁵ with fluorescence recovery after photobleaching (FRAP) studies demonstrating cellular association. X-Ray crystal structures demonstrate simultaneous recognition of two KAc residues on an H4 peptide separated by two or three linking residues (PDB: 3UVW, 3UVX, 3UVY).¹⁵ A tetra-acetylated H4 peptide has been used in the BRD4(1) AlphaScreen to identify small molecule inhibitors of this domain.^{16,17} Furthermore, two separate ITC studies reported the dissociation constant for the same peptide (H4₁₋₂₀KAc₄) with BRD4(1): $K_D = 2.8 \mu\text{M}$ ¹⁵ and $K_D = 3.1 \mu\text{M}$ ¹⁶. In

summary, the interaction of BRD4(1) with the histone H4₁₋₂₀(KAc)₄ peptide has been well characterised and was therefore chosen as a model system to investigate the effects of incorporation of KAc-mimicking amino acids.

5.1.3 3,5-Dimethylisoxazoles as KAc mimics

Small molecule probes have been developed in order to disrupt the bromodomain-KAc interaction,¹⁸ and subsequent studies have contributed to the elucidation of the biological roles of these proteins. Amongst others discussed in the Introduction, isoxazoles were discovered as an effective KAc mimic that displaces histone peptides from bromodomains.¹⁹ This low molecular weight motif demonstrated good ligand efficiency and has proven suitable for further development to give potent ligands of bromodomains.^{20,21} A 3,5-dimethylisoxazole (DMI) containing compound reported by Hewings *et al.* was optimised to inhibit BRD4(1) KAc recognition with an IC₅₀ of 371 nM (Figure 5.3).²²

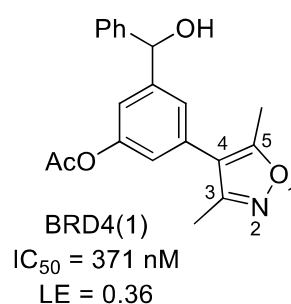


Figure 5.3 Optimised isoxazole containing inhibitor of BRD4. Numbering system included for 3,5-dimethyl isoxazole core.

Methodology for preparing unnatural amino acids containing the isoxazole moiety has previously been reported by both Falorni *et al.*²³ and Giacomelli *et al.*,²⁴ however, in these compounds the isoxazole is attached directly to the α -carbon of the amino acid. The overlaid crystal structures of a

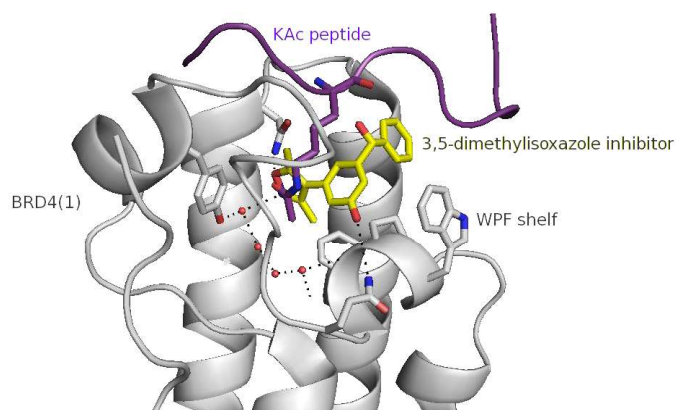


Figure 5.4 Comparison of peptide and small molecule binding in BRD4(1) KAc pocket (PDB: 3UVW). Diaryl methanol derivative of 3,5-dimethylisoxazole (DMI – yellow) binds in hydrophobic pocket, with aryl substituent residing on ‘WPF’ shelf, and polar isoxazole forming hydrogen bonding interactions in a similar fashion to KAc residue overlaid (purple)

KAc containing peptide and DMI containing ligand to the BRD4(1) bromodomain suggest this attachment is unlikely to be of sufficient length to penetrate the binding pocket of the bromodomain effectively (Figure 5.4).

5.1.4 Aims

Well-characterised, peptide-integrated KAc mimics, with comparable affinities to bromodomains, would be valuable for further study on the functional significance of bromodomain-KAc interactions. Increasing the affinity of competitor peptides in displacement assays such as AlphaScreen, might improve reliability of existing assays,¹⁶ and allow new assays to be developed in cases when a cognate peptide of sufficient affinity is not known.²⁵ Incorporating a KAc-mimicking amino acid in a peptide in place of KAc is a hitherto unexplored strategy for enhancing the affinity for binding to bromodomains (Figure 5.5). In this fashion, they might prove useful in developing assays for bromodomains without known ligands,²⁶ for example sub-family VI in the bromodomain classification including TRIM28 and MLL, in order to facilitate small-molecule screening.

The ability to incorporate a KAc mimic *in situ*, in site-selective manner, would enable precise activation of KAc mediated PPIs to study downstream effects. Since modified peptides would resemble natural binding partners more closely, it is possible this would cause fewer off-target effects or disruptions.

It was hypothesised that appropriately linked isoxazole-containing amino acids might bind to bromodomains. A range of unprotected isoxazole-containing L-amino acids were designed and synthesised by David S. Hewings, Natalie H. Theodoulou, and Lukass Jursins (University of Oxford), with varying chain lengths, degree of saturation and location of hydrogen bond donors and acceptors (unpublished). Substitution of an isoxazole core at the 3-, 4- or 5-position could increase potency and selectivity of these small molecules, relative to an individual KAc amino acid. Installation of amide functionality on the linkage could allow the formation of

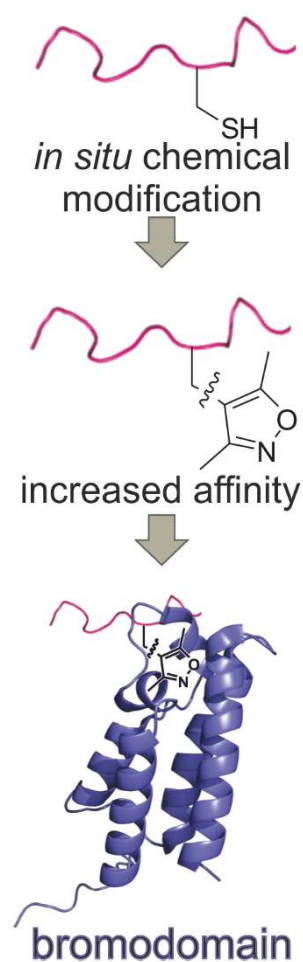


Figure 5.5 Hypothesis that incorporation of DMI unit as KAc mimicking residue could increase affinity of peptide for bromodomains.

further hydrogen bonds to an adjacent area of the binding pocket, which a second KAc has been shown to occupy.¹⁵ *Meta*-substitution of a phenylalanine derivative was predicted to overlap effectively with the second KAc and interact with the hydrophobic residues within the ZA channel region of the binding pocket.

The aims of this study were to:

- Assess the affinity of isoxazole containing non-natural amino acids, by transferring the previously developed AlphaScreen assay to the BRD4(1)-H4(KAc)₄ system.
- Quantify the effect of integration of a selection of these amino acids as replacements for KAc residues at different positions of H4(KAc)₄-like peptides.
- Optimise an isoxazole-based alkylating agent to modify a cysteine residue *in situ* and assess the subsequent effect on affinity.

5.2 Inhibitory values of free amino acids

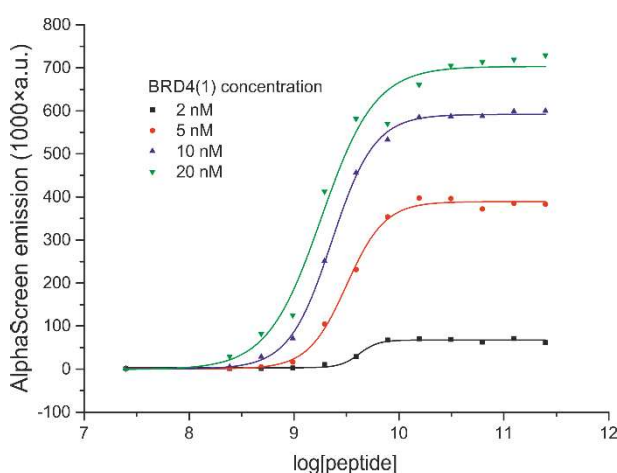


Figure 5.6 Biotinylated peptide signal response curves for BRD4(1) with H4₁₋₂₀(KAc)₄. AlphaScreen emission responses with varying concentrations of protein and competitor peptide.

Initial biological evaluation of isoxazole-containing free amino acids was performed using the AlphaScreen assay to assess the potential for increasing affinity relative to KAc. Transfer of the assay protocol to this system was achieved by measurement of signal response curves to define BRD4(1) protein and H4₁₋₂₀(KAc)₄ competitor peptide concentrations that resulted in optimal

response (Figure 5.6). Assay parameters were selected to generate a maximum signal of over 300,000 units, providing an appropriate assay window to avoid artefacts, and also lie close to the drop-off point in the curves, which would result in good assay sensitivity. With an AlphaScreen bead dilution of 1:400, concentrations of 4 nM and 10 nM, respectively, were therefore selected as optimal assay conditions. Insufficient signal was observed with all KAc competitor peptides previously used: H3₁₋₂₇K14Ac, H3₁₋₂₇K18Ac and H3₁₋₂₇K23Ac gave emission of less than 40,000 units even up to 100 nM BRD4(1) concentrations (data not shown).

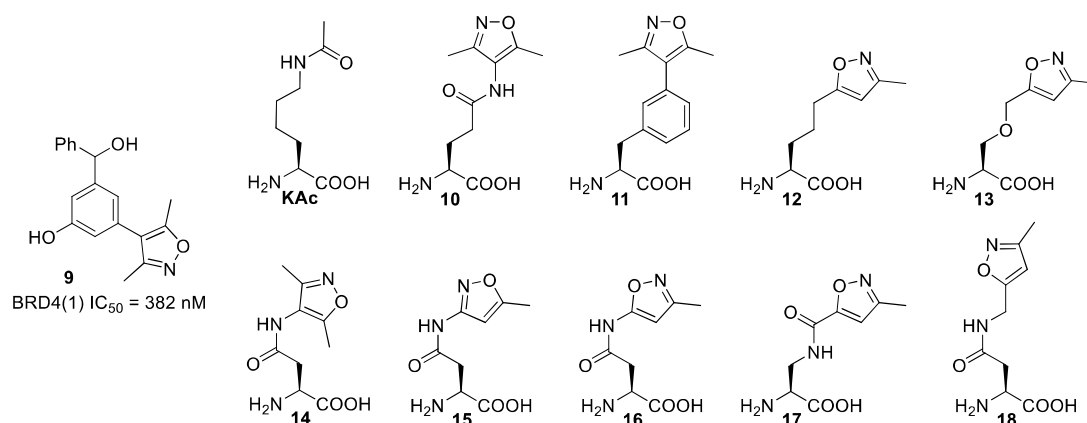
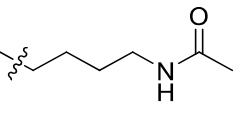
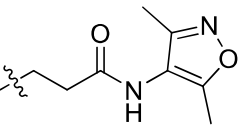
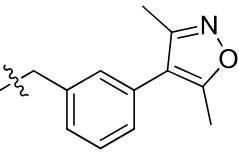
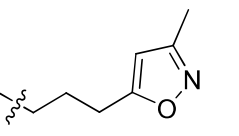
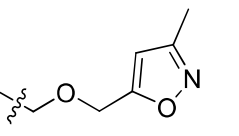
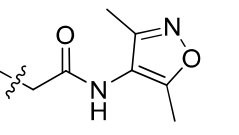
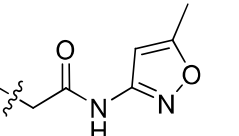
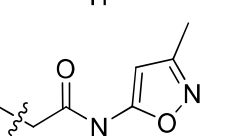
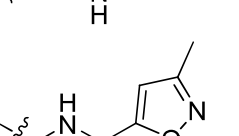
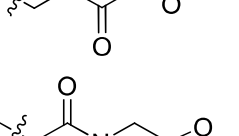


Figure 5.7 Amino acids synthesised within the group as acetyl lysine mimetics. Comparison to acetyl lysine (KAc), with reference phenol DMI derivative included as positive control shown left. Dimethyl isoxazole units were attached through aromatic, alkyl, ether, and amide linkers to the alpha position of a carboxylic acid.

Inhibition of BRD4(1) by unnatural amino acids (Figure 5.7) was measured as an average of two separate assays performed in triplicate, normalised against references (Table 5.1: buffer blank – 0% inhibition; 25% DMSO – 100% inhibition). The phenol derivative (**9**) bearing a 3,5-dimethylisoxazole moiety, with an IC_{50} of 382 nM,²² was included on each plate as a positive control, which in every case demonstrated 100% inhibition. Along with previous data against the bromodomains of BRD9 and BAZ2A (obtained by the SGC), this work confirmed that these compounds are the first unnatural amino acids known to successfully inhibit bromodomain recognition of an acetyl-lysine-containing peptide. Several amino acids resulted in >50% inhibition at 250 μ M, with significant inhibitory effect detected at lower concentrations of 25/50 μ M. These findings represent an improvement on the binding affinity of acetyl-lysine, which displayed negligible inhibition at 250 μ M.

Three amino acids incorporating an isoxazole attached by the 3- or 5- position to a three-atom linker (**15-17**) showed a preference for BAZ2A inhibition, with >80% inhibition at 250 μ M. These compounds also demonstrated some affinity for BRD4(1), with ~30% inhibition at the same concentration. Attachment of a isoxazole with a three atom linker seemed preferable for BRD4(1) inhibition over a four atom linker: cf. 23.6% and 7.4% inhibition at 250 μ M for amides linked at 4-position to three (**14**) or four (**10**) atom linkers, respectively. These data suggest that this length is sufficient, and extended lengths are less well tolerated. Amide connections demonstrated a higher percentage of BRD4(1) inhibition than alkyl or ether, with a preferred orientation of the carbonyl adjacent to the isoxazole (31.2% inhibition at 250 μ M for **17**).

Table 5.1 Inhibitory values of free amino acids against bromodomains. Single-shot inhibition values (AlphaScreen assay) given as average percentage inhibition against each bromodomain, from measurements in triplicate. Inhibition values highlighted as a heatmap from green (low inhibition) to red (high inhibition). Data for BAZ2A and BRD9 were obtained by the SGC. n.d. values not determined.

Protein	BRD4(1)		BAZ2A		BRD9	
Competitor peptide	H4(KAc) ₄		H3K14(Ac)		H2(KAc) ₃	
Concentration (μM)	250	50	250	25	250	25
R-group on L-amino acid	Percentage inhibition					
 KAc	10.8	4.6	n.d	n.d	n.d	n.d
 10	7.4	4.1	25.1	1.4	-3.1	-4.8
 11	76.3	18.3	45.3	5.8	0.5	-7.7
 12	16.4	2.3	n.d	n.d	n.d	n.d
 13	10.3	0.6	n.d	n.d	n.d	n.d
 14	23.6	1.3	66.9	10.4	12.6	-2.1
 15	26.3	0.9	96.1	12.6	47.9	-1.0
 16	n.d	n.d	84.3	14.5	14.6	1.4
 17	31.2	2.5	98.7	58.1	73.6	28.6
 18	18.6	4.7	53.8	1.9	-0.1	-8.1

The amino acid with the strongest inhibitory activity against BRD4(1) was the phenyl-linked isoxazole (**11**), which showed >70% inhibition of binding of tetra-acetylated H4₁₋₂₀ to BRD4(1) at 250 μ M. The apparent selectivity over BRD9, for which no significant inhibition was observed, might be due to the recognition of multiple acetylation marks by BRD4(1), as discussed above. It appears that the aromatic linker may be able to interact with the hydrophobic region of the ZA channel and displace multiple, proximal KAc residues more effectively. Additionally, compared to the alkyl chains with high degrees of freedom, the phenyl ring would also act to rigidify the linker, which may account for the increased selectivity.

5.3 Incorporation into peptides

It has been established that a single BET bromodomain recognises two adjacent KAc residues on the same peptide.¹⁵ X-Ray crystal structures of multiply-acetylated peptides with BRD4(1) revealed the N-terminal KAc forms the hydrogen-bonding interaction with the conserved asparagine residue, while flexible glycine residues allow a peptide conformation where the C-terminal KAc fits into the grooves created by the ZA and BC loops. It is not clear whether an isoxazole-containing amino acid could effectively mimic a KAc residue in either of these binding modes.

Each of the four KAc positions were replaced individually with alkyl- (**12**), ether- (**13**), or amide- (**16**) linked isoxazoles. In order to compare the effects from the linker alone, attachment was *via* the 5-methyl in each case, which previous data suggested would lead to detectable levels of inhibition. A control peptide, H4₁₋₂₀KAc₄ was also synthesised to determine whether the proposed KAc mimics could interact more strongly with bromodomains than KAc itself. In order to accurately quantify peptide concentrations by UV absorbance, an N-terminal tyrosine residue was included in all sequences ($\epsilon_{280} = 1280 \text{ M}^{-1}\text{cm}^{-1}$ 6.0 M guanidine·hydrochloride). This additional residue was also present in the H4₁₋₂₀(KAc)₄ sequence employed by Filippakopoulos *et al.* in ITC studies,¹⁵ and had little apparent effect on affinity.

5.3.1 IC₅₀ value determination

IC₅₀ values were obtained in the AlphaScreen assay by a 1:2 serial dilution of the peptides, performed in triplicate. Results contained low error between replicates (average standard deviation for all replicate values on plate was 2.8% of maximum signal), with a high Z value of 0.9. The dilution series covered a good range of the curves, and fitting was performed according to an IC₅₀ - 4 Parameter Fit in GraFit²⁷ (Figure 5.8).

H4₁₋₂₀KAc₄ synthesised by David Hewings (Oxford) was included as a control, along with the same peptide purchased commercially (GL Biochem), with identical sequence and modification pattern. These gave closely fitting IC₅₀ values of 17.1 μ M and 15.7 μ M, respectively. The 3,5-dimethylisoxazole phenol derivative, also used as a positive control on each individual assay plate, afforded IC₅₀ values of 280 nM and 257 nM, which was in line with published values (IC₅₀ 382 nM²²).

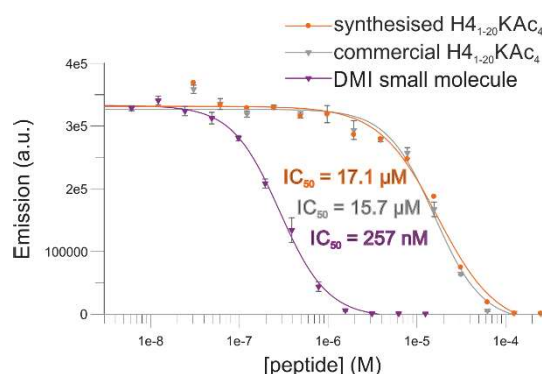


Figure 5.8 AlphaScreen concentration response curves for dilution series of positive controls: tetra-acetylated H4 peptide H4₁₋₂₀KAc₄, and a 3,5-dimethylisoxazole phenol derivative as a small molecule control with a reported IC₅₀. Results are the average of readings taken in triplicate, with fittings are overlaid, performed as IC₅₀ - 4 Parameter Fit in GraFit. Form of curve and fitting is representative of results from other peptide concentration series.

Affinities of modified peptides were strongly dependent on the position of the unnatural residue. Substitution at K5 or K8 greatly reduced the affinity compared to the cognate tetra-acetylated peptide (IC₅₀ range 23.5-63.9 μ M). The alkyl linker had the least detrimental effect in these positions, with only an approximately 1.5 fold increase in IC₅₀. However, substitution at K12 or K16 resulted in peptides with roughly equal affinity to this reference peptide (IC₅₀ range 6.8-18.1 μ M). Additionally, it appears that an amide-linked isoxazole in these positions bound more strongly than the control with an IC₅₀ of 6.8 and 8.7 μ M for the K12 and K16 position, respectively.

Positional sensitivity seemed to be in good agreement with previous studies on this peptide, which demonstrated that dual alkylation at K5 and K8 was almost as potent as the tetra-acetylated peptide (H4₁₋₁₁K5AcK8Ac K_D = 6.8 $\pm$ 0.1 by ITC).¹⁵ Given the high contribution of these positions to

overall affinity, it seemed reasonable that their substitution was less well tolerated. In the same study, alkylation at K12 and K16 had a similar synergistic effect on affinity, though to a lesser extent ($H4_{11-21}K12AcK16Ac$ $K_D = 46.1 \pm 0.9$ by ITC), and the data suggested a greater potential for increasing affinity with KAc mimics in these positions.

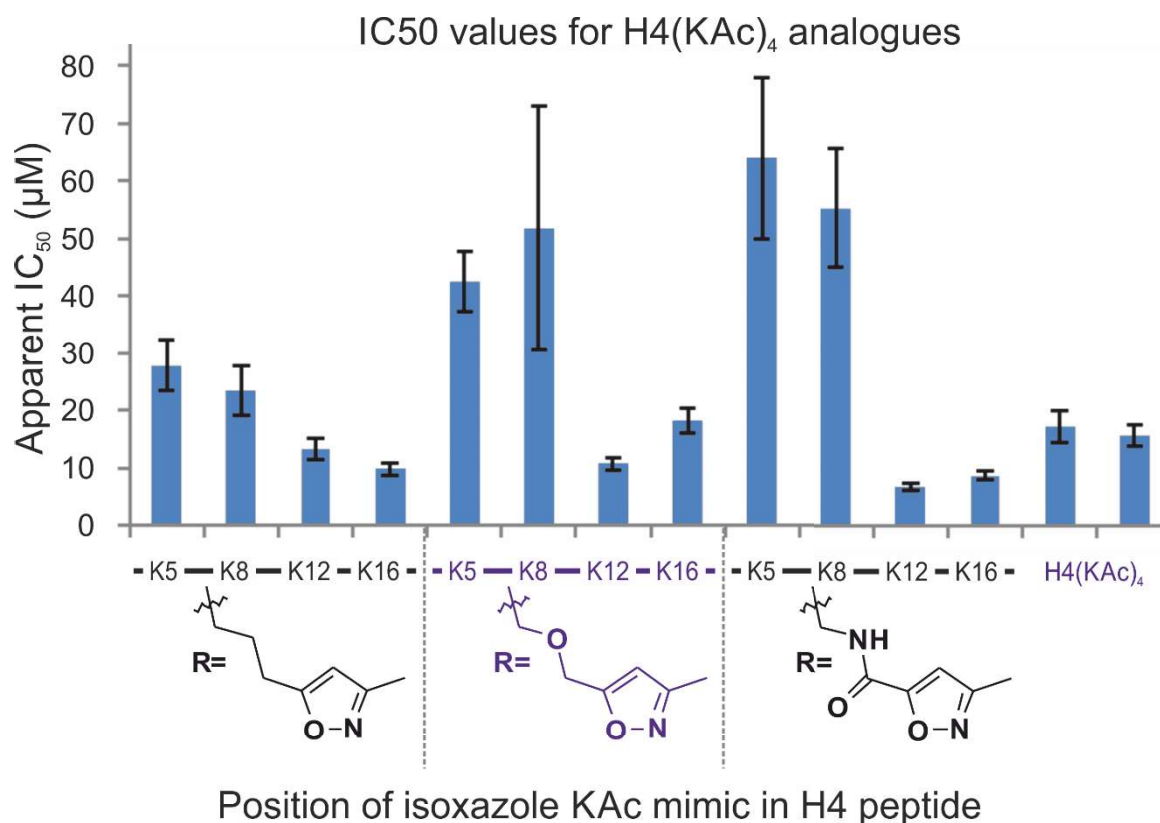


Figure 5.9 IC₅₀ values for H4(KAc)₄ peptides bearing isoxazole containing non-natural amino acid substitutions. Substitution positions are indicated. Data obtained from serial dilutions in AlphaScreen assay inhibiting the interaction of BRD4(1) with biotinylated H4(KAc)₄. Results are the average of readings taken in triplicate, with fitting performed as IC₅₀ - 4 Parameter Fit in GraFit; error bars shown as the standard error of the fit.

5.3.2 Validation of controls

The measured IC₅₀ values for the control H4₁₋₂₀(KAc)₄ peptides, whether synthesised internally or commercially purchased, were much higher than expected. In order to confirm the reliability of this assay we sought to quantify the affinities in an orthogonal, non-fluorescence based assay. Isothermal calorimetry (ITC) is still considered the ‘gold’ standard for quantifying binding affinity,²⁸ directly measuring thermodynamic contributions to affinity, and has frequently been used to characterise affinities of both peptides²⁹ and novel ligands.³⁰ Hence, ITC data were obtained for the control peptide to compare quantitative values (Figure 5.10). This experiment produced a good

titration across the range of interaction, and curve fitting according to a one site binding model resulted in a K_D of 14.39 μM , which was in good agreement with our AlphaScreen values. The measured stoichiometry of 0.88 sites suggested a 1:1 binding interaction, which may account for the difference in affinity to the values reported in the literature, which suggest a 1:2 binding interaction (2.8 μM reported with $N = 0.57$ sites;¹⁵ 3.1 μM reported with $N = 0.61$ sites¹⁶). This result validated the AlphaScreen assay for detecting quantitative affinities of modified H4 peptides for BRD4(1), and substantiated the affinities obtained for peptides bearing isoxazole containing non-natural amino acids.

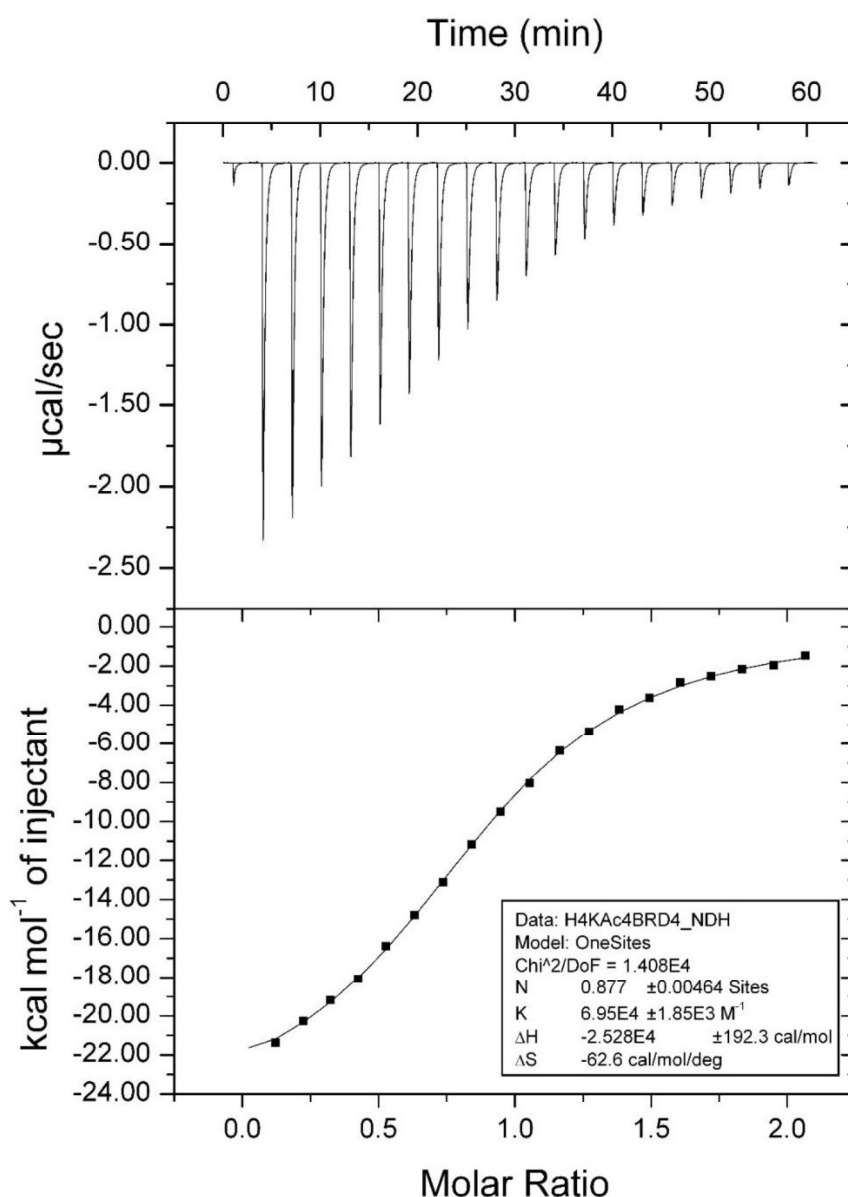


Figure 5.10 ITC curve for interaction between BRD4(1) and H41-20(KAc)4. Peptide ligand (60 μL ; 1.0 mM) was titrated into the cell containing the protein (250 μL ; 100 μM) in a series of 20 $\times$ 2 μL injections (first injection 0.4 μL). Error values represent errors of the nonlinear least-squares fit. Buffer conditions used for the titrations were 50 mM HEPES (pH 7.5), 500 mM NaCl.

5.4 *In situ* alkylation

Having developed a reliable AlphaScreen assay that could quantitatively compare affinities of isoxazole containing peptides, the study progressed towards attempts to increase affinity of a peptide for a bromodomain *in situ*. It was thought that a similar isoxazole alkylating agent to that used during synthesis of individual amino acids, would be able to modify an amino acid in the context of a peptide, thus incorporating a KAc-mimicking residue. Given the positional dependence demonstrated in IC₅₀ measurements, it was decided to attempt a modification of the K12 position of a H4 peptide (Figure 5.11). As well as a greater increase in affinity afforded by substitution at K12 or K16 relative to K5 or K8, the K12 position had the advantage of increased peptide flexibility and accessibility in that region, with multiple glycine residues towards both the N- and C-terminus.

Tyr-Ser-Gly-Arg-Gly-Lys(Ac)-Gly-Gly-Lys(Ac)-Gly-Leu-Gly-Cys-Gly-Gly-Ala-Lys(Ac)-Arg-His-Arg-Lys-NH₂

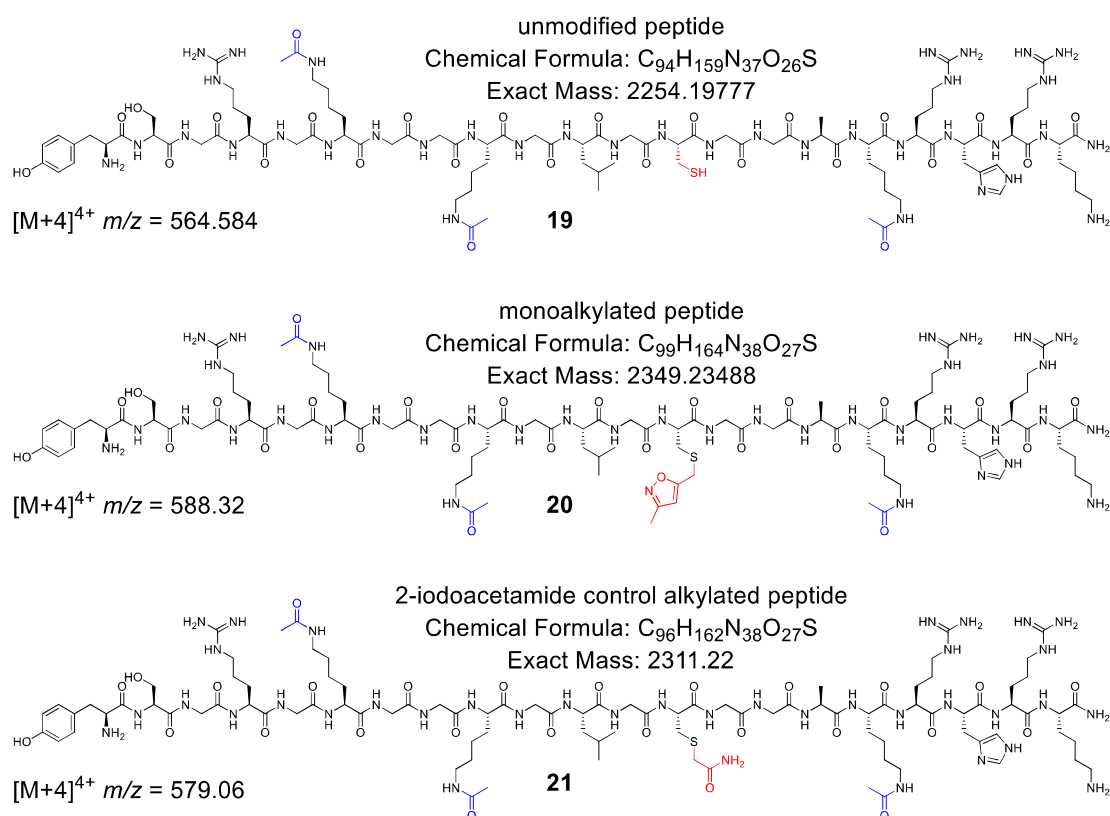


Figure 5.11 Chemical structures of unmodified and alkylated H4-like peptides. Position 12 of a H4(KAc)₄ peptide was replaced with a cysteine (**19**), which could be modified with a halogen-based isoxazole alkylating agent, giving **20**. 2-iodoacetamide was used as an alternative alkylating agent to provide a positive control, resulting in **21**. Acetylation shown in blue; K12C substitution and alkylation shown in red.

Common alkylation reactions occur *via* nucleophilic attack at an electron deficient site of an alkylating agent, either by bimolecular nucleophilic substitution (S_N2 , e.g. with haloalkanes), or Michael-type addition (e.g. maleimides containing an activated, conjugated alkene). Cysteine is the most nucleophilic of the 20 naturally coded amino acids, and one of the least abundant. Indeed, C110 of histone H3 is the only cysteine present in human histone proteins, and a C110A substitution does not disrupt nucleosome function.⁶ Therefore, a “tag-and-modify” approach was chosen to substitute the K12Ac with a cysteine which could be alkylated with an isoxazole. Regioselectivity would be determined by the competition between this inserted cysteine and other common nucleophilic moieties found in naturally occurring amino acids: hydroxyl, thioether and primary or secondary amine groups.

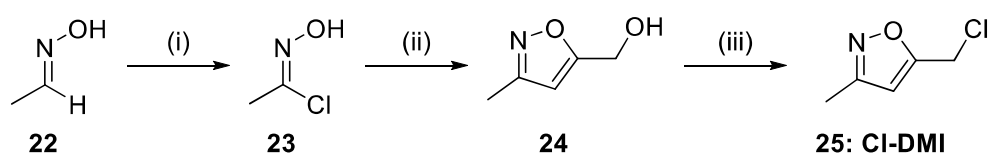


Figure 5.12 Synthesis of 5-(chloromethyl)-3-methylisoxazole (Cl-DMI) alkylating agent. *Reagents and conditions:* (i) NCS, DMF, 0 °C, then rt, 3 h, 78%; (ii) Propargyl alcohol, NEt₃, CH₂Cl₂, rt, dropwise addition of **23** in CH₂Cl₂ over 10 h, then rt, 24 h, dark, 11%; (iii) Thionyl chloride, argon atm., 80 °C, 20 minutes, quantitative by NMR.

Before alkylation was attempted, the affinity of the modified H4(KAc)₃K12C peptide (Figure 5.11, **19**) was assessed, in order to attribute any improved interactions after alkylation with the isoxazole KAc mimic, rather than with the K12C substitution. ITC data confirmed the expected decrease in affinity relative to H4(KAc)₄, with a measured K_D of 32.4 μ M (Figure 5.13).

Previous studies suggested that 5-(chloromethyl)-3-methylisoxazole (Cl-DMI), which was used during the synthesis of the unnatural amino acids, was an efficient alkylating agent, and was therefore employed in this study. Starting reaction conditions were adapted from those in the literature.³¹ The buffer used was strongly denaturing to ensure an accessible, open conformation of the peptide (4 M guanidine·hydrochloride) and included tris(2-carboxyethyl)phosphine (TCEP) as a reducing agent, as well as methionine (10 mM) as an extra oxygen scavenger. Reducing conditions were maintained to avoid oxidation of the thioether to a sulfoxide, and to prevent unwanted disulphide bond formation. Selectivity of alkylation was then assessed over a range of buffer, pH, reducing and temperature conditions.

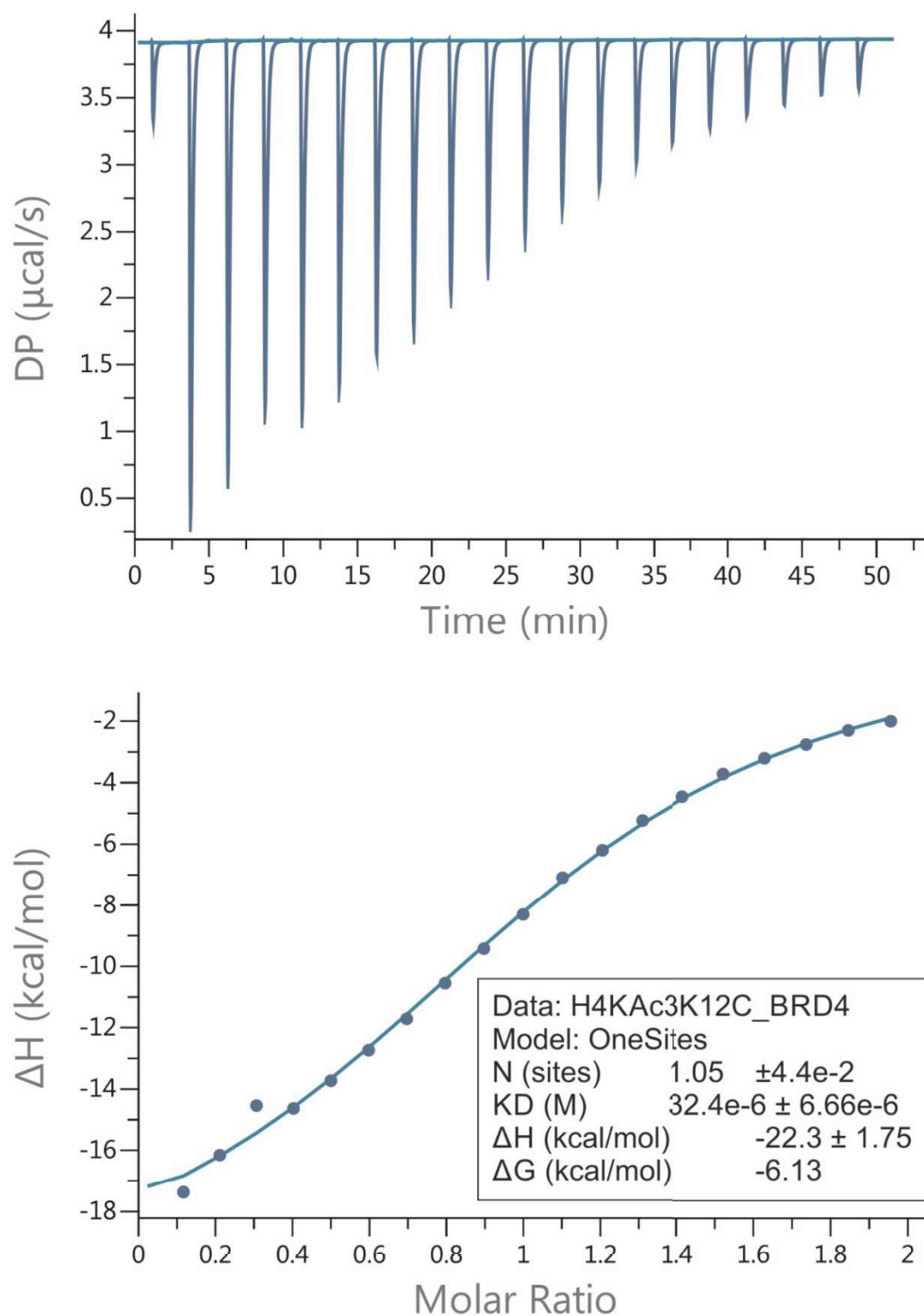


Figure 5.13 ITC curve for interaction between BRD4(1) and H4₁₋₂₀(KAc)₃K12C (19). Peptide ligand (60 μL ; 1.27 mM) was titrated into the cell containing the protein (250 μL ; 135 μM) in a series of $20 \times 2 \mu\text{L}$ injections (first injection 0.4 μL). Error values represent errors of the nonlinear least-squares fit. Buffer conditions used for the titrations were 50 mM HEPES (pH 7.5), 500 mM NaCl.

5.4.1 Reaction monitoring

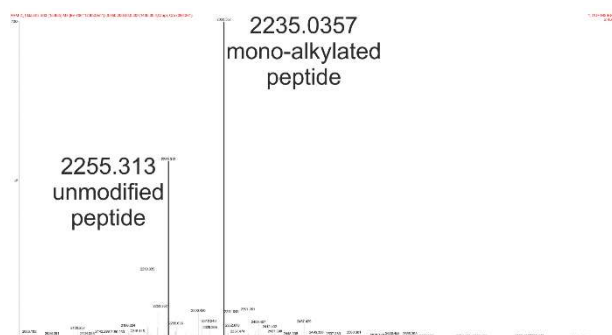


Figure 5.14 Deconvoluted ESI-MS spectra of peptide alkylation reaction obtained by LC/MS. Combined from elution chromatogram and deconvoluted (2000-3000; 1.0 Da resolution). Found 2255.313; unmodified peptide $[M+H]^+$ requires 2255.21. Found 2235.0357; mono-alkylated peptide $[M+H]^+$ requires 2235.24.

Initial attempts to monitor the reaction were performed by liquid chromatography (LC) with tandem mass spectrometry (MS/MS) on a nano LC/MS/MS (Waters NanoAcquity UPLC TOF-MS). The total ion count (TIC) chromatogram showed no distinct peaks, however, deconvoluted spectra identified whole masses corresponding to both the

unmodified and mono-alkylated product in the reaction mixture (Figure 5.15), which was consistent with the anticipated result. Masses corresponding to that of the $[M+4H]^{4+}$ species for the mono-alkylated product were also observed by targeted MS/MS, however, peaks were broad and ill-defined, and attempts to improve resolution by adjusting the elution profile were unsuccessful. To prevent undesirable electrostatic interactions as a result of the highly basic nature of amino acids in the peptide chain, which contained three arginine residues, trifluoroacetic acid (0.1%) in the elution buffer was substituted with ammonia (0.1%), but also had no effect on peak width (chromatogram not shown).

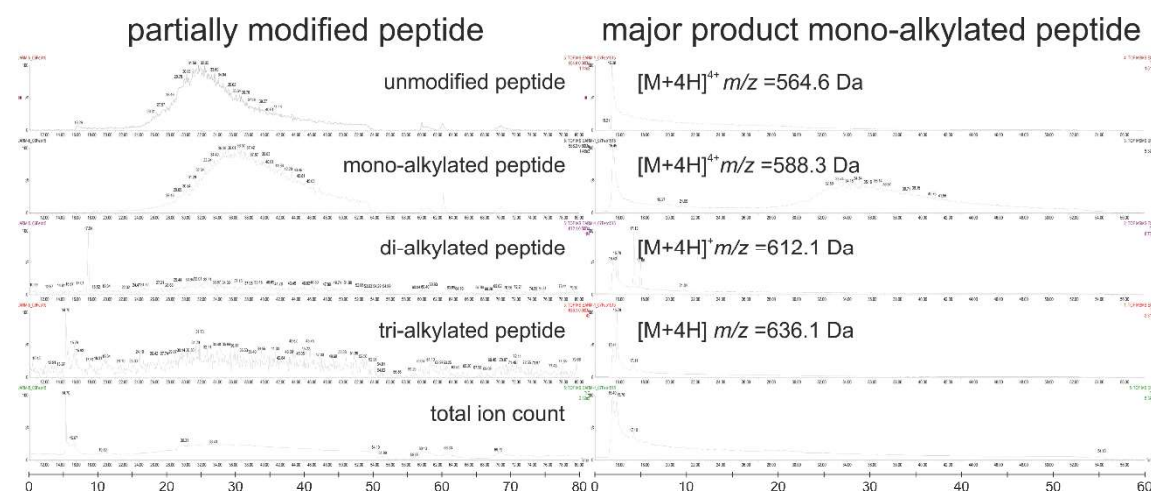


Figure 5.15 Total ion count (TIC) chromatograms, from targeted MS/MS based on the masses of $[M+4H]^{4+}$ ions showing presence of unreacted and mono-alkylated peptide, with undetectable signal from multiply alkylated peptides (species at retention time ~14-17 min were considered to be buffer-related).

Targeted MS/MS analysis was performed to identify the location of alkylation. Isolation of precursor ions in the 4⁺ charge state, followed by collision-induced dissociation (CID), allowed observation of fragmentation ions corresponding to those calculated for peptide bond cleavage of the starting peptide (from precursor *m/z* 564.6) and desired product (from precursor *m/z* 588.3). For the alkylated product these included a-, b-, y- and z-type fragment ions from cleavage adjacent to the alkylated cysteine (Figure 5.16). However, signal-to-noise ratios were not high, and several prominent ions were not assigned, which was likely due to a large overlap of species in the elution profile observed.

Ultimately, although this method was able to confirm that the correct alkylated product was formed, the elution times, relative to the reaction rate, meant that sample time points could not be accurately observed. This method was thus deemed insufficient for reaction monitoring, and additionally had little potential for product separation.

Therefore, it was attempted to determine the conversion to the mono-alkylated product by observation of intact peptide ions by matrix-assisted laser-desorption ionisation–time-of-flight mass spectrometry (MALDI-TOF-MS). This technique would enable rapid observation of reaction samples, avoid the difficulties experienced in liquid chromatography of the peptides, and the soft ionisation technique would allow singly-charged ions to be monitored rather than a selected precursor ionic state. Furthermore, it was hypothesised that the efficiency of ionisation of the peptide was unlikely to be greatly affected by this single modification, since protonation is more likely to occur on basic residues which remained constant, so ratios of peak integrals could be used semi-quantitatively.

Modifications: amidation deacetylation (+) lysine acetylation (+)

Modifications: amidation deacetylation (+) lysine acetylation (+) cysteine isoxazole (+)

Figure 5.16 Fragmentation ions identified by LC/MS/MS. Calculated and observed ions corresponding to a-, b-, y- and z-type ions from fragmentation across peptide bonds and subsequent decarboxylation or deamidation. Observed ions (colour) show deviation from calculated values (Da). Good coverage across the peptide, including ions corresponding to cleavage adjacent to the K12C position show existence of unmodified and mono-alkylated product.

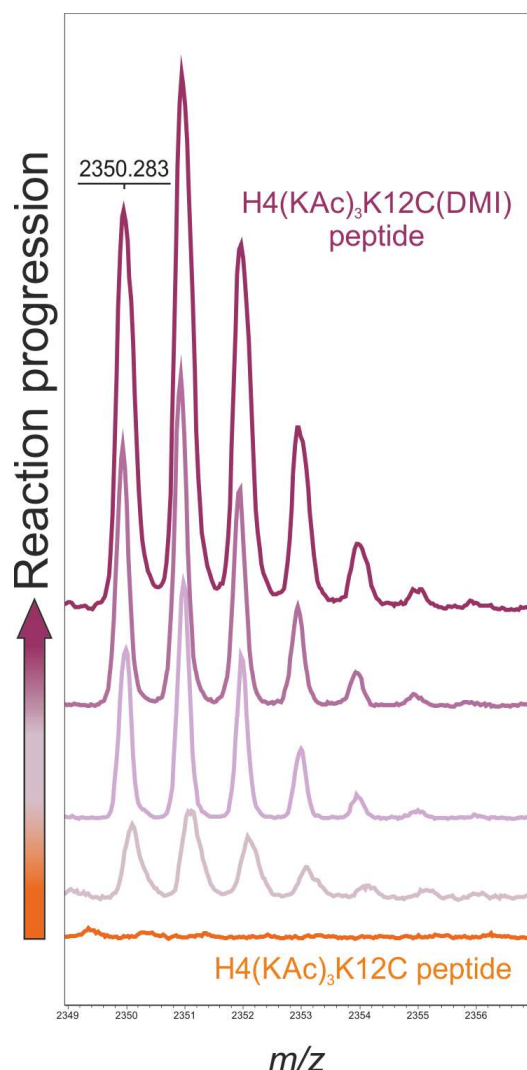


Figure 5.17 MALDI-TOF spectra showing increasing intensity of peak at 2350 m/z corresponding to the intact mass of H4K12CDMI peptide (purple), which is not observed in the starting peptide (orange). Good resolution shows product isotopic peaks with separation of one atomic mass unit.

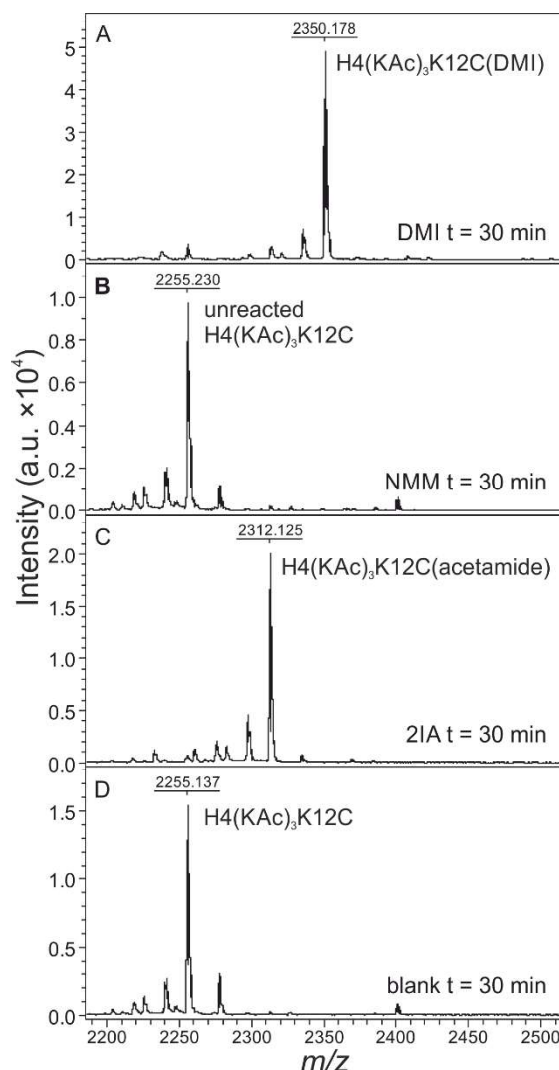


Figure 5.18 Selected regions of MALDI-TOF mass spectra obtained by for alkylation with different alkylating agents. Samples taken from reaction mixtures [1 mM H4(KAc)₃K12C, 1 M CHES (pH 10), 4 M guanidine-hydrochloride, 10 mM methionine, 10 mM TCEP, 5 mM alkylating agent] after 30 minutes, comparing alkylation with: A, 5-(chloromethyl)-3-methylisoxazole (DMI); B, N-methylmaleimide (NMM); C, 2-iodoacetamide (2IA); D, buffer.

Gratifyingly, initial observations of the reaction course confirmed our hopes for this technique.

Spectra for the unreacted H4K12C peptide (**19**) contained a single dominant peak at 2255 m/z . A time-course experiment following the alkylation reaction with Cl-DMI showed clear evolution of a peak at 2350 m/z , corresponding to the mono-alkylated peptide (**20**), which could be used to monitor reaction progression (Figure 5.17).

Several steps were taken to ensure that the peak under observation was the desired reaction product (Figure 5.18). The peptide was subjected to the same reaction conditions in the absence of an alkylating agent, to confirm its stability. Additionally, alkylation with Cl-DMI was contrasted with

the use of 2-iodoacetamide (2IA), a commonly used alkylating agent for mass spectrometry studies, and *N*-methylmaleimide (NMM), a less reactive maleimide derivative shown to be a selective alkylating agent for thiols.^{32,33} Observation of these products in reaction mixtures would validate this method of monitoring the reaction, and provide some insight into the relative efficiency of alkylation. Under conditions tested, 2IA caused almost complete conversion of the H4K12C peptide (**19**) into the alkylated product (**21**) within the first sampling interval: a mass shift of +57 Da was observed, corresponding to addition of the acetamide, with the percentage of product peak integrals (2312.19 Da) compared to the starting material (2254.66 Da) 93.9% after 5 min. NMM, however, formed negligible amounts of alkylated product, with no peak observed at the expected mass of 2366.24 Da within the duration of experiment (2 h, Figure 5.18B). There was no mass shift for the peptide under the reaction conditions alone, confirming that alkylation observed was indeed specific to the CI-DMI agent.

5.4.2 Alkylation optimisation

Reaction conditions were optimised for a fast, efficient, selective method of alkylation by varying the reducing conditions, pH, temperature, and number of equivalents of alkylating agent (Table 5.2). Efficiency was assessed by the percentage ratio of integrals of peaks corresponding to H4(KAc)₄K12C (**19**) and H4(KAc)₄K12C(DMI) (**20**): $100\% \times \frac{I_{2350}}{(I_{2350} + I_{2255})}$. In every case, these were the highest intensity peaks in the spectra.

Since reduction of any disulfides to cysteine must be maintained to allow thiolate formation, we investigated the effect the number of equivalents of the TCEP reducing agent may have on rate of reaction. Analysis of reaction products containing 0-10 equivalents of TCEP (conditions a-d) suggested that optimal conditions resulted from use of 5 equivalents.

Table 5.2 Results from reaction condition optimisation. Reaction mixture (20 μ L) containing 1 mM H₄(KAc)₃K12C, 1 M buffer, 4 M guanidine·hydrochloride, 10 mM methionine was prepared in buffer with TCEP and Cl-DMI as shown in the table. Note that condition a) was selected as the locus for exhaustive local optimisation, and is listed several times in the table for ease of comparison, with variations highlighted in grey. n.d. value not determined. [†]ratio after 60 min; *ratio after 16 h.

Variable	Condition	TCEP (eq)	Buffer (pH)	T (°C)	Cl-DMI (eq)	% ratio (120 min)
eq. of reducing agent	a	10	CHES (10)	25	1	41*
	b	5	CHES (10)	25	1	79
	c	2.5	CHES (10)	25	1	71
	d	0	CHES (10)	25	1	26
pH of reaction and buffer dependence	e	10	HEPES (7)	25	1	3
	f	10	HEPES (8)	25	1	n.d
	g	10	CHES (8)	25	1	38
	h	10	CHES (9)	25	1	58
	a	10	CHES (10)	25	1	41*
	i	10	CAPS (10)	25	1	47 [†]
	j	10	CAPS (11)	25	1	59
temperature	k	10	CHES (10)	4	1	17*
	a	10	CHES (10)	25	1	41*
	l	10	CHES (10)	42	1	78
eq. of alkylating agent	m	10	CHES (10)	25	0.5	34
	a	10	CHES (10)	25	1	41*
	n	10	CHES (10)	25	2	74
	o	10	CHES (10)	25	5	92

Free cysteine has a pK_a ~9, however, this value can vary depending on its local environment: it has been shown that exposed cysteines on proteins often exhibit a lower pK_a value of ~7.5.³⁴ Slight increases in pH within this range will affect cysteine protonation state, and thus the population of exposed thiolates, which have a far higher nucleophilicity. This feature was considered important to control selectivity, especially in the context of the H4 peptide containing other potentially nucleophilic amino acids: the C-terminal free lysine (pK_a ~10.5), and the N-terminal tyrosine

(pKa ~10.5). Alkylation was studied between pH 7-11, using HEPES, CHES or CAPS buffer, with overlap between the pH ranges to control for any effects caused by buffer selection rather than directly a result of pH (conditions a,e-j). Samples prepared in HEPES buffer could not be well observed by MALDI-MS: high concentrations of this buffer apparently interacted with the CHCA matrix preventing crystallisation, and therefore ionisation of the sample. However, at low pH conditions there was evidence of minimal alkylation, with only 3% conversion after 60 min at pH 7. CHES buffer at pH 9 resulted in the highest level of alkylation after 120 min (58%), which was as predicted from this value lying between the pKas of cysteine and the competitive tyrosine and lysine residues. These conditions would presumably promote maximal thiolate formation, without loss of selectivity from phenolate or neutral primary amine formation.

Although levels of alkylation seemed to increase with higher temperatures (conditions a,k,l), after 16 h no product peak was observed by MALDI, and it is likely that peptide decomposition at this temperature may be problematic, especially when attempting to extend this methodology towards alternative, more complex protein based systems. The degree of alkylation at 25 °C seemed sufficient, and was taken as an appropriate balance between rate and stability.

Excess thionyl chloride used in the synthesis of the alkylating agent was removed *in vacuo*, however, volatility of the desired product prevented use of pressures <300 mbar. Although this would be rapidly hydrolysed in buffer, it is likely that residual chlorinating agent reduces the stoichiometry

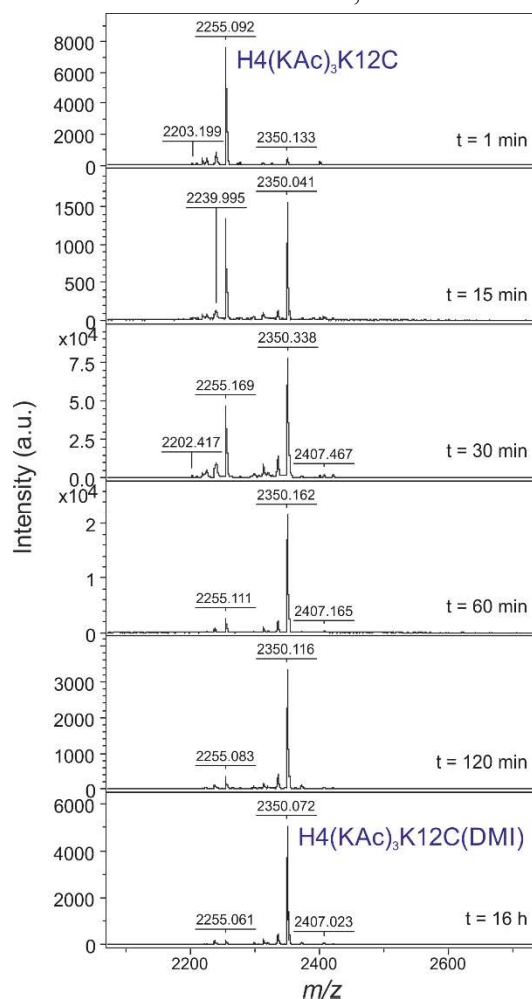


Figure 5.19 Regions of spectra obtained by MALDI-TOF following alkylation reaction under optimised conditions. Samples taken from reaction mixtures at time intervals shown in bottom right of spectra: 1 mM H4(KAc)₃K12C, 4 M guanidine-hydrochloride, 10 mM methionine, 5 mM TCEP, 1 M CHES (pH 9), 5 mM Cl-DMI. Time course shows conversion from starting peptide (observed *m/z* 2255.09; expected *m/z* 2255.21) to alkylated H4(KAc)₃K12C(DMI) peptide (observed *m/z* 2350.07; expected *m/z* 2350.24).

predicted by molecular mass calculations. Assuming a 50% net content of Cl-DMI by mass, it was then investigated how many equivalents of alkylating agent were required for efficient, selective mono-alkylation (conditions a, m-o). Under MALDI analysis after 120 min, the percentage ratio of 2350/2255 m/z increased approximately linearly with the number of equivalents of alkylating agent, and over-alkylated products (expected at 2312.22 m/z) were never observed. Seeking to drive the reaction towards over-alkylating conditions, in order to definitively demonstrate mono-alkylation, the number of equivalents was further increased. Peaks with the greatest intensity were still observed at 2255 and 2350 m/z , without any emergence of products at 2312 m/z which may correspond to multiply alkylated peptides. Alkylated product ratios never exceeded 95% after 120 min, which was obtained with 5 eq, and decreased to 87% for 15 eq.

Under optimised conditions (1 mM H4(KAc)₃K12C, 4 M guanidine·hydrochloride, 10 mM methionine, 5 mM TCEP, 1 M CHES (pH 9), 5 mM Cl-DMI) the reaction proceeded in a peak-to-peak fashion, and mono-alkylated peptide became the dominant species after 30 min. After 60 min the percentage ratio of this peak, relative to the starting peptide, was 90%, and after 16 h 96% was observed as mono-alkylated peptide (Figure 5.19).

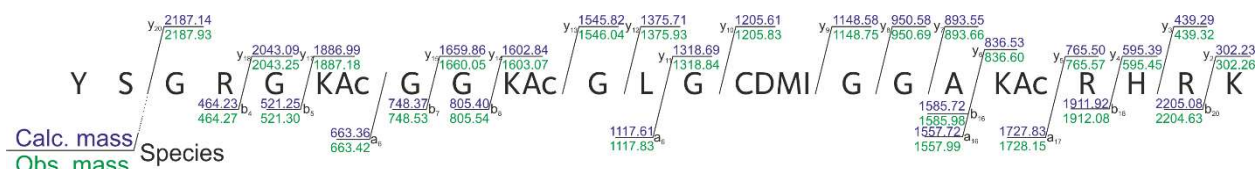


Figure 5.20 MALDI-TOF-TOF fragmentation species observed after alkylation confirmed the site of alkylation as that of the cysteine residue (CDMI). Labels indicate a-, b-, or y--type ion and position. Calculated mass (Da) – blue; observed mass (Da) – green.

Tandem mass spectrometry (MS/MS), with laser-induced fragmentation (LIFT), confirmed the site of alkylation as that of the cysteine residue (Figure 5.20). Crucially, the characteristic peak resulting from amide bond cleavage adjacent to the alkylated cysteine was observed at 1148.75 m/z (expected m/z 1148.58) and was not present in the starting peptide. Good coverage of y-type ions were detected along the full length of the peptide, with a- and b-type ions towards both the N- and C-terminus.

In order to confirm the quantification of peak intensities, and the hypothesis of similar ionisation efficiencies for unmodified and mono-alkylated peptides, integral ratios were compared by high performance liquid chromatography (HPLC). End-points of reaction mixtures were diluted 1:4 into reaction buffer (1 M CHES, 4 M guanidine·hydrochloride, 10 mM methionine; pH 7.5) and injected onto a reversed phase Dionex Acclaim® C18 column on a PerkinElmer Flexar HPLC system.

This setup was able to separate reactant and product peptide peaks, and also identify further products not previously observed by MALDI-TOF (Figure 5.21). The same alkylating agents were used (buffer, 2IA, NMM, and Cl-DMI), and afforded peaks eluting at detectably different retention times, including a small amount of alkylation by NMM, forming cysteine-succinamide (NMS), which was not detected by MS alone. Furthermore, alkylation with an excess of Cl-DMI (5 eq, which eluted after the 10 min gradient at 100% B) was shown to produce a second peak of an unknown species. Since a mass corresponding to polyalkylation was never detected, it was assumed that this is likely a product of misalkylation, i.e. substitution of the

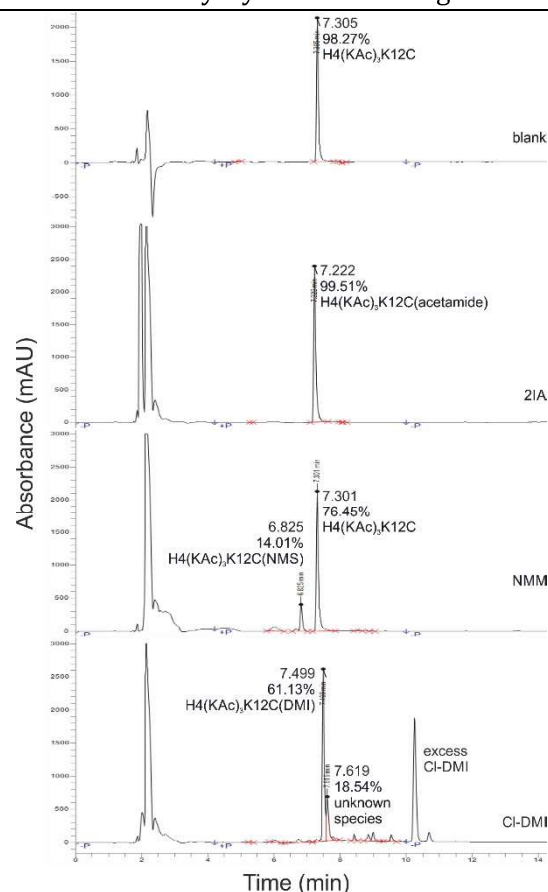


Figure 5.21 Comparison of products from H4(KAc)₃K12C with various alkylating agents showing specific species elution at different retention times. Reaction mixtures after 2 h: 1 mM H4(KAc)₃K12C, 4 M guanidine·hydrochloride, 10 mM methionine, 5 mM TCEP, 1 M CHES (pH 9). Alkylating agent shown bottom right: buffer; 2-iodoacetamide (2IA); *N*-methylmaleimide (NMM); 5-(chloromethyl)-3-methylisoxazole (Cl-DMI).

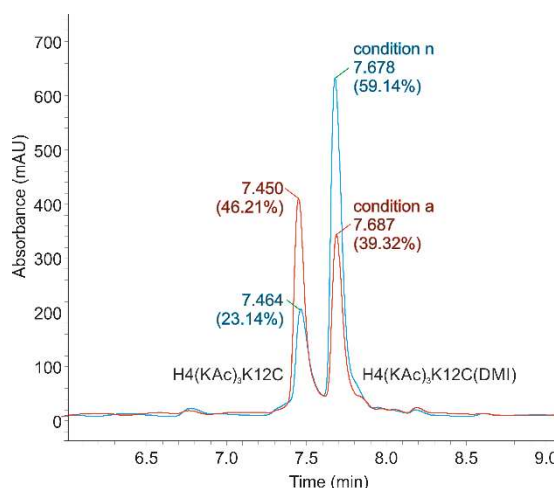


Figure 5.22 Region of interest in HPLC trace of partial alkylation conditions showing conversion from H4(KAc)₃K12C (7.5 min) to H4(KAc)₃K12C(DMI) (7.7 min). Reaction: 1 mM H4(KAc)₃K12C, 4 M guanidine·hydrochloride, 10 mM methionine, 5 mM TCEP, 1 M CHES (pH 9). Condition a: 1 mM Cl-DMI; product peak integral ratio 45.97% (41.04% by MALDI). Condition n: 2 mM Cl-DMI; product peak integral ratio 71.88% (74.42% by MALDI) Samples injected after 16 h.

isoxazole at a position other than K12C. Gratifyingly, integral ratios showed good correlation with those observed by MALDI-MS, exemplified by comparison of partial reaction mixtures from conditions a and n (Figure 5.22).

In order to further investigate overalkylation, the starting peptide was subjected to high concentrations of Cl-DMI and analysed by HPLC. With increasing equivalents of Cl-DMI several unknown impurities were formed with retention times of >6 min, the most significant eluting at a slightly higher retention time (7.8 min) than the H4(KAc)₃K12C(DMI) peptide (7.7 min).

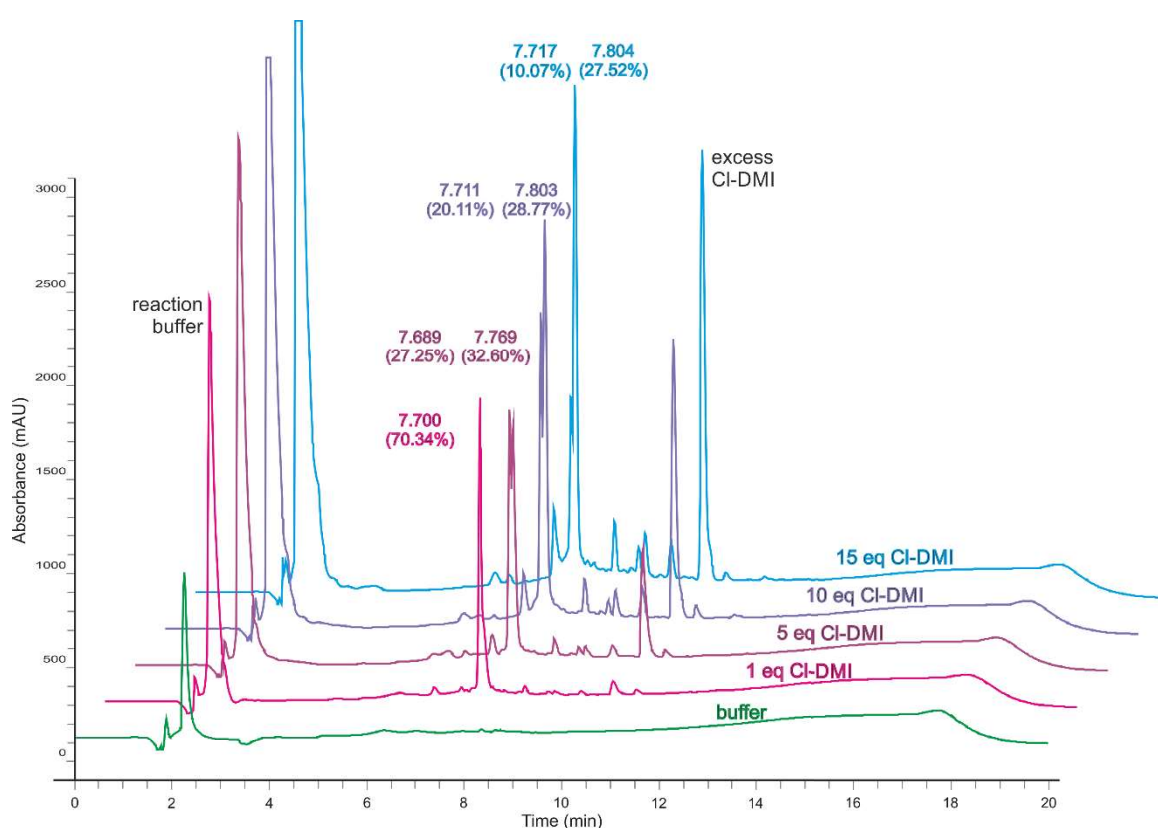


Figure 5.23 Full elution profile from analytical-HPLC of alkylation reaction with large excess of Cl-DMI to observe overalkylation. Offset spectra demonstrate selective alkylation to single peak at 7.7 min with the use of 1 eq, and increasing proportions of an undesired product at 7.8 min with greater numbers of equivalents, as well as multiple small peaks in the range of 6–10 min. Gradient elution 0–100% B, 2–10 min (eluent A: 95% H₂O, 5% MeCN, 0.1% TFA; eluent B: 5% H₂O, 95% MeCN, 0.1% TFA). Large peaks at the start of elution are caused by high concentrations of reaction buffer; excess Cl-DMI eluted >10 min at 100% B.

5.4.3 Isolation and affinity measurements

Since it was demonstrated that the starting peptide and the alkylated product were observable at different retention times by HPLC, once the reaction had been optimised to a sufficient yield (96% product/starting peptide integral ratio by MALDI-MS; 70% purity by analytical HPLC), a time-course experiment was carried out to simultaneously monitor and isolate the resulting peptide.

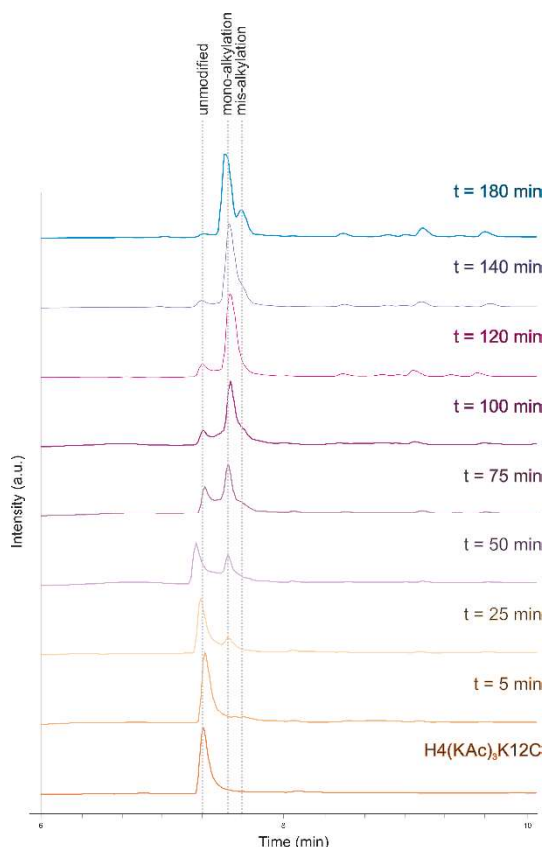


Figure 5.24 Time course of peptide alkylation reaction monitored by HPLC. UV absorption at 220 nm detected by analytical HPLC, time points indicated bottom right, with starting peptide H4(KAc)₃K12C included for comparison. Reaction progression shows generation of desired mono-alkylated peptide at a retention time of 7.47 min followed by mis-alkylation after 140 min.

Alkylation proceeded in line with previous studies, with the desired product as the dominant peak after 75 min (65%) and with a high conversion after 140 min (77%). Attempts to remove residual starting material, by addition of extra equivalents after 140 min, resulted in mis-alkylation (7.605 min, 19%).

Peaks corresponding to the desired product were manually collected from the analytical HPLC system and reinjected to ensure that they were isolated as a single species.

Combined, isolated H4(KAc)₃K12C(DMI) (**20**) from multiple injections was lyophilised and re-dissolved in 50 mM HEPES, 500 mM NaCl (pH 7.5), and subsequently observed as a single peak by analytical HPLC at a retention time of

7.083 min; (a second peak at ~9.5 min was present in both this sample and the buffer blank, and was attributed to a contaminant on the system). Estimation of peptide content by UV absorbance (Nanodrop® spectrophotometer), allowed this solution to be used in the previously described AlphaScreen assay to quantify the affinity.

Alkylated peptide containing a thioether-linked dimethyl isoxazole, was isolated in sufficient quantity for preparation of a 1.5 mM solution for a dilution series. AlphaScreen emission data for this range spanned the inhibition curve well for all peptides evaluated, and the standard error for curve fitting was $\leq 10\%$ of the IC_{50} values obtained. Although the control H4(KAc)₄ peptide had a higher affinity than previously detected (IC_{50} 5.5 μ M c.f. 15.7 μ M), there was high correlation of quantitative values for K12 positional modifications (previous IC_{50} values for alkyl-, ether-, and amide-DMI were 13.3 μ M, 10.7 μ M, and 6.8 μ M, respectively).

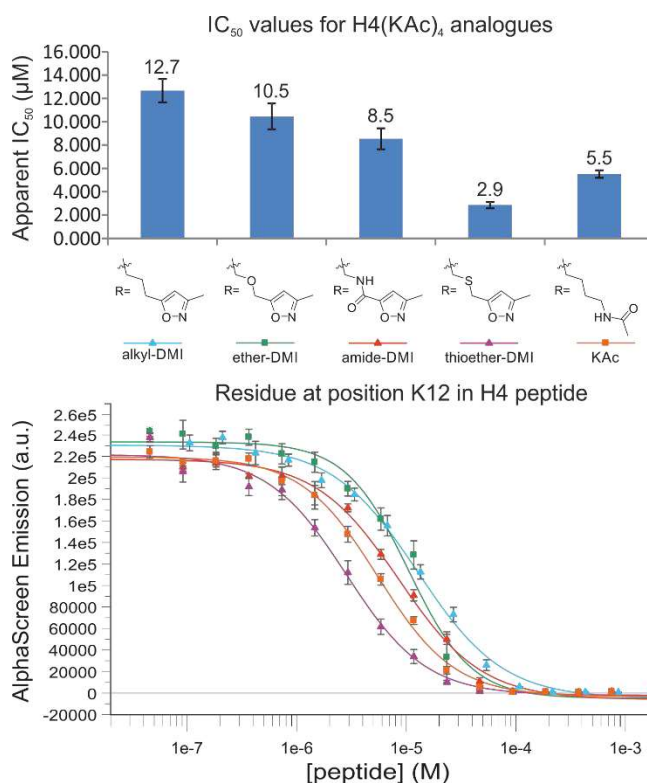


Figure 5.25 IC_{50} values for H4(KAc)₄ analogue peptides bearing isoxazole containing non-natural amino acid substitutions at K12. Data obtained from serial dilutions in AlphaScreen assay inhibiting the interaction of BRD4 with biotinylated H4(KAc)₄: values in μ M above bars; identity of substitution below; error bars shown as the standard error of the fit. Results are the average of readings taken in triplicate, with fitting performed as IC_{50} - 4 Parameter Fit in GraFit. AlphaScreen emission readings with fitting overlaid are shown below.

The thioether-linked DMI not only demonstrated good inhibitory activity of BRD4(1), the IC_{50} value was the lowest measured in this assay. This confirms the ability of this isoxazole-containing non-natural amino acid to mimic the interactions of KAc with bromodomains, and the potential to increase affinity to peptides naturally bearing this modification.

5.6 Conclusions

In conclusion, several isoxazole containing non-natural amino acids were identified as acetyl lysine mimics in their unprotected monomeric form. After incorporation of these amino acids into H4-like peptides, measurements of affinities to BRD4(1) confirmed the potential to form a stronger interaction with bromodomains than the cognate acetylated form, and additionally demonstrated a strong positional dependence. Substitutions of the KAc in positions 12 or 16 often resulted in peptides with comparable, if not greater affinity than the tetra-acetylated peptide.

The increased affinity as a result of isoxazole integration at the K12 position was applied to a cysteine mutant H4(KAc)₃K12C peptide, in a “tag-and-modify” approach, involving alkylation with 5-(chloromethyl)-3-methylisoxazole. Site-selective, *in situ* mono-alkylation was achieved under mild conditions in high yields after 120 min, which could be monitored by both MALDI-TOF MS and analytical reverse phase HPLC (>90% 2350/2255 *m/z* peak integral ratio by MALDI; >70% product at retention time 7.7 min by analytical HPLC). Regiochemistry of the alkylation was confirmed by tandem MS/MS (nanoLC/MS/MS and MALDI/MS/MS with LIFT fragmentation). Isolation of the alkylated peptide enabled comparison of its affinity to other K12 mutants, and the measured IC₅₀ was approximately half that of the cognate H4(KAc)₄ control peptide.

An X-ray crystal structure or NMR structure of the peptide bound to a bromodomain would confirm that the modified amino acid mimics KAc, and might provide a rationale for positional selectivity. A suitable amino acid-based acetyl lysine mimic could be employed as an irreversible mark on histone proteins in order to probe bromodomain biology *in vivo*. If these amino acids are neither substrates nor high-affinity competitive inhibitors of HDACs, such KAc mimics would be useful for the study of lysine acetylation in cell extracts without the complicating effects of deacetylation.

5.7 References

- (1) Dawson, P. E., W., M. T., Clark-Lewis, L., and Kent*, S. B. H. (1994) Synthesis of proteins by native chemical ligation. *Science* (80-.). 266, 34–45.
- (2) Muir, T. W., Sondhi, D., and Cole, P. a. (1998) Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6705–6710.
- (3) Chalker, J., Bernardes, G., and Davis, B. (2011) A “tag-and-modify” approach to site-selective protein modification. *Acc. Chem. Res.* 44, 730–741.
- (4) Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* 4, 232–4.
- (5) Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D., and Chin, J. W. (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* 36, 153–63.
- (6) Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* 128, 1003–12.
- (7) Huang, R., Holbert, M. A., Tarrant, M. K., Curtet, S., Colquhoun, D. R., Dancy, B. C. B. M., Dancy, B. C. B. M., Hwang, Y., Tang, Y., Meeth, K., Marmorstein, R., Cole, R. N., Khochbin, S., and Cole, P. A. (2010) Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J. Am. Chem. Soc.* 132, 9986–9987.
- (8) Chatterjee, C., and Muir, T. W. (2010) Chemical approaches for studying histone modifications. *J. Biol. Chem.* 285, 11045–50.
- (9) Dhall, A., and Chatterjee, C. (2011) Chemical Approaches To Understand the Language of Histone. *ACS Chem. Biol.* 6, 987–999.
- (10) Dawson, M. A., Kouzarides, T., and Huntly, B. J. P. (2012) Targeting Epigenetic Readers in Cancer. *N. Engl. J. Med.* 367, 647–657.
- (11) French, C. A., Miyoshi, I., Kubonishi, I., Grier, H. E., Perez-atayde, A. R., and Fletcher, J. A. (2003) BRD4-NUT Fusion Oncogene : A Novel Mechanism in Aggressive Carcinoma. *Cancer Res.* 63, 304–307.
- (12) Zuber, J., Shi, J., Wang, E., Rappaport, A. R., Herrmann, H., Sison, E. a., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., Taylor, M. J., Johns, C., Chicas, A., Mulloy, J. C., Kogan, S. C., Brown, P., Valent, P., Bradner, J. E., Lowe, S. W., and Vakoc, C. R. (2011) RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* 478, 524–528.
- (13) Muller, S., Filippakopoulos, P., and Knapp, S. (2011) Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* 13, e29.
- (14) Jung, M., Philpott, M., Müller, S., Schulze, J., Badock, V., Eberspächer, U., Moosmayer, D., Bader, B., Schmees, N., Fernández-Montalván, A., and Haendler, B. (2014) Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem.* 289, 9304–19.
- (15) Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C. H., Knapp, S., and Mueller, S. (2012) Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* 149, 214–231.
- (16) Philpott, M., Yang, J., Tumber, T., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Keates, T., Felletar, I., Ciulli, A., Knapp, S., and Heightman, T. D. (2011) Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* 7, 2899–2908.
- (17) Vidler, L. R., Filippakopoulos, P., Fedorov, O., Picaud, S., Martin, S., Tomsett, M., Woodward, H., Brown, N., Knapp, S., and Hoelder, S. (2013) Discovery of novel small-molecule

inhibitors of BRD4 using structure-based virtual screening. *J. Med. Chem.* 56, 8073–88.

(18) Brand, M., Measures, A. R., Wilson, B. G., Cortopassi, W. A., Alexander, R., Höss, M., Hewings, D. S., Rooney, T. P. C., Paton, R. S., and Conway, S. J. (2014) Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chem. Biol.* 10, 22–39.

(19) Hewings, D. S., Wang, M., Philpott, M., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Vuppusetty, C., Marsden, B., Knapp, S., Conway, S. J., and Heightman, T. D. (2011) 3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands. *J. Med. Chem.* 54, 6761–6770.

(20) Dawson, M. A., Prinjha, R. K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W.-I., Robson, S. C., Chung, C., Hopf, C., Savitski, M. M., Huthmacher, C., Gudgin, E., Lugo, D., Beinke, S., Chapman, T. D., Roberts, E. J., Soden, P. E., Auger, K. R., Mirguet, O., Doehner, K., Delwel, R., Burnett, A. K., Jeffrey, P., Drewes, G., Lee, K., Huntly, B. J. P., and Kouzarides, T. (2011) Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* 478, 529–33.

(21) Bamborough, P., Diallo, H., Goodacre, J. D., Gordon, L., Lewis, A., Seal, J. T., Wilson, D. M., Woodrow, M. D., and Chung, C.-W. (2012) Fragment-based discovery of bromodomain inhibitors part 2: optimization of phenylisoxazole sulfonamides. *J. Med. Chem.* 55, 587–96.

(22) Hewings, D. S., Fedorov, O., Filippakopoulos, P., Martin, S., Picaud, S., Tumber, A., Wells, C., Olcina, M. M., Freeman, K., Gill, A., and Others. (2013) Optimization of 3, 5-Dimethylisoxazole Derivatives as Potent Bromodomain Ligands. *J. Med. Chem.* 56, 3217–3227.

(23) Falorni, M., Giacomelli, G., and Spanu, E. (1998) Synthesis of new α -amino- acids containing the isoxazole moiety. *Tetrahedron Lett.* 39, 9241–9244.

(24) De Luca, L., Giacomelli, G., and Riu, a. (2001) Solid-phase synthesis of isoxazole-based amino acids: A new scaffold for molecular diversity. *J. Org. Chem.* 66, 6823–6825.

(25) Chung, C.-W., and Witherington, J. (2011) Progress in the discovery of small-molecule inhibitors of bromodomain--histone interactions. *J. Biomol. Screen.* 16, 1170–85.

(26) Loharch, S., Bhutani, I., Jain, K., Gupta, P., Sahoo, D. K., and Parkesh, R. (2015) EpiDBase: a manually curated database for small molecule modulators of epigenetic landscape. *Database* 10.1093/database/bav013.

(27) Leatherbarrow, R. J. (2009) GraFit Version 7. Erithacus Software Ltd. Horley, U.K.

(28) Falconer, R. J., Penkova, A., Jelesarov, I., and Collins, B. M. (2008) Survey of the year 2008: applications of isothermal titration calorimetry. *J. Mol. Recognit.* 23, 395–413.

(29) Tsai, W.-W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M.-C., Patel, D. J., and Barton, M. C. (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927–932.

(30) Lucas, X., Wohlwend, D., Hügler, M., Schmidtkunz, K., Gerhardt, S., Schüle, R., Jung, M., Einsle, O., and Günther, S. (2013) 4-Acyl pyrroles: mimicking acetylated lysines in histone code reading. *Angew. Chem. Int. Ed. Engl.* 52, 14055–9.

(31) Simon, M. D., and Shokat, K. M. (2012) A method to site-specifically incorporate methyl-lysine analogues into recombinant proteins. *Methods Enzymol.* 1st ed. Elsevier Inc.

(32) Luche, S., Diemer, H., Tastet, C., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T. (2004) About thiol derivatization and resolution of basic proteins in two-dimensional electrophoresis. *Proteomics* 4, 551–561.

(33) Paulech, J., Solis, N., and Cordwell, S. J. (2013) Characterization of reaction conditions providing rapid and specific cysteine alkylation for peptide-based mass spectrometry. *Biochim. Biophys. Acta - Proteins Proteomics* 1834, 372–379.

(34) Marino, S. M., and Gladyshev, V. N. (2010) Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *J. Mol. Biol.* 404, 902–916.

Chapter 6: Conclusions

The aim of this study was to gain insight into mechanisms of epigenetic control for the tandem reader domains of TRIM24 and TRIM33. This was achieved through a combination of biological, computational and chemical methods, which resulted in: increased knowledge of affinity for cognate binding partners; identification of novel ligands for these homologous proteins; further structural understanding of TRIM24, including attempts to contextualise readership of nucleosome modifications; and methods for increasing affinity of bromodomains towards peptides bearing KAc-mimicking residues.

Data presented in Chapter 2 suggested that the PHD finger and bromodomain of TRIM24 and TRIM33 operate co-operatively, with a very strong increase in affinity towards peptides bearing multiple modifications. The TRIM24 and TRIM33 PHD-bromodomain cassettes were expressed with an intact affinity label suitable for indirect assay development, and an AlphaScreen assay was optimised to quantify affinity for modified H3-like peptides. The differential recognition preferences of TRIM24 and TRIM33 were profiled by determining affinity of these tandem domains towards various combinations of modifications. Both TRIM24 and TRIM33 demonstrated a significant increase in binding interaction with peptides bearing the K9Me₃ modification, in addition to acetylation. Since the domains can also function independently, i.e. recognition of either acetylation or trimethylation in isolation, this suggests that combined readout of histone modifications could result in alternative downstream effects.

In Chapter 3, a computational docking system was developed for TRIM24 and TRIM33 to create an enriched screening library of 200 compounds, based on the Maybridge fragment screening library (Fisher Scientific). The AlphaScreen format developed for peptide recognition profiling in Chapter 2, was then employed to identify inhibitors of histone peptide recognition for TRIM24 and TRIM33. No known screening capacity previously existed for TRIM33, and at the time of commencing this work, no small-molecule ligands were known for either target. High-throughput small-molecule screening was carried out on both the bespoke enriched library (Maybridge), and

the PPI-Net compound collection of 1,536 compounds designed to inhibit PPIs, in order to identify potential novel ligands. Validation of hit molecules, through several orthogonal assays, identified four promising lead compounds for probe development from the Maybridge collection, and eight molecules from the PPI-net collection. Ligands were identified which demonstrated selectivity of inhibition for TRIM24 or TRIM33. Moreover, some compounds demonstrated inhibition of either methyl- or acetyl-lysine binding, suggesting a level of domain selectivity.

Chapter 4 detailed cloning of *TRIM24* and *TRIM33* genes into alternative expression vectors, with increased stability after cleavage, which allowed expression and purification of unlabelled TRIM24 and TRIM33, with sufficient purity and structural homogeneity for crystallisation. A highly repeatable procedure was outlined for obtaining single-crystals of TRIM24, which can be utilised in compound soaking of previously identified hits: crystal structures containing bound small-molecules would enable structure-guided optimisation of probes. A high resolution crystal structure of *apo*-TRIM24 was obtained, and nucleosome association studies demonstrated binding of TRIM24 to histone modifications in the context of a nucleosome.

In Chapter 5, the previously developed AlphaScreen assay was transferred to the BRD4(1)-H4(KAc)₄ system, and used to assess the ability of isoxazole containing non-natural amino acids to inhibit peptide recognition. Integration of three of these amino acids as replacements for KAc residues at different positions of H4(KAc)₄ peptides, resulted in peptides with comparable, if not greater affinity than the tetra-acetylated peptide. This study confirmed the potential to form a stronger interaction with bromodomains than the cognate acetylated form, and additionally demonstrated a strong positional dependence of substitutions. An isoxazole-based alkylating agent was optimised to selectively modify a cysteine residue *in situ*, under mild conditions and in high yields. Use of this agent on an H4(KAc)₄-like peptide bearing a K12C substitution resulted in increased affinity for the BRD4(1) bromodomain: the measured IC₅₀ of an alkylated peptide was approximately half that of the cognate H4(KAc)₄ control peptide. This will facilitate screening of bromodomain targets for which binding partners are unknown, or of prohibitively low affinity for assay development.

Collectively, this work has provided fundamental information concerning tandem epigenetic reader domains, in a field of multivalent recognition which could have broad reaching relevance for biological systems. Epigenetic reader domains have commonly been studied in isolation, and consideration of wider protein architecture and subsequent recognition profiles could enable identification of more subtle downstream effects of recognition. To extend the metaphor, it would enable the study of ‘reading’ phrases rather than words, and distinguish between phrases such as ‘they left’ and ‘he left’, rather than simply ‘left’. It could be envisioned that inhibition of multiple reader domains would be beneficial for therapeutically efficacious compounds, which would have the advantage of a specific mode of action, rather than an inhibition target. The combined recognition of multiple modifications, as demonstrated here for TRIM proteins, increases the capacity for transmission and storage of information without the requirement for novel reader domains, which can be evolutionarily rationalised. Furthermore, it presents a possible contributing factor to mechanisms of epigenome maintenance through mitosis, whereby cross talk within nucleosome cores formed from random recombination of modified parental histone proteins can re-establish a pre-existing state. Moving forwards, the study of higher order systems, containing several reader or effector domains, and generation of a toolkit of compounds for specific inhibition, could cause a paradigm shift in the field of epigenetics.

Chapter 7: Materials and methods

7.1 Biological methods

7.1.1 Materials and reagents

Materials were obtained from the following suppliers:

Material	Supplier
Agarose	Bioline
SDS-PAGE reagents	Bio-Rad and Sigma
T4 DNA ligase and PfuTurbo® DNA polymerase	Stratagene
Q5® High-Fidelity DNA Polymerase	New England Biolabs (NEB)
IPTG and kanamycin	LaserBio Labs
dNTPs for PCR	Promega
HEPES Free Acid >99.5 %	Apollo Scientific
CHAPS	Calbiochem
AlphaScreen Histidine (NI chelate detection kit)	Perkin Elmer
384-well proxiplates plus	Perkin Elmer
Thermowell Sealing Tape	Corning

All other chemicals were obtained from Sigma-Aldrich unless otherwise stated. Purified (Milli-Q) water was obtained from a Millipore Elix® Reverse Osmosis system which was further purified by a Millipore Milli-Q® Synthesis system with a 0.22 µm filter on the outlet. Reagent quantities given as weight per volume (w/v) correspond to g/mL unless otherwise stated.

Plasmids

The pNIC28-Bsa4 TRIM24 (TIF1α) construct was kindly provided by the SGC, containing His₆-TRIM24D872 (PHD-Bromo). The pGEX-4T1 TRIM33 (TIF1γ) construct was kindly provided by Prof. Joan Massagué, containing Gst-TRIM33 (PHD-Bromo). SENP1 in a pNIC28-Bsa4 vector was kindly provided by the SGC.

7.1.2 General methods

pH measurements

The pH of solutions was determined using a Jenway pH Meter 3305, with an Aldrich glass/calomel combination electrode. Calibration was carried out between pH 4.0-7.0 or 7.0-10.0 immediately before use, with buffer solutions of phthalate (pH 4.0), phosphate (7.0) and borate (pH 10.0) from Fischer Scientific. Electrodes were stored in a 4 M aqueous potassium chloride solution.

Centrifugation

Samples of volume less than 1.5 mL were centrifuged at room temperature using an accuSpin™ benchtop centrifuge (Fischer Scientific) at 13,000 rpm. Protein samples of up to 2.0 mL were centrifuged in a Beckman Microfuge 22R Centrifuge (14,000 rpm) at 4 °C. Cell growth media and cell lysates were centrifuged in a Beckman Avanti J-25 (8500 rpm: JA-10 rotor; 15,000 rpm: JA-16.250 rotor and 24,000 rpm: JA-25.50 rotor, respectively). Proteins were concentrated using a Beckman Allegra X-30R centrifuge (4000 rpm, SX4400 swing bucket) at 4 °C. Plates were centrifuged using a Beckman Allegra X-30R centrifuge (2000 rpm, S6096 rotor) at 25 °C.

7.1.3 Microbiology

Standard sterile practices were followed thoroughly, using the Heracleus laminar flow hood when necessary. Media and equipment were sterilised by autoclaving at 121 °C for 20 min. Solutions of IPTG, antibiotics and other labile compounds were sterilised by filtering through 0.2 µm filters (Minisart®, Sartorius Stedim). Final concentrations of antibiotics unless indicated otherwise were: Ampicillin 100 µg/mL; Kanamycin 30 µg/mL.

Agarose gel electrophoresis

DNA samples were analysed and purified by gel electrophoresis of the DNA mixture on 1% agarose gels prepared with SYBR Safe™ DNA gel stain (Invitrogen). Samples were loaded after mixing with 6× loading buffer and run in TAE buffer on a Bio-Rad system at a constant potential of 80 V until desired fragment separation was achieved. Gels were visualised using a UV transilluminator

(Gel logic 200) and the size of the fragments was referenced to one of the following molecular weight ladders: Quick-Load DNA marker, broad range (NEB); GeneRuler DNA Ladder, 1 kb (Thermo Scientific); GeneRuler DNA Ladder, 100 bp (Thermo Scientific).

Buffer	Reagent	Amount required
50× TAE running buffer (per litre)	Tris-HCl	242.0 g
	Acetic acid (glacial)	57.0 mL
	EDTA (0.5 M)	100 mL
6× DNA loading buffer (per 10 mL)	ddH ₂ O	6.7 mL
	Bromophenol blue	25 mg
	Glycerol	3.3 mL
	Xylene cyanol FF	25 mg
DNA electrophoresis gel (per gel)	Agarose	0.5 g
	TAE buffer (1×)	50 mL
	SYBR® Safe DNA gel stain	5 µL

Digested DNA was purified using a QIAQuick Gel extraction Kit (QIAGEN) according to the manufacturer's instructions. DNA concentration was estimated by measuring the absorbance at 260 nm using a Nanodrop® ND-1000 spectrophotometer (Nanodrop® Technologies Inc.) with the 'DNA-50' program module according to the manufacturer's instructions.

DNA sequencing

Plasmid sequencing was performed by Source BioScience Sequencing facility in the Department of Biochemistry (University of Oxford). T7 promoter and terminator sequence primers were used unless otherwise stated, on a 3730xl DNA Analyser-Titania.

Site-directed mutagenesis

Oligonucleotide primers were supplied by Sigma® Life Science as a dried pellet, resuspended in 10 mM Tris-HCl (pH 7.5) to a final primer concentration of 1 mM, and stored at -20°C . These were designed with a minimal base-pair mismatch and maximal GC content, with a staggered overlap to prevent primer-dimer formation. All oligo T_m are calculated using the nearest neighbor method with values of 50 mM for cation concentration and 0.5 micromolar for the strand concentration.

Protein	Mutation	Direction	Sequence : (5' to 3')	Length	T_m ($^{\circ}\text{C}$)	GC%
TRIM24	Y872D	Forward	CTTTCTGCCGAGACTTATCTA AACCAGAAGTTG	33	71.5	42.4
TRIM24	Y872D	Reverse	GGTTTAGATAAGTCTCGGCAG AAAGTGCAAATCCACTC	38	76.8	44.7
TRIM24	W828A	Forward	TCCATGAATGAGGACGCGTGT GCAGTTTGTC	31	80.9	51.6
TRIM24	W828A	Reverse	TGACAAACTGCACACGCGTCC TCATTCATGG	31	81.6	51.6
TRIM24	N980A	Forward	CTGTGCTGAATTTCGCTGAGCC TGATTCAGAAG	32	78.1	50
TRIM24	N980A	Reverse	CTGAATCAGGCTCAGCGAATT CAGCACAGTTTT	33	77.7	45.4
TRIM24	N980F	Forward	CAAACTGTGCTGAATTCTTT GAGCCTGATTC	32	73.9	40.6
TRIM24	N980F	Reverse	ACTTCTGAATCAGGCTCAAAG AATTCAGACA	32	74.3	40.6
TRIM33	W889A	Forward	CTTCCAATCCATGGACGCGTG TGCTGTCTG	30	81.5	56.6
TRIM33	W889A	Reverse	TTGGCAGACAGCACACGCGTC CATGGATTG	30	84.0	56.6
TRIM33	N1034A	Forward	GTCCGTTTGATCTTCAAGGCC TGTGAAAGGTTTAATG	37	77.9	43.2
TRIM33	N1034A	Reverse	TCATTAAACCTTTCACAGGCC TTGAAGATCAAACGGA	37	78.6	40.5
TRIM33	N1039A	Forward	TCAAGAACTGTGAAAGGTTTG CTGAAATGATGAAAGTTGTC	42	78.1	35.7
TRIM33	N1039A	Reverse	GAACAACTTTCATCATTTTCAG CAAACCTTTCACAGTTCTTGA	42	78.1	35.7
TRIM33	N1039F	Forward	GAACTGTGAAAGGTTTTTTGA AATGATGAAAGTTGTTT	38	73.6	31.5
TRIM33	N1039F	Reverse	CAACTTTCATCATTTCAAAAA ACCTTTCACAGTTCTTG	38	74.1	31.5

Site-directed mutagenesis of plasmids was performed following the protocol of the QuikChange™ site-directed mutagenesis kit (Stratagene) with the following quantities of reagents being used:

Reagent	Concentration	Amount used (μL)
DNA template	~100 ng/μL	0.5
5' forward primer	10 pmol/μL	1.25
3' reverse primer	10 pmol/μL	1.25
dNTP mix	10 mM of each	1.0
10 × PCR buffer	n. a.	5.0
Sterile Milli-Q water	n. a.	40.0

A Techne Genius thermal cycler was programmed for an initial hold at 95 °C for 2 min to denature double stranded DNA, after which 1.0 μL of PfuTurbo® DNA polymerase was added. This was followed by 18 cycles of denaturing at 95 °C for 30 sec, annealing at 54 °C for 1 min, extension at 68 °C for 8 min. For inactivating mutations, 1.0 μL of PfuTurbo® DNA polymerase was added to the reaction mixture before incubation in a Labnet Multigene II Personal Thermal Cycler, with an initial denaturation step at 95 °C for 30 sec, followed by 16 cycles of denaturing at 95 °C for 30 sec, annealing at 55 °C for 1 min, and extension at 68 °C for 7 min. The resulting samples were held at 4 °C before removal from the cycler, and digestion of parental DNA (centrifugation with 1 μL of restriction endonuclease *DpnI*, followed by incubation at 37 °C for 1 h). Immediate transformation into competent *E. coli* XL10-Gold cells, yielded colonies containing plasmids with the desired sequence.

Cloning

Gene inserts were generated by PCR from parental DNA templates using the following primers:

Protein	Primer	Direction	Sequence : (5' to 3')	Length	T _m (°C)	GC%
TRIM24	LIC	Forward	TACTTCCAATCCATGAATG AGGACTGGTGTG	31	74.6	45.1
TRIM24	LIC	Reverse	TATCCACCTTTACTGTCAT TCTGGATAGAGG	31	68.9	41.9

Protein	Primer	Direction	Sequence : (5' to 3')	Length	T _m (°C)	GC%
TRIM33	LIC	Forward	TACTTCCAATCCATGGACT GGTGTGCTGTCTGC	33	79.4	51.5
TRIM33	LIC	Reverse	TATCCACCTTTACTGTTAC TTTATATGTACTGGTCTCT C	39	68.1	35.8
TRIM24	SLIC	Forward	CAGACCGGTGGATCCATG AATGAGGACTGGTGTG	34	82.9	55.8
TRIM24	SLIC	Reverse	TGCGGCCGCAAGCTTTTAC TGTCATTCTGGATAGAGG	37	82.3	51.3
TRIM33	SLIC	Forward	CAGACCGGTGGATCCATG GACTGGTGTGCTGTCTGC	36	86.9	61.1
TRIM33	SLIC	Reverse	TGCGGCCGCAAGCTTTTAC TTTATATGTACTGGTCTCT C	39	78.5	46.1
TRIM24	extended SLIC	Forward	CCATTGACGTGTTCCAGCA ACAGACCGGTATGAATGA GGACTGGTGTGCAGTTTGT C	57	90.3	50.8
TRIM24	extended SLIC	Reverse	GAAGATTCTTGGAGATAG GTCTTACTGTCATTTTCGA ACGCCGGCGTGAGCTCGTG GTG	60	89.1	50
TRIM33	extended SLIC	Forward	CCATTGACGTGTTCCAGCA ACAGACCGGTATGGACTG GTGTGCTGTCTGCC	51	91.4	56.8
TRIM33	extended SLIC	Reverse	GTCTACTCTCTGGTCATGT ATATTTTATTGTTTCAACG CCGGCGTGAGCTCGTGGTG	57	88.7	50.8
TRIM24	BamHI-gene	Forward	TGGATCCATGAATGAGGA CTGGTGTGCAGTT	31	78.4	48.3
TRIM24	gene-HindIII	Reverse	CGCAAGCTTTTACTGTCAT TCTGGATAGAGGTTCTTTA GAAGTTC	45	76.7	40
TRIM33	BamHI-gene	Forward	TGGATCCATGGACTGGTGT GCTGTCTGC	28	79.5	57.1
TRIM33	gene-HindIII	Reverse	CGCAAGCTTTTACTGTTAC TTTATATGTACTGGTCTCT CATCTGA	45	75.0	37.7
TRIM33	gene-NotI	Reverse	GACTGCGGGCGCTTACTGT TACTTTATATGTACTGGTC TCTCATCTGA	48	80.4	45.8
SUMO	NdeI-gene	Reverse	AGGAGATATCATATGAAA CATCACCATCACCATCAC	36	73.9	38.8
SUMO	gene-NcoI	Reverse	CTTGCCCATGGATCCACCG GTCT	23	75.6	60.8

Protein	Primer	Direction	Sequence : (5' to 3')	Length	T _m (°C)	GC%
TRIM24	Gibson	Forward	AGCAACAGACCGGTGGAT CCATGAATGAGGACTGGT G	37	84.3	54
TRIM24	Gibson	Reverse	TCGAGTGC GGCCGCAAGCT TACTGTCATTCTGGATAGA G	39	83.9	53.8
TRIM33	Gibson	Forward	AGCAACAGACCGGTGGAT CCATGGACTGGTGTGCTGT CTG	40	87.6	57.5
TRIM33	Gibson	Reverse	GTGCGGCCGCAAGCTTTTA CTTTATATGTACTGGTCTC TCATCTG	45	81.0	46.6

Amplification was carried out in a Labnet Multigene II Personal Thermal Cycler, with an initial denaturation step at 95 °C for 2 min, after which 1.0 µL of Q5® High-Fidelity DNA Polymerase (NEB) was added, followed by 25 cycles of denaturing at 95 °C for 30 sec, annealing at 55 °C for 1 min, and extension at 68 °C for 8 min. The resulting samples were held at 4 °C before removal from the cycler, and digestion of parental DNA (centrifugation with 1 µL of restriction endonuclease *DpnI*, followed by incubation at 37 °C for 1 h).

Reagent	Concentration	Amount used (µL)
DNA template	~100 ng/µL	0.5
5' forward primer	10 pmol/µL	1.25
3' reverse primer	10 pmol/µL	1.25
dNTP mix	10 mM of each	1.0
5 × Q5 buffer	n. a.	4.0
Sterile Milli-Q water	n. a.	11.0

Vector linearisation

Approximately 5 µg of vector was incubated at 37 °C for 4 h, with 1 µL of each restriction enzyme in a 100 µL reaction (NEBuffer 4: cutsmart). pGTvL1-SGC and pNIC28-Bsa4 vectors were linearised for LIC of TRIM24 or TRIM33 by treatment with BseRI. pNIC28-Bsa4 vectors were linearised for insertion of *SUMO* gene by treatment with NdeI-HF and NcoI. pETM11-SUMO3GFP

vectors were linearised for insertion of TRIM24 or TRIM33 by treatment with BamHI-HF and HindIII-HF. All vectors were analysed and purified by gel electrophoresis immediately after digest.

Ligation

T4 DNA ligase (NEB) was used in a 20 μ L reaction (T4 DNA ligase buffer) according to the manufacturer's protocol with a 1:3 molar ratio of vector to insert.

Cohesive end generation

Cohesive ends were generated with T4 DNA polymerase in presence of dGTP: 18 μ L DNA (0.5-1.0 μ g), 1 μ L T4 DNA polymerase (0.5 U), 2 μ L 10 $\times$ NEBuffer 2, 0.2 μ L BSA (10 mg/mL), incubated at 22 $^{\circ}$ C for 30 min before addition of 2 μ L dGTP (10 mM).

(Sequence) ligation-independent cloning

After cohesive end generation, gene inserts were annealed with linearised vectors in a 1:1 or 3:1 ratio at 37 $^{\circ}$ C for 30 min, followed directly by transformation onto plates containing appropriate antibiotics and 5% sucrose for negative selection of SacB gene.

Gibson assembly

Cloning was performed using a one-step isothermal *in vitro* recombination reaction consisting of T5 exonuclease, Phusion DNA polymerase and Taq DNA ligase.¹ Linear vectors ($\sim$ 0.03 pmol) and gene inserts ($\sim$ 0.1 pmol) with complementary regions were incubated at 50 $^{\circ}$ C for 15-60 min in a 1:3 ratio with Gibson Assembly master mix (NEB): 20 μ L reaction volume, 10 μ L 2 $\times$ GA master mix. 2 μ L of the reaction mixture was used directly to transform NEB 5-alpha (#C2987) competent cells, or diluted 1:3 to transform BL21 (DE3) Gold competent cells.

Competent cells

Escherichia coli (*E. coli*) strains of the following genotypes were used:

BL21 (DE3) Gold: F⁻ ompT hsdS_B (r_B-m_B⁻) gal dcm (DE3). DE3 denotes a chromosomal copy of the T7 RNA polymerase gene.

XL10 (Gold): Tet^r D(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac The [F' proAB lacIqZDM15 Tn10 (Tet^r) Amy Cam^r].

NEB 10-beta: araD139 D(ara-leu)7697 fhuA lacX74 galK(D(lacZ)M15) mcrA galU recA1 nupG rpsI (Str^r) D(mrr-hsdRMS-mcrBC).

NEB 5-alpha (#C2987): fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17

Competent cells were thawed on ice for 20 min before 1-5 μL of plasmid DNA was added to 20 μL of the competent cells in pre-chilled 50 mL Falcon tubes. The tubes were left on ice for 30 min before placement into a water bath at 42 °C ('heat shocked') for 20 sec for BL21 DE3 (Gold) cells or 30 sec for XL10 (Gold) and NEB cells. The tubes were returned to ice for 5 min, 300 μL 2-TY media was added gently to each tube which was then incubated at 37 °C for 1 h before 100 μL of the transformation mixture was plated onto agar plates containing the appropriate antibiotic and incubated overnight.

7.1.4 Constructs

Vector maps are available in Appendix 8.2. DNA sequences for all plasmids are available in Appendix 8.3, with chromatogram files (.ab1) available in the CD Appendix: Plasmid_sequencing_data.

TRIM24

The pNIC28-Bsa4 vector containing His₆-TRIM24 D872Y (PHD-Bromo) was reverted to a WT construct by site-directed mutagenesis with Y872D primers. LIC cloning into a pGTvL1-SGC vector afforded Gst-TRIM24 (PHD-Bromo). Gibson assembly into the pETM11-SUMO3GFP vector, with excision of the *GFP* gene, afforded His₆Sumo-TRIM24 (PHD-Bromo).

TRIM33

pGEX-4T1 TIF1 γ plasmid-infused filter paper was rehydrated with 50 μ L 10 mM Tris buffer (pH 7.5) for 5 min. Following a brief centrifugation, the supernatant contained a DNA concentration of 97.4 ng/ μ L. Successive transformation into NEB 10-beta, XL10 (Gold), and BL21(DE3) Gold cells afforded cultures for growth and expression. LIC cloning into a pGTvL1-SGC vector afforded Gst-TRIM33 (PHD-Bromo). Gibson assembly into the pETM11-SUMO3GFP vector, with excision of the *GFP* gene, afforded His₆Sumo-TRIM33 (PHD-Bromo).

7.1.5 Protein expression and purification

All constructs were transformed into *E. coli* BL21 (DE3) Gold cells for expression.

Bacterial growths

Media	Reagent	Amount required per litre (g)
Terrific Broth	Tryptone (pancreatic digest of casein)	12.0
	Yeast extract	24.0
	K ₂ HPO ₄	9.4
	KH ₂ PO ₄	2.2
2×Tryptone-Yeast (2-TY) Extract	Bacto tryptone	16.0
	Yeast extract	10.0
	NaCl	5.0

All growth media were autoclaved (*Thermo Live Sciences* MAT 490 LEI) at 121 °C for 45 min before use. Optical density readings were taken in 1.6 mL cuvettes with 5 × dilution in Milli-Q water against a reference sample of the growth media at zero time. The absorbance was measured at 600 nm using a Novaspec® II spectrophotometer.

Bacterial plate cultures were prepared using solid 2-TY medium with the appropriate antibiotic, and streaked in a laminar flow hood (HereSafe® Model KS12 Class II Biosafety Cabinet) using sterile equipment. Following inoculation, these were inverted and grown overnight at 37 °C in a Heraeus® TypB 6030 incubator. Bacterial plates with resulting bacterial colonies were stored in

the dark at 4 °C. Small-scale growths for starter cultures (100 mL of 2-TY media containing the appropriate antibiotic in 500 mL Duran flasks), were inoculated with a single colony from an agar plate, and grown at 37 °C overnight.

Expression trials

Starter cultures were used to inoculate 100 mL (in 500 mL Duran flasks) of 2-TY media containing the appropriate antibiotic. Cultures were incubated at 37 °C until they reached an OD₆₀₀ of 1.0. Cultures were transferred to incubators held at 18 °C or 28 °C and allowed to equilibrate for 15 min before induction with ZnCl₂ to a final concentration of 0.1 mM and IPTG to final concentrations of 0.1 mM, 0.25 mM or 0.5 mM. Cultures were incubated at these temperatures overnight, before being pelleted by centrifugation, resuspended in 300 µL HisTrap column binding buffer for whole-cell analysis by SDS-PAGE.

Large-scale growth

Large-scale recombinant protein expression was performed using 2000 mL PYREX® narrow-mouth graduated Erlenmeyer flasks containing 600 mL of 2-TY medium containing the appropriate antibiotic. Flasks were inoculated with 1 % of the bacterial starter culture and grown at 37 °C until they reached an OD₆₀₀ of 1.0. Cultures were equilibrated to 18 °C and induced with ZnCl₂ (0.1 mM) and IPTG (0.25 mM), before incubation overnight. After this time, cultures were centrifuged and the resulting bacterial pellet frozen in a sealable plastic bag at –80 °C.

Protein purification

Cell pellets were dissolved in binding buffer (5× mass of cell pellet w/v) and lysed by sonication, applying thirty 10 sec bursts interrupted by 10 sec pauses. The lysate was clarified by centrifugation and the supernatant decanted, filtered through a 0.45 µm filter, and applied to a purification column.

Reagents	IMAC buffers (pH 7.6)				GST affinity column buffers (pH 7.8)	
	Binding	Wash	Elution	Strip	Binding/wash	Elution
HEPES	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM
NaCl	0.5 M	0.5 M	0.5 M	0.5 M	150 mM	150 mM
Imidazole	5 mM	30 mM	500 mM	–	–	–
EDTA	–	–	–	100 mM	–	–
DTT	–	–	–	–	0.5 mM	0.5 mM
Glutathione	–	–	–	–		20 mM

Immobilised metal affinity chromatography (IMAC) was performed with a HisTrap™ (6 mL) column (GE Healthcare) charged with 20 mL of 100 mM NiSO₄ and washed with 50 mL of binding buffer prior to loading the cell lysate at 1 mL/min. The column was then washed with approximately 100 mL binding buffer. When all residual products of bacterial fermentation were eluted from the column by binding buffer, the protein of interest was batch-eluted with 30 mL elution buffer. Eluted protein was collected and fractionated; fractions containing the highest levels of pure protein (as determined by SDS-PAGE gels) were concentrated using Amicon Ultracell-30k concentrators (Millipore) in a Beckmann Allegra™ 21R centrifuge.

GST affinity column: 10 mL was equilibrated with 30 mL of binding buffer prior to loading the cell lysate at 1 mL/min. The column was then washed with approximately 100 mL binding buffer. When all residual products of bacterial fermentation were eluted from the column by binding buffer, the protein of interest was batch eluted with 60 mL elution buffer. Eluted protein was collected and fractionated; fractions containing the highest levels of pure protein (as determined by SDS-PAGE gels) were concentrated using Amicon Ultracell-30k concentrators (Millipore) in a Beckmann Allegra™ 21R centrifuge.

Further purification was achieved by gel filtration chromatography. Concentrated fractions (<2 mL) were loaded on a 150/300 mL Superdex 75 size exclusion chromatography column (Amersham). Eluted protein was collected and fractionated; fractions containing the highest levels of pure protein

(as determined by SDS-PAGE gels) were concentrated (to 50-500 μ M) using Amicon Ultracell-30k concentrators (Millipore) in a Beckmann Allegra™ 21R centrifuge.

Protein concentration was estimated by measuring the absorbance at 280 nm using a Nanodrop® ND-1000 spectrophotometer (Nanodrop® Technologies Inc.) with the ‘Protein A280’ program module according to the manufacturer’s instructions. Protein molecular weight and molar extinction coefficient were calculated using ProtParam on the ExPASy Bioinformatics Resource Portal.

Purification by strong anion exchange

Attempts were made to purify TRIM33 using strong anion exchange. After affinity purification, protein was exchanged into buffer A with a PD-10 buffer exchange column according to the manufacturer’s instructions. The sample (10 mL) was loaded onto a CIM® QA-1 tube monolithic column (BIA Separations) with a bed volume of 1 mL, which had been pre-equilibrated in buffer A, and eluted at 1 mL/min, with collection into 1 mL fractions.

Elution gradients used were as follows, concluding with an isocratic elution of any strongly bound anions using 100% B. Gradient A: 0% B, 5 mL; 0-50 % B, 20 mL; 100 % B, 10 mL. Gradient B: 0% B, 5 mL; 0-10% B, 10 mL; 10-50% B, 20 mL; 100% B, 10 mL. Gradient C: 0-18% B, 10 mL; 18-22% B, 10 mL; 22-32% B, 10 mL; 32-60% B, 10 mL; 100% B, 10 mL. Gradient D: 0-20% B, 10 mL; 20-30% B, 20 mL; 30-60% B, 10 mL; 100% B, 10 mL.

Reagent	QA column buffers (pH 7.6)		
	Concentrated buffer	Buffer A (low salt)	Buffer B (high salt)
HEPES	500 mM	20 mM	20 mM
NaCl	1 M	50 mM	1 M
β -ME		1 mM	1 mM

Whole-protein mass spectrometry

Intact masses of proteins, purified as described above, were recorded by liquid chromatography/mass spectrometry (LC/MS) using a ZMD (Waters®) single quadrupole mass spectrometer, interfaced with a Hewlett Packard® Series 1050 liquid chromatography and sample

handling system. The protein sample (7 μ L) was injected onto a Grace Vydac 214TP C4 (250 mm $\times$ 4.6 mm) reverse-phase HPLC column and eluted at 1 mL/min using a gradient system from Solvent A (0.1 % formic acid in H₂O) to Solvent B (0.1% formic acid in acetonitrile). Samples were injected directly into the mass spectrometer. The following MS parameters were used: polarity: ES⁺; capillary voltage: 3,000 V; sample cone voltage: 35 V; extraction cone voltage: 1V; desolvation temperature: 250 °C; source temperature: 100 °C; cone gas flow: 100 L/h; desolvation gas flow (N₂): 830 L/h. Sodium formate was used to calibrate the instrument. Spectra were processed using MassLynx™ v4.0 (Waters Corporation) with the Maximum Entropy method (MaxEnt1) employed for deconvolution of intact mass (1 Da resolution). Expected masses were calculated using ExPASy ProtParam tool on the Bioinformatics Resource Portal.

SDS-PAGE

The purity of proteins was analysed by PAGE. Gels were run on a mini-PROTEAN Tetra Electrophoresis System (Bio-Rad) at a constant potential of 200 V. Following electrophoresis, gels were stained for 20 min with InstantBlue™ Coomassie® stain (Expedeon). For denaturing PAGE analyses, samples were prepared by mixing with sample loading buffer (2 $\times$) and incubation at 100 °C for 3 min. All gels were loaded with one of the following molecular weight ladders: SeeBlue® prestained standard (Thermo Fisher Scientific); BLUEstain™ protein ladder, 11-245 kDa (Gold Biotechnology); PageRuler Plus prestained protein Ladder (Thermo Fisher Scientific); blue protein standard, broad range (NEB); colour protein standard, broad range (NEB). Gel compositions (quantities for 1 gel):

Reagent	Resolving gel		Stacking gel
	12.5 %	15 %	4 %
Tris-HCl (1.5 M, pH 8.8)	1.25 mL	1.25 mL	–
Tris-HCl (0.5 M, pH 6.8)	–	–	0.63 mL
Milli-Q water	1.57 mL	1.15 mL	1.57 mL
30 % (w/v) Acrylamide	2.08 mL	2.50 mL	0.25 mL
SDS (10 % w/v)	50 μ L	50 μ L	25 μ L
10 % (w/v) APS	50 μ L	50 μ L	25 μ L
TEMED	4 μ L	4 μ L	2.5 μ L

Gels were prepared using 70 mm × 100 mm glass plates, with 0.75 mm spacers. TEMED and freshly prepared ammonium persulfate were added just prior to pouring the gels. The resolving gel was cast with the addition of a separate layer of methanol to ensure a level surface. Once the resolving layer had polymerized, methanol was removed and the stacking gel was cast.

Buffer	Reagent	Composition	Amount required per litre
SDS-PAGE running buffer	Tris-HCl		30 g
	Glycine		144 g
	SDS		10 g
SDS-PAGE sample loading buffer (pH 6.8)	Tris-HCl	0.5 M	10 mL
	Bromophenol blue	0.2 % w/v	0.2 g
	SDS	10 % w/v	20 mL
	Glycerol		12 mL
	β-mercaptoethanol		5.0 mL
SDS-PAGE stain	Coomassie® brilliant blue R	0.25 %	2.5 mL
	acetic acid	10 %	100 mL
	methanol	30 %	300 mL

SDS-PAGE destain was of the same composition without the presence of Coomassie® brilliant blue R.

Affinity tag proteolytic cleavage

Buffer	Reagent	Composition
TEV cleavage buffer (pH 7.2 at 25 °C)	Tris HCl	25 mM
	ZnCl ₂	10 μM
	NaCl	500 mM
	Trisodium citrate dihydrate	5 mM
	2-Mercaptoethanol	5 mM
SENP1 cleavage buffer (pH 7.8 at 25 °C)	HEPES	50 mM
	NaCl	0.5 M
	Imidazole	5 mM

Protein samples were buffer exchanged using a PD-10 Desalting Column containing 8.3 mL of Sephadex™ G-25 Medium by elution with 3.5 mL of cleavage buffer. Protein constructs with a TEV cleavage site were incubated at 4 °C with 20 μL of TEV protease (9 mg/mL). Samples were

taken for analysis by mass spectroscopy and SDS-PAGE gels over a period of 2 days and afforded only partial cleavage. Protein constructs with a SUMO expression label were incubated at 4 °C with 1% w/w SENP1 for 2 h: complete cleavage was observed by SDS-PAGE. The cleaved His₆-SUMO and His₆-SENP1 were removed by IMAC chromatography, isolating cleaved protein in the flow-through before further purification by size exclusion chromatography.

Circular dichroism

For investigation of protein secondary structure using circular dichroism (CD), measurements were acquired using a Chirascan CD spectrometer (Applied Photophysics) with a Peltier temperature-controlled cell holder. All experiments were performed in a 0.1 cm path length cuvette using 0.1-0.2 mg/mL protein in 10 mM sodium phosphate buffer (pH 8.0). Data were recorded from 240 to 185 nm, at 0.5 nm intervals, and each data point was averaged for 3 sec. Spectra were base-line corrected and smoothed using the Savitzky-Golay filter.

7.1.6 Assay procedures

Buffer	Reagent	Composition	Amount required per litre
AlphaScreen buffer (pH 7.6)	HEPES	25 mM	5.95 g
	NaCl	100 mM	5.84 g
	CHAPS	0.05 % w/v	0.5 g
	BSA	0.1 w/v	1 g
Fluorescence buffer (pH 7.4)	dry HEPES	25 mM	5.95 g
	NaCl	100 mM	5.84 g
	dry CHAPS	0.05 % w/v	0.5 g
HBS-EP buffer (pH 7.4)	HEPES	10 mM	2.38 g
	NaCl	150 mM	8.76 g
	Surfactant P20	0.05 % v/v	0.5 mL
	EDTA	3 mM	877 mg
HBS-P+ buffer (pH 7.4)	HEPES	10 mM	2.38 g
	NaCl	150 mM	8.76 g
	Surfactant P20	0.05 % v/v	0.5 mL

7.1.6.1 Tryptophan fluorescence

Fluorescence was measured on a Perkin-Elmer LS50B luminescence spectrometer in a 400 μ L micro absorption quartz cuvette. Emission was measured over 300-400 nm after excitation at either 280 nm, to excite tryptophan and tyrosine residues, or 295 nm, to selectively excite tryptophan residues. Data were acquired and processed using Carey WinUV Scan. For concentration dependence a 32 μ M solution of His₆-TRIM24 was used for serial two-fold dilutions in fluorescence buffer. For peptide titrations 1-4 μ L increments of 100 μ M peptide, in fluorescence buffer, were added to a 240 μ L sample of 2 μ M protein.

7.1.6.2 AlphaScreen

General procedure

Assay buffer was freshly prepared each day by supplementing HEPES base with BSA, filter sterilized through a 0.22 μ m filter, and stored at 4 °C until use. For incubation steps, the plate was sealed, shaken for 10 sec at 600 rpm on a plate oscillator, and incubated at room temperature in the dark. Compounds (5 μ L) were dispensed into a ProxiPlate-384 Plus (Perkin Elmer) in duplicate or triplicate, as point concentrations or serial buffer dilutions from 50 mM DMSO solutions. Protein-peptide mix (7 μ L) was dispensed into each well using a Thermo-Fisher electronic multi-channel pipette, and incubated for 30 min. AlphaScreen bead mix (8 μ L; 25 μ g/mL of acceptor and donor beads) was added to each well, taking adequate precautions to avoid light exposure, before being incubated for 1 h. The plate was read using a Synergy™ 2 MultiMode Microplate Reader using the built-in AlphaScreen 384 ProxiPlate function: excitation 680 nm, 0.18 sec; emission 570 nm, 0.37 sec.

Single-shot screening against BAZ2A and BRD9A was carried out by the Structural Genomics Consortium (Oxford) as described previously.^{2,3} Single shot screening against BRD4 was performed in triplicate. Compounds were tested at 25 μ M and 250 μ M with <0.5% DMSO. Amino acids were tested at 50 μ M and 250 μ M with 0.5% DMSO. Inhibition is reported as a reduction in signal arising from peptide-bromodomain interaction, normalised to a DMSO control. Signal

response curves were obtained for peptides or compounds in triplicate, with serial 1:2 dilutions. Samples containing buffer or DMSO in place of compound were included in all assays for normalisation. Data were processed by fitting a four-parameter equation to calculate IC₅₀ values using GraFit version 7.0.3.⁴

Assay composition

Ni²⁺ chelate acceptor and streptavidin donor beads were prepared as a mixture in a 1:300 or 1:400 dilution as required (0.005-0.007 mg/mL FAC); biotinylated peptide employed, and final assay concentrations were as detailed in table below. Assay well volume was 20 µL.

Protein (His ₆ -)	FAC protein (nM)	Peptide	Sequence	FAC peptide (nM)	FAC beads (µg/mL)
TRIM24	50	B1	H3 ₁₋₂₇ K9Me3	15	5
	10	B2	H3 ₁₋₂₇ K18Ac	8	5
	10	B3	H3 ₁₋₂₇ K23Ac	8	5
	10	B4	H3 ₁₋₂₇ K9Me ₃ K18Ac	4	5
	10	B5	H3 ₁₋₂₇ K9Me ₃ K23Ac	8	5
	50	B6	H3 ₅₋₂₄ K14Ac	15	5
	30	B7	H3 ₁₋₂₇ K14Ac	8	5
TRIM33	500	B1	H3 ₁₋₂₇ K9Me3	15	7
	250	B2	H3 ₁₋₂₇ K18Ac	30	7
	50	B4	H3 ₁₋₂₇ K9Me ₃ K18Ac	15	7
	500	B5	H3 ₁₋₂₇ K9Me ₃ K23Ac	15	7
BRD4	10	H4(KAc)₄	S-H4 ₁₋₂₀ K5AcK8AcK12AcK16Ac	4	5
BRD9	40	H2A(KAc)₃	H2A ₁₋₂₁ K9AcK13AcK15Ac	25	5
BAZ2A	25	B6	H3 ₅₋₂₄ K14Ac	50	5

Peptides

Peptides were supplied as 1 mg quantities with >95% purity (as determined by HPLC; Cambridge Research Biochemicals). Residual water and counter ions were predicted to comprise 20% of the net mass, and peptides were dissolved to 10, 1 and 0.1 mM stocks in buffer (50 mM HEPES, 500 mM NaCl; pH 7.6) before being stored at -80 °C in 5 µL aliquots.

Affinities of the following peptides with varying modifications were measured:

Peptide	Sequence	Amino acid sequence
U1	H3 ₁₋₂₇ K14Ac	ARTKQTARKSTGGK(Ac)APRKQLATKAARK-acid
U2	H3 ₁₋₂₇ K18Ac	ARTKQTARKSTGGKAPRK(Ac)QLATKAARK-acid
U3	H3 ₁₋₂₇ K23Ac	ARTKQTARKSTGGKAPRKQLATK(Ac)AARK-acid
U4	H3 ₁₋₂₇ K9Me3	ARTKQTARK(Me ₃)STGGKAPRKQLATKAARK-acid
U5	H3 ₁₋₂₇ K9Me3K18Ac	ARTKQTARK(Me ₃)STGGKAPRK(Ac)QLATKAARK-acid
U6	H3 ₁₋₂₇ K9Me3K23Ac	ARTKQTARK(Me ₃)STGGKAPRKQLATK(Ac)AARK-acid
U7	H3 ₁₋₂₇ K18AcK23Ac	ARTKQTARKSTGGKAPRK(Ac)QLATK(Ac)AARK-acid
U8	H3 ₁₋₂₇ K9Me3K18Ac K23Ac	ARTKQTARK(Me ₃)STGGKAPRK(Ac)QLATK(Ac)AARK-acid
U9	H3 ₁₋₂₇	ARTKQTARKSTGGKAPRKQLATKAARK-acid

Compound screening

Compounds were provided in, or transferred to, LabCyte 384-LDV source plates, which were compatible with an Echo acoustic liquid handler (Labcyte), and stored in controlled environments under nitrogen with <0.5% O₂ and <0.5% humidity when not in use. All transfers to assay plates were backfilled with DMSO to the same final volume (100 nL). Assays were performed according to the general procedure.

201 compounds from the Maybridge collection were provided as solubilised stock (30 mM in DMSO) on LabCyte 384-LDV source plates. Samples of 20 or 100 nL were transferred to assay plates in duplicate to test at 30 or 150 μ M FAC. 16 compounds were subsequently provided as solid samples of 1-2 mg, and solubilised in DMSO (50 mM; c07 – 40 mM; c14 – 30 mM). Titrations were manually prepared using a Matrix electronic multichannel pipette (Thermo Scientific) as dilutions in fresh DMSO from stock concentrations, and transferred to LabCyte 384-LDV source plates. 1536 compounds from the PPI-net compound collection were provided (10 mM in DMSO) in 96-well V-bottomed microplates (Greiner). These compounds were first transferred to 20 $\times$ LabCyte 384-LDV plates for use with the Echo acoustic liquid handler (Labcyte) using a Bravo

Automated Liquid-Handling Platform (Agilent). Samples of 20 or 100 nL were transferred to assay plates in duplicate to test at 10 or 50 μ M FAC.

7.1.6.3 Differential scanning fluorimetry (DSF)

DSF was used to determine T_m values for proteins under different buffer conditions, or in the presence of compounds. Experiments were performed using a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad).

Buffers used were: MES pH 5.5; Bis-Tris pH 6.0, 6.5, 7.0; HEPES pH 7.0, 7.5, 8.0; EPPS pH 7.5, 8.0, 8.5; TAPS pH 8.5, 9.0; CHES pH 9.0, 9.5; CAPSO pH 9.5, 10. 46 μ L of buffer was added to each well in triplicate. 2 μ L of Sypro Orange dye (50 $\times$), and 2 μ L of the protein of interest (50 μ M; 50 mM HEPES, pH 7.5) were added consecutively to the same wells and mixed gently by pipetting.

For compound stabilisation screening, 5 μ L aliquots of predissolved compound (maximum 10% DMSO) were dispensed in each well of a 48-well plate, in triplicate. To prepare a Master Mix, 3.3 μ L of Sypro Orange solution was added to 2.98 mL buffer (500 mM HEPES, 0.5 M NaCl, 1 mM β -ME, pH 7.6) and mixed by inversion, before the protein of interest was added to a concentration of 2.2 μ M. 45 μ L of this Master Mix was added to each well and mixed gently by pipetting: final assay concentrations were 5 $\times$ Sypro Orange, 10/20 μ M compounds, 2 μ M protein.

Plates were sealed with a clear polypropylene seal (StarLab) and centrifuged for 2 min at 2,000 rpm in a Beckman Coulter centrifuge. The chamber equilibrated to 25 $^{\circ}$ C before inserting the sample plate and the readings of fluorescence were taken between 25-85 $^{\circ}$ C, in one degree increments, with a 30 sec hold for each step. FAM (492 nm) and ROX (610 nm) filters were used for excitation and emission, respectively. Software automatically performed global minimum subtraction. The T_m was

$$y = LL + \frac{UL - LL}{1 + e^{-ax}}$$

calculated using the Boltzmann equation $y = LL + \frac{UL - LL}{1 + e^{-ax}}$, where LL and UL are values of minimum and maximum intensities and a corresponds to the slope of the curve within T_m . Boltzmann fitting was performed using OriginPro 9.1, and the T_m shift was calculated relative to reference wells on the same plate.

7.1.6.4 Isothermal calorimetry (ITC)

All calorimetric experiments were performed on a MicroCal iTC200 or MicroCal PEAQ-ITC (Malvern) and analysed with the MicroCal ORIGIN software package using a single binding site model. The first data point was excluded from the analysis. Proteins and peptides containing an N-terminal tyrosine were dialysed at room temperature overnight in a Slide-A-Lyzer™ MINI Dialysis Device (2000 MWCO; Thermo Scientific Life Technologies) into 50 mM HEPES, 500 mM NaCl; pH 7.5. Samples were centrifuged to remove aggregates (20 min, 14,000 rpm, 25 °C), and concentrations estimated by measuring the absorbance at 280 nm using a Nanodrop® ND-1000 spectrophotometer (Nanodrop® Technologies Inc.) with the ‘Protein A280’ program module according to the manufacturer’s instructions (tyrosine $\epsilon_{280} = 1280 \text{ M}^{-1}\text{cm}^{-1}$). Samples were diluted to the required concentration using dialysis buffer, and lyophilized histone peptides were dissolved in the same buffer. The cell was stirred at 750 rpm, with reference power set to 4 $\mu\text{cal/sec}$ and temperature held at 298 K. After an initial delay of 60 sec, 20 $\times$ 2 μL injections (first injection 0.4 μL) were performed with a spacing of 180 sec. Heats of dilution were measured under the same conditions and subtracted for analysis.

Protein solutions in the calorimetric cell (250 μL) were titrated with the peptide solutions in the syringe (60 μL) at the following concentrations: 100 μM BRD4(1) and 1.0 mM H4₁₋₂₀(KAc)₄; 135 μM BRD4(1) and 1.27 mM H4₁₋₂₀(KAc)₃K12C; 20 μM TRIM24 and 200 μM H3₁₋₂₇K18Ac; 20 μM TRIM33 and 200 μM H3₁₋₂₇K18Ac; 50 μM TRIM24 and 750 μM H3₁₋₂₇K9Me₃; 150 μM TRIM33 and 1.67 mM H3₁₋₂₇K9Me₃.

7.1.6.5 Surface plasmon resonance (SPR)

SPR experiments were performed on a Biacore™ T200, under the supervision of Dr David Staunton, in the Molecular Biophysics Suite, Department of Biochemistry, University of Oxford. All buffers were filtered and degassed immediately before use.

Carboxymethylated dextran (CM5) streptavidin immobilisation

A fresh series S CM5 sensor chip (Biacore Life Sciences) was loaded into the system and primed with HBS-P+ buffer, before four cells were coated with streptavidin. The following injection cycle resulted in an increase of 920 RU in baseline signal: NaOH (50 mM in HBS-P+), 5 μ L/min, 90 sec $\times$ 2; 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 0.4 M in H₂O) and *N*-hydroxysuccinimide (NHS, 0.1 M in H₂O), 10 μ L/min, 420 sec; streptavidin (Sigma S-4762, 0.5 mg/mL in 10 mM NaOAc, pH 4.0), 5 μ L/min, 840 sec; ethanolamine HCl (1 M, pH 8.5), 10 μ L/min, 420 sec; glycine HCl (100 mM, pH 2.0), 30 μ L/min, 30 sec $\times$ 3.

CM5 analysis

Biotinylated peptide ligand was immobilised on a single cell (F2) of the streptavidin coated CM5 chip, with non-specific binding removed *via* multiple weak acid washes: Biotin-H3₁₋₂₇K9Me₃ (100 mM), 30 μ L/min, 30 sec $\times$ 2; glycine HCl (100 mM, pH 3.0), 30 μ L/min, 30 sec $\times$ 3. Protein analyte titrations were performed across active and reference cells (F2-F1), from a series of ten 1:2 dilutions from 102.4 μ M His₆-TRIM24, run sequentially with increasing concentrations (30 μ L/min, 120 sec). A recovery time of 600 sec between injections was followed by a weak acid wash to regenerate the surface: glycine HCl (100 mM, pH 3.0), 30 μ L/min, 30 sec.

An alternative biotinylated peptide ligand was immobilised on a single cell (F4) of the streptavidin coated CM5 chip, with non-specific binding removed *via* multiple weak acid washes: Biotin-H3₁₋₂₇K9Me₃K18Ac (100 mM), 30 μ L/min, 30 sec $\times$ 2; glycine HCl (100 mM, pH 2.0), 30 μ L/min, 30 sec $\times$ 3. Protein analyte titrations were performed across active and reference cells (F4-F3), from a series of ten 1:2 dilutions from 102.4 μ M His₆-TRIM24, run sequentially with increasing concentrations (30 μ L/min, 120 sec). A recovery time of 600 sec between injections was followed by a weak acid wash to regenerate the surface: glycine HCl (100 mM, pH 2.0), 30 μ L/min, 30 sec.

Nitrilotriacetic acid (NTA) analysis

A fresh series S NTA sensor chip (Biacore Life Sciences) was loaded into the system and primed with HBS-EP buffer, followed by a short wash: EDTA (3 mM in HBS-EP buffer), 10 μ L/min, 60 sec. For each cycle, a single cell of the chip (F2) was activated *via* nickel chelation, before ligand capture: NiCl₂ (0.5 mM in HBS-EP buffer), 10 μ L/min, 60 sec; His₆-TRIM24 (0.1 μ M), 10 μ L/min, 60 sec. Peptide analyte titrations were performed across active and reference cells (F2-F1), from a series of ten 1:2 dilutions from 2 μ M His₆-H3₁₋₂₇K23Ac, run sequentially with increasing concentrations (30 μ L/min, 120 sec). After each cycle, with a 500 sec recovery time, the surface was regenerated: imidazole (1M in H₂O), 10 μ L/min, 60 sec; EDTA (350 mM), 10 μ L/min, 120 sec.

7.1.6.6 WaterLOGSY

NMR studies on TRIM-ligand interactions were carried out with the supervision of Amjad Khan (University of Oxford). Proteins were buffer exchanged from storage buffer into 20 mM Na/K phosphate buffer (150 mM NaCl, pH 7.5) using PD SpinTrap G-25 columns (GE Healthcare). For each compound, reference ¹H spectra, and WaterLOGSY spectra were obtained in the absence and presence of TRIM24 and TRIM33. Samples were prepared in 160 μ L volumes, from 50 mM DMSO stock solutions, with final concentrations: 500 μ M compound, $\pm$ 20 μ M protein, 10% D₂O for signal locking. After loading into 3 mm Ultra Precision NMR tubes (GPE Scientific), samples were briefly centrifuged in a manual benchtop rotor to collect at base, and identify any precipitates. Spectra were recorded on a Bruker 700 MHz spectrometer equipped with a cryoprobe, with a standard WaterLOGSY pulse sequence.⁵

7.1.7 Crystallisation

All crystallisation trials were carried out at 20 °C. Protein was concentrated to ~20 mg/mL immediately after purification, and plated using a Phoenix Liquid Handling System (Art Robbins Instruments) onto Intelli-Plate 96-3 low-profile crystallisation plates (Hampton Research) in three ratios with precipitation buffer.

Initially, a crystallisation screen was conducted using JCSG+ (Hampton Research) crystallisation condition plates, resulting primarily in precipitation. Manual compilation of a screening selection based around literature conditions,⁶ 100/200 mM HEPES (pH 7.5) with 1.7-2.2 M ammonium sulfate and 0.5-4% PEG400, produced crystals in ~20% of wells. A narrow screening plate of 100 mM HEPES (pH 7.5) with 2.0-2.2 M ammonium sulfate and 1.5-3% PEG400 resulted in the formation of large plates of overlapping crystals within 2-3 days, which reached full size within one week.

Single crystals (precipitant 100 mM HEPES, 2.16 M ammonium sulfate, 2.4% PEG400 pH 7.5) were mounted onto a CrystalCap magnetic sample mount (Hampton Research), with an appropriate diameter nylon loop, cryoprotected by transferring to a solution of 20% glycerol in mother liquor, before being cyro-cooled in liquid N₂.

Data were collected on beamline I03 at the Diamond Light Source (by Dr Michael McDonough, University of Oxford) under the proposal mx9306 (collected March, 2015). The structure of TRIM24 was solved (by David Zollman, University of Oxford) by molecular replacement using an *apo*-TRIM24 structure as the search model (PDB: 3O33). Structural refinement was carried out in PHENIX⁷ with iterative rebuilding of the models using COOT⁸ until $R_{\text{work}}/R_{\text{free}}$ converged to final values of 0.168/0.197. The space group was C2 with six molecules in the symmetrical unit.

7.1.8 Non-denaturing mass spectrometry

Non-denaturing mass spectrometry studies on TRIM24-nucleosome protein-protein complexes were carried out by Nisha Patel (University of Oxford). Nucleosomes were prepared by David Zollman (University of Oxford). Mass spectrometry measurements were performed on a modified Synapt G1 HDMS instrument with the travelling-wave ion mobility cell replaced by an 18-cm drift cell with radial RF ion confinement and a linear voltage gradient to direct ions along the axis of transmission to the time-of-flight (TOF) mass analyser. The mass spectrometer was typically set to a source pressure of 5-7 mBar, capillary voltage of 1.4-1.7 kV, capillary nanoflow of 0.03-0.2 mBar and argon as collision gas with flow rate set to 5-8 mL/min (~6.6-6.7 e⁻² mBar). Collision voltage

ranged from 5 V. The sample and extraction cone and trap bias voltages, quadrupole profile, and collision gas pressure were optimised for maximal ion intensity of the target membrane protein complex. Helium was the drift cell gas and set to a pressure of ~1.6-1.8 Torr (40 mL/min flow rate). Pusher time, pulse width and TDC inhibit were set to 180, 7, and 7 μ sec, respectively.

7.2 Computational methods

Scripts were run from the terminal of a Linux Ubuntu v13.04 guest on a Windows 7 host system running Oracle VM VirtualBox. Sample scripts are available as run in Appendix 8.1, and in the CD Appendix: Computational_scripts. Software was installed as a local stand-alone version unless stated otherwise.

Sequence alignments for comparisons of homology and mutation analyses were performed using ClustalX.⁹ Secondary structure predictions were performed using PSIPRED¹⁰ and GlobProt¹¹ web servers, with an input sequence of TRIM33 α (GenBank Accession: NP_056990).

PHD finger sequence alignments were performed in the Internal Coordinate Mechanics (ICM: Copyright 1989-2012, MolSoft L.L.C.) environment,¹² with a gap opening penalty of 2.4 and a gap extension penalty of 0.15 based on optimal settings for structural similarity identification with the Gonnet comparison matrix.¹³ Alignments were manually adjusted to maintain the characteristic Zn²⁺ chelating residue positions. A phylogenetic tree was created from the resultant alignment using PhyloDraw,¹⁴ labels were manually repositioned for clarity, with branch lengths maintained.

7.2.1 Docking file preparation

Receptor files

The TRIM24 receptor file was prepared from chain A of the *apo*-crystal structure of the protein (PDB: 3O33). Water molecules were assessed for likelihood of displacement using WaterClassifier,¹⁵ with scores appended to PDB files as a new property (in place of the B-factor) for graphical representation in PyMOL. Water molecules located in the base of the KAc pocket

were manually corrected to account for the hydrogen bonding network, before those with a displaceability of >80 % were removed using the built in `extract` function.

The TRIM33 receptor file was prepared from chain A of the *apo*-crystal structure of the protein (PDB: 3U5M). Water molecule positions were not resolved in this crystal structure of TRIM33 due to low resolution, and were modelled using WaterDock2.1.¹⁵ Reliability of this method, particularly for hydrogen bonded networks, was not validated, so comparable water positions were overlaid from the TRIM24 receptor hydration state. Water positions for each domain – PHD finger and bromodomain – were aligned separately in PyMOL using the `cealign` function.

Protonation states were set for pH 7.0 using the AMBER force field in PROPKA's online server (<http://nbc.net/pdb2pqr>), including optimisation of hydrogen bonding and steric restrictions on proximity of atoms. Charges were merged for non-polar hydrogens and lone pairs using the AutoDockTools script `prepare_receptor4.py`.

Ligand files

The Maybridge Fragment Collection, provided by Fisher Scientific, was downloaded as a structure-data file (`.sdf`) of 31,973 molecules. Molecular co-ordinates were replaced in the ICM-Pro environment (MolSoft), and separated into `.pdb` files with the prefix `MaybridgeF`, using the Open Babel Package, version 2.3.1 <http://openbabel.org> (accessed Jan 2013).¹⁶ Charges were merged for non-polar hydrogens and lone pairs using the AutoDockTools script `prepare_ligand4.py`.

Configuration files

The docking areas, parameters, and location of the receptor files, were specified in a `config.txt` file for each simulation. Docking areas of 20 Å² were centred on positions of conserved water molecules: in the base of the KAc pocket (HOH109: TRIM24 x=18.18, y=30.80 z=-3.30; TRIM33 x=8.56, y=31.42 z=-3.13), or in the interdomain region avoiding inclusion of the KAc pocket (HOH144: TRIM24 x=27.19, y=42.64 z=-2.44; TRIM33 x=29.59, y=41.30 z=-1.51).

7.2.2 Docking

Docking was performed using AutoDock Vina¹⁷ on the ARCUS hub of the University of Oxford Advanced Research Computing Facility. The ARCUS (Dell SandyBridge) system comprises of 1344 (84*16) 2.2GHz Xeon SandyBridge cores, 80x64GiB nodes, 4x128GiB 4 nodes RAM, 20TB scratch disk, running on Red Hat Enterprise Linux. Screening files were uploaded to the server with the `pscp` function bundled within the PuTTY network file transfer application. A pre-defined module for AutoDock Vina was loaded into the working environment.

Screening was initiated by the Torque Resource Manager, an open-source, portable batch system (PBS) based job scheduler, on submission of `vina_screen_local_multi.sh`. Simulations generated 20 modes per computation, with affinity and RMSD (relative to lowest energy mode) stored in a log file, and the lowest energy conformation written as a `.pdbqt` file.

7.2.3 Results analysis

Results were downloaded from the ARCUS hub using the script `zip_results.sh` which called `vina_screen_get_top_quant.py` to define the top 1000 results from each screen, and extract both log files and `.pdbqt` molecular results. Binding affinity data were extracted from docking results, and the distributions visualised as histograms. Energetic cut-offs were selected to reduce the number to a sum suitable for manual curation. Reduction criteria and resultant numbers of molecules were: TRIM24 KAc, -7.3 kcal/mol, 381 molecules; TRIM24 cleft, -6.4 kcal/mol, 430 molecules; TRIM33 KAc, -7.5 kcal/mol, 435 molecules; TRIM33 cleft, -6.6 kcal/mol, 294 molecules.

The collection of all identified compounds was combined (1540), reduced to remove duplicates (1312), and manually reduced to a collection of the top 400 putative hits based on predicted affinity, assessed reliability of binding mode, and synthetic tractability. These top 400 results were extracted from the Maybridge fragment screening collection, using the script `make_sdf.sh` which appended the rank and energy from docking simulations. Ligand docking poses for the top 400

results were loaded into PyMOL with a script `view_top400_results_RECEPTOR_LIGAND.pml` generated using the script `make_pml.sh`

Comparison of results to identify a common binding pharmacophore was facilitated by grouping molecules according to similarity. Tanimoto coefficients were calculated in Open Babel for all molecules in the `.sdf`, using the default (path-based) FP2 fingerprint relative to asimplified molecular-input line-entry system (SMILES) string.

The following input SMILES strings were used for calculation of tanimoto distances:

Molecular reference	SMILES string
T1	<chem>C1C(N=C2C=Cc3ccccc3N12)c4ccc5ccccc5c4</chem>
T2	<chem>Cc1c(cc(nc1c2ccccc2)c3ccccc3)c4ccc(Br)cc4</chem>
T3	<chem>O=C1C2=C(C=CC=C2)C3=C1C=CC=C3</chem>
T4	<chem>CC1C(C2=C(C)C=CC2)=CC=C1</chem>
T5	<chem>Cc1ccc(cc1)n2nnnc2c3cnoc3c4ccccc(F)c4</chem>
T6	<chem>Cc1cccc(NC(=O)c2ccc(s2)c3ccccc3)c1C(=O)O</chem>

Docking conformations of highly ranked ligands were compared using LigPlot⁺ v1.4 (run with Java 7.0.250) to generate multiple ligand-protein interaction diagrams.¹⁸

7.3 Chemical methods

7.3.1 General experimental

Reactions were carried out under a nitrogen or argon atmosphere in oven-dried or flame-dried glassware unless otherwise stated. Standard inert atmosphere techniques were used in handling all air and moisture sensitive reagents. All data were gathered at 300 K unless otherwise stated.

Solvents and reagents

Solvents and reagents were used as supplied by major chemical suppliers. Petroleum ether refers to distilled light petroleum of fraction (30 °C – 40 °C). Inorganic solutions are saturated aqueous solutions unless otherwise stated. Brine refers to a saturated aqueous sodium chloride solution. Concentrations of solvents and solutions in general procedures are given relative to 1 eq of starting material.

Silica gel chromatography

Silica gel chromatography was carried out using VWR Kieselgel 60 silica gel (60-63 μm). Thin-layer chromatography (TLC) was carried out using Merck Kieselgel 60 F254 (230-400 mesh) aluminium-supported, fluorescent treated silica, visualised under UV light (λ_{max} 254), followed by thermal development after dipping in one of: a) aqueous solution of ninhydrin, acetic acid, and *n*-butanol; b) aqueous solution of potassium permanganate and hydrochloric acid.

Infrared spectroscopy

IR spectra were recorded on a Bruker Tensor 27 FT-IR spectrometer from a thin film deposited onto a sodium chloride plate or a diamond ATR module. Only selected maximum absorbances (ν_{max}) of the most intense peaks are reported (cm^{-1}): br = broad, s = strong.

Nuclear magnetic resonance (NMR)

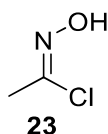
^1H spectra were recorded on a Bruker AVII400 spectrometer (400 MHz) using the deuterated solvent specified as a reference for the internal deuterium lock. The chemical shift data for each

signal (δ_{H}) are reported as follows: (spectrometer frequency; solvent) δ_{H} (no. of protons, multiplicity, J coupling constant where applicable, assignment). Multiplicity is abbreviated as follows: s = singlet, br = broad. ^{13}C spectra were recorded on a Bruker AVII400 spectrometer (101 MHz) using a DEPT Q pulse sequence with broadband proton decoupling and internal deuterium lock. The chemical shift data for each signal (δ_{C}) are reported as follows: (spectrometer frequency; solvent) δ_{C} (assignment). All chemical shifts are quoted on the δ -scale in parts per million (ppm) using residual solvent as the internal standard. NMR spectra are available in Appendix 8.4.

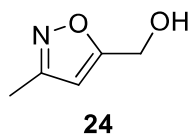
7.3.2 Synthesis of alkylating agent

Compound names are those generated by ChemBioDraw™ (CambridgeSoft) following IUPAC nomenclature.

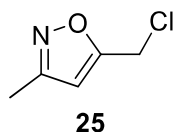
N-Hydroxyacetimidoyl chloride (**23**)



To a stirred solution of acetaldoxime (3.00 g, 50.8 mmol, 1.0 eq) in DMF (54 mL) at 0 °C, was added *N*-chlorosuccinimide (8.82 g, 66.1 mmol, 1.3 eq). The reaction mixture was warmed to rt and stirred for 3 h. The pale green solution was diluted with brine (120 mL), and the product extracted with Et₂O (2 × 100 mL). The combined extracts were then washed with H₂O (3 × 80 mL), dried (MgSO₄), filtered, and concentrated *in vacuo* to afford **23** as a cyan blue oil (6.53 g, 1:1 product:Et₂O, 40 mol, 78%) that was used without further purification, shortly after preparation. ^1H NMR (400 MHz, CDCl₃) δ_{H} 2.18 (3 H, s, CH₃), 9.86 (1 H, br s, OH). Further data could not be collected due to instability, however, these data are in good agreement with the literature values.¹⁹

(3-Methylisoxazol-5-yl)methanol (24**)**

To a stirred solution of propargyl alcohol (2.65 mL, 2.55 g, 45.5 mmol, 1.0 eq) in CH₂Cl₂ (70 mL), protected from light, was added NEt₃ (7.6 mL, 5.52 g, 54.6 mmol, 1.2 eq). A solution of *N*-hydroxyacetimidoyl chloride (**23**, 4.7 g, 50 mmol, 1.1 eq) in CH₂Cl₂ (20 mL) was added dropwise, over 10 h, using a syringe pump. The reaction solution continued to stir for a further 12 h at rt. The reaction mixture was diluted with CH₂Cl₂ (100 mL) and washed with H₂O (3 × 100 mL), before the combined organic layers were dried (MgSO₄), filtered, and concentrated *in vacuo* to afford a dark red oil. Purification by silica gel column chromatography using 30-70% Et₂O/petroleum ether yielded **24** as a yellow oil (582 mg, 5.14 mmol, 10%). R_f 0.28 (40% Et₂O/petroleum ether); ν_{max} (thin film)/cm⁻¹: 3356 (O-H) (br), 1610 (C=C) (s); ¹H NMR (400 MHz, CDCl₃) δ_H 2.13 (3 H, s, C(3)CH₃), 4.55 (2 H, s, CH₂OH), 4.89 (1 H, br s, OH), 5.97 (1 H, s, C(4)H); ¹³C NMR (101 MHz, CDCl₃) δ_C 11.1 (CH₃), 55.8 (CH₂), 102.4 (CH), 159.8 (CO), 171.6 (CN). These data are in good agreement with the literature values.²⁰

5-(Chloromethyl)-3-methylisoxazole (25**: Cl-DMI)**

Freshly prepared (3-methylisoxazol-5-yl)methanol (572 mg, 5.06 mmol, 1.0 eq) was dissolved in thionyl chloride (7.4 mL, 12.0 g, 101 mmol, 20 eq) and the solution heated under reflux for 20 min. The resulting reaction mixture was concentrated *in vacuo*, to yield **25** as a dark brown oil (172 mg, 1.31 mmol, 26% yield) that was stored at -80 °C, and used (assuming a net content of 50% by weight) without further purification. R_f 0.20 (20% MeOH/CH₂Cl₂); ν_{max} (thin film)/cm⁻¹: 1609 (C=C) (s), 1411 (s); ¹H NMR (400 MHz CDCl₃) δ_H 2.29 (3 H, s, CH₃), 4.56 (2 H, s, CH₂),

6.16 (1 H, s, CH); ^{13}C NMR (101 MHz, CDCl_3) δ_{C} 11.3 (CH_3), 34.8 (CH_2), 104.3 (CH), 159.9 (CO), 166.9 (CN). These data are in good agreement with the literature values.²¹

7.3.3 Alkylation

Alkylation reactions were performed in thin walled 500 μL PCR tubes, with a reaction volume of 20 μL to allow adequate volume for samples taken during trials. Conditions corresponding to the various temperatures for alkylation were: 4 $^{\circ}\text{C}$, fridge; 25 $^{\circ}\text{C}$, PCR cycler; 42 $^{\circ}\text{C}$, water bath. pH dependence was assessed with a range of buffers: HEPES pH 7.0, 8.0; CHES pH 9.0, 10.0; CAPS pH 10.0, 11.0. High concentrations of buffer (1 M) minimise variations in pH caused by the molar equivalent of acid liberated during alkylation. Buffers were degassed by sonication under argon immediately prior to use. TCEP was freshly prepared as a 100 mM stock in the corresponding buffer, and 0-2 μL added to reaction mixtures to afford 0-10 eq. of reducing agent. Alkylating agents (Cl-DMI, *N*-methyl maleimide and 2-iodoacetamide) were accurately weighed and prepared as 20 mM stocks, by dissolving in DMSO (20 % of final volume) before dilution into alkylation buffer. Time series commenced upon addition of the alkylating agent to a PCR tube containing all other reagents, pre-equilibrated to the temperature of interest.

Optimised alkylation reaction conditions were: 1 M CHES (pH 9.0), 4 M guanidine hydrochloride, 10 mM l-methionine, 5 mM TCEP, 1 mM Y-H4₁₋₂₀(KAc)₃K12C peptide, 5 mM Cl-DMI, 25 $^{\circ}\text{C}$, 2 h.

7.3.3.1 MALDI

The extent of substrate alkylation was analysed by MALDI-TOF-MS (matrix-assisted laser-desorption ionization–time-of-flight mass spectrometry) using a Bruker Ultraflex instrument. Peptide samples were prepared in a 1:4 ratio from quenched reaction mixture (1:3 dilution of reaction into 95% H_2O , 4.5% acetonitrile, 0.5% formic acid) with CHCA matrix (20 mg/mL α -Cyano-4-hydroxycinnamic acid in 50% acetonitrile) and spotted onto an MTP 384 Massive target plate using the dried droplet method. The instrument was calibrated directly prior to data acquisition using mono-isotopic peptide masses with Peptide Calibration Standard II (Bruker Daltonics,

Coventry, UK). Sample ionization was achieved with a nitrogen laser (337 nm) at 35-50% laser energy and MS spectra were acquired by manual operation in reflectron mode: 2000-3000 data points were collected to make a total intensity of approximately 8×10^4 a.u. Alkylation sites were assigned unambiguously by MALDI-TOF/TOF mass spectrometry: MS/MS spectra were acquired by Laser-induced fragmentation (LIFT).²² Predicted fragmentation masses were calculated using the Fragment Ion Calculator, as part of the Proteomics Toolkit from the Institute for Systems Biology, (<http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html>), and compared manually to spectrum generated.

7.3.3.2 *nano LC/MS/MS*

Liquid chromatography-tandem mass spectrometry was performed using an Ultimate 3000 nanoHPLC system (Dionex, Sunnyvale, CA, USA) comprising a WPS-3000 micro auto sampler, a FLM-3000 flow manager and column compartment, an LPG-3600 dual-gradient micro-pump, and an SRD-3600 solvent rack controlled by Hystar (Bruker Daltonics, Billerica, MA, USA) and DCMS link 2.0 software. For the separation with a C18 Pepmap column (75 μm i.d., 15 cm, Dionex), a flow rate of 250 nL/min was used as generated by a cap-flow splitter cartridge (1/1000). Peptides were eluted by the application of a 80 min gradient using solvents A (98% H_2O , 2% acetonitrile, 0.1% formic acid) and B (80% acetonitrile, 20% H_2O , 0.1% formic acid): 2-10% B, 0-3 min; 10-25% B, 3-18 min; 25-50% B, 18-30 min, 50-90% B, 30-30.2 min. The liquid chromatography was interfaced directly with a 3D high capacity ion trap mass spectrometer (amaZon; Bruker Daltonics) utilizing 10 μm i.d. distal coated SilicaTips (New Objective, Woburn, MA, USA) and nano-ESI mode. Smart parameter settings (SPS) on the ion trap were tuned for the target mass of the $[\text{M}+4\text{H}]^{4+}$ ion: 564.6, 588.3, 612.1, 636.1 m/z for unmodified, mono-, di-, and tri-alkylated peptides respectively. MS/MS analysis was initiated on a contact closure signal triggered by HyStar software (version 3.2). Precursor ions were selected with active exclusion (0.5 min) in collision-induced dissociation (CID) mode. CID fragmentation was achieved using helium gas and a 30%–200% collision energy sweep with amplitude 1.0 (ions are ejected from the trap as soon as they fragment). Raw LC-MS/MS data were processed using DataAnalysis 4.0 software (Bruker Daltonics). Manual

sequence input was combined with fixed modifications (Lysine Acetylation, KAc; Amidation Deacylation, C-terminal KAc) and variable modifications (Cysteine Isoxazole, CDMI).

7.3.3.3 HPLC

Samples from alkylation reactions were monitored by HPLC on a PerkinElmer Flexar system with a Binary LC Pump and UV/VIS LC Detector. 20 μ L samples were loaded onto a pre-equilibrated reversed phase Dionex Acclaim® 120 column (C18, 5 μ m, 120 Å, 4.6 $\times$ 150 mm), and eluted with a standard gradient elution of 0-100% B (A: 95% H₂O, 5% MeCN, 0.1% TFA; B: 5% H₂O, 95% MeCN, 0.1% TFA), at 1 mL/min over 10 min, monitoring absorption at 230 nm (absorption of peptide bonds). Peaks corresponding to mono-alkylation were collected separately, and lyophilised, before resuspension for use in AlphaScreen assays. Collected samples were ran on the same gradient to confirm isolated peptide was a single species.

7.4 References

- (1) Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. a, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–5.
- (2) Philpott, M., Yang, J., Tumber, T., Fedorov, O., Uttarkar, S., Filippakopoulos, P., Picaud, S., Keates, T., Felletar, I., Ciulli, A., Knapp, S., and Heightman, T. D. (2011) Bromodomain-peptide displacement assays for interactome mapping and inhibitor discovery. *Mol. Biosyst.* 7, 2899–2908.
- (3) Theodoulou, N. H. (2012) Development of Unnatural Amino Acid Acetyl-Lysine Mimics. University of Oxford.
- (4) Leatherbarrow, R. J. (2009) GraFit Version 7. Erithacus Software Ltd. Horley, U.K.
- (5) Dalvit, C. (1998) Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *J. Biomol. NMR* 11, 437–444.
- (6) Tsai, W.-W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C.-Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M.-C., Patel, D. J., and Barton, M. C. (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927–932.
- (7) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 213–221.
- (8) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2126–2132.
- (9) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D.

- G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- (10) Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.
- (11) Linding, R. (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31, 3701–3708.
- (12) Abagyan, R., Totrov, M., and Kuznetsov, D. (1993) ICM - A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J. Comput. Chem.* 15, 488–506.
- (13) Gonnet, G. H., Cohen, M. a, and Benner, S. a. (1992) Exhaustive matching of the entire protein sequence database. *Science* 256, 1443–1445.
- (14) Choi, J. H., Jung, H. Y., Kim, H. S., and Cho, H. G. (2000) PhyloDraw: a phylogenetic tree drawing system. *Bioinformatics* 16, 1056–1058.
- (15) Ross, G. A., Morris, G. M., and Biggin, P. C. (2012) Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* 7, e32036.
- (16) O’Boyle, N. M., Banck, M., James, C. a., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An Open chemical toolbox. *J. Cheminform.* 3, 33.
- (17) Trott, O., and Olson, A. J. (2010) AutoDock Vina : Improving the Speed and Accuracy of Docking with a New Scoring Function , Efficient Optimization , and Multithreading. *J Comput Chem* 31, 455–461.
- (18) Laskowski, R. A., and Swindells, M. B. (2011) LigPlot + : Multiple Ligand Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* 2778–2786.
- (19) Zhang, J., and Curran, D. P. (1991) Stereoselective synthesis of 1,2-diols by the cycloadditive strategy: total synthesis of (±)-exo-brevicomine and (±)-and (–)-pestalotin. *J. Chem. Soc. Perkin Trans. 1* 2627–2631.
- (20) Al-Busafi, S., Al-Belushi, M., and Al-Muqbali, K. (2010) Short Synthesis of COX-2 Inhibitor Inotilone. *Synth. Commun.* 40, 1088–1092.
- (21) Gadzhily, R. A., and Aliev, A. A. (2002) Synthesis and Properties of 3-Alkyl (aryl) -5-chloromethylisoxazoles 38, 415–418.
- (22) Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J., and Holle, A. (2003) A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal. Chem.* 376, 952–965.

Chapter 8: Appendices

8.1 Computational processing scripts

colour by displaceability

```

1 27 March 2013
  Angelina Measures

3
4 Separate hydrated Apo crystal structure in protein and water position files i.e.
5   PyMOL>fetch 3o33
6   PyMOL>extract 3o33A, 3o33 & chain A
7   PyMOL>extract 3o33waters, 3o33A & resn HOH
8   PyMOL>cmd.save("3o33waters.pdb", "3o33waters")
9   PyMOL>cmd.save("3o33A.pdb", "3o33A")
10 Note: although this cannot include water-water interactions, over a large dataset nearby
11     waters generate too much noise
12 displaceability was only significant when other waters were not present and interactions were
13     to desolvated protein only

14
15 Convert protein file into pdbqt:
16   pdb2pqr via online server *PROPKA* - http://nbcrc.net/pdb2pqr
17 Note: could be done with MOE
18   upload 3o34A.pdb
19   AMBER force field
20   AMBER output naming scheme
21   x new atoms not too close
22   x optimise H-bonding
23   x protonation at pH 7.0

24
25 Convert .pqr to .pdbqt with prepare_receptor4.py
26   /home/angelina/MGLTools-latest/MGLToolsPckgs/AutoDockTools/Utilities24/prepare_receptor4.py -r
27     3o33A.pqr -v -o 3o33A.pdbqt -A checkhydrogens -U nphs_lps -d 3o33A_dictionary.txt
28 # -r input -v verbose -o output -A repairs only if no hydrogens -U merge non-polar hydrogens
29   and lone pairs but DON'T remove waters -d dictionary

30
31 Run WaterClassifier
32   $ chmod +x WaterClassifier.sh
33   $ ./WaterClassifier.sh 3o33waters.pdb 3o33A.pdbqt

34
35 Parse results for properties to append to pdb file
36 Take from '#' to end of file, -s squeeze the spaces into 1 character, keep 2nd field from
37   space delimited cut, delete header P(D)
38   $ sed -n '/^#/, $p' WaterProbabilities.txt | tr -s ' ' | cut -d ' ' -f 2 | sed '1d' >
39     WaterDisplaceabilityOnly.txt

40
41 Reassign B-factors to displacability and colour PyMOL representation
42 # load the protein
43   cmd.load("3o33A.pdb")

44
45 # open the file of new values (just 1 column of numbers, one for each alpha carbon)
46   inFile = open("WaterDisplaceabilityOnly.txt", 'r')

47
48 # create the global, stored array
49   stored.newB = []

50
51 # read the new B factors from file
52   for line in inFile.readlines(): stored.newB.append( float(line) )

53
54 # close the input file
55   inFile.close()

56
57 # clear out the old B Factors
58   alter 3o33waters, b=0.0

59
60 # update the B Factors with new properties
61   alter 3o33waters and resn HOH, b=stored.newB.pop(0)

62
63 # color the protein based on the new B Factors of the alpha carbons
64   cmd.spectrum("b", "red_white_green", "3o33waters")

```

```

59 # save file with new B-factors
60 select 3o33waters_displacability, 3o33A + 3o33waters
61 cmd.save("3o33waters_displacability.pdb", "3o33waters_displacability")
62
63 Delete waters based on displaceability threshold
64 Note: Lack of water-water interactions means probability fails for highly conserved,
65 structural KAc pocket waters
66 remove these manually from selection using a Bayesian statistical rationale with low
67 displaceability prior for 4xHOH
68 # select by threshold
69 extract displaceable, 3o33waters & b > 0.62 &! resi 127+109+117
70 # save as pdb the protein with remaining waters
71 cmd.save("3o33damp.pdb", "3o33A + 3o33waters")
72
73 Convert this to pdbqt for use in docking
74 http://nbcrc.net/pdb2pqr then prepare_receptor4.py
75
76 23 July 2012
77 By Gregory Ross
78
79 With these files and scripts, one is able to perform the water classification as described in
80 "G. Ross et al., PLOS One, Vol 7, Issue 3, e32036".
81
82 First, you need water positions (which should be in a pdb or pdbqt file format) and the
83 protein structure you wish to score against (which has to be a pdbqt file). The protein
84 structure file should be of the apo protein and should not contain any overlapping water
85 positions.
86
87 The classification can be carried out using the shell script "WaterClassifier.sh". This script
88 calls a python script that gets the water molecules'
89 structural information and also calls the water classifiers using the statistical language R.
90 It is easiest to give this file executable permission, ie.
91
92 $ chmod +x WaterClassifier.sh
93
94 If the file with water positions is called "WaterPositions.pdb" and the apo protein structure
95 file is called "ApoProtein.pdbqt", then the script is
96 used in the following manner:
97
98 $ ./WaterClassifier.sh WaterPositions.pdb ApoProtein.pdbqt
99
100 The output of the above process is called "WaterProbabilities.txt". It gives the water
101 displacement probabilities in the order that the water molecules
102 appeared in "WaterPositions.pdb". Note that two probabilities are outputted for each water
103 molecule. The second column is the probability a water is displaced by a ligand. The
104 probability it is conserved is one minus this value. The third column is the probability a
105 water is displaced by a non-polar
106 group. As before, the probability it is displaced by a polar group is one minus this value.
107
108 It is important to note that the accuracy of the classifiers (within the test sets)
109 improved with higher or lower predicted probabilities. That is, if a water molecule is
110 predicted to be displaced with a probability between 0.4 and 0.6, the outcome is relatively
111 uncertain. Likewise, if a water molecule is predicted to be displaced with a probability
112 above 0.8 or below 0.2, one can expect the prediction to be more trustworthy.

```

config.txt

```

# location of receptor file relative to working directory
2 receptor = ./3o33damp.pdbqt
3
4 # coordinates of centre of docking area
5 center_x = 46.57
6 center_y = 41.06
7 center_z = -23.78
8
9 # grid size for docking
10 size_x = 20
11 size_y = 20
12 size_z = 20
13
14 # energetic exhaustivity
15 num_modes = 20

```

vina_screen_local.sh

```

1 #!/bin/bash
3 #PBS -V
4 #PBS -l nodes=1:ppn=16
5 #PBS -l walltime=100:00:00
6 #PBS -N TRIM33PHD
7 #PBS -m bea
8 #PBS -M angelina.measures@dtc.ox.ac.uk
9 #PBS -V
11 # go to the directory where job is submitted
12 cd $PBS_O_WORKDIR
13
14 for f in Maybridge*.pdbqt; do
15     b=`basename $f .pdbqt`
16     mkdir -p $b
17     vina --cpu 16 --config ./config_C1009_T33PHD.txt --ligand $f --out ${b}/out.pdbqt --log ${b}/log.txt
18 done

```

zip_results.sh

```

2 ##### written: 130328
3 ##### latest edit: 130328
4 ##### ZIP TOP RESULTS #####
5 ## script to analyse results from docking screen on supercomputer
6
7 ## Instructions for use
8 # replace PROTEIN with protein used e.g. TRIM24dryZA
9 # replace LIGAND with database used e.g. MaybD/MaybF
10 # replace NN with number of LIGANDs you want to extract
11 # use in folder containing:
12 # #
13 # in run by typing
14
15 ## Analyse results
16 ./vina_screen_get_top.py NN > PROTEIN_LIGAND_results.csv
17 # grep Maybridge lines, only keep directory names, add 'here' to start of lines and save as a
18 # directories file
19 grep Maybridge PROTEIN_LIGAND_results.csv | cut -d '\t' -f1 | sed 's/^./\t/' > directory_list.
20 # use this list to zip them
21 cat directory_list.txt | zip -r PROTEIN_LIGAND_docked.zip -@

```


vina_screen_get_top_quant.py

```

2  #!/usr/bin/env python
# change LIGANDFOLDERS (line 8) to standard prefix e.g. SD/*.pdbqt
# if you have already converted files to .pdb change that too
6 import sys
import glob
8
def doit(n):
10     file_names = glob.glob('LIGANDFOLDERS/*.pdbqt')
    # define empty matrices for results
12     everything = []
    failures = []
14     # the number of files is the length of matrix
    print 'Found', len(file_names), 'pdbqt files'
16     for file_name in file_names:
        file = open(file_name)
18     # read file split by line
        lines = file.readlines()
        file.close()
20        try:
            # the first line in the file
22            line = lines[1]
24            #
            result = float(line.split(':')[1].split()[0])
            everything.append([result, file_name])
26        except:
            # if this doesn't work put it in the error matrix
            failures.append(file_name)
30    everything.sort(lambda x,y: cmp(x[0], y[0]))
    part = everything[:n]
32    # part is everything from the start to n (number specified)
    for p in part: # for each one
34        # print the file name. Comma means no newline
        print p[1],
36        # print the result (affinity)
        print p[0]
38    print
    if len(failures) > 0:
40        print 'WARNING:', len(failures), 'pdbqt files could not be processed'
42
if __name__ == '__main__':
    doit(int(sys.argv[1]))

```

make_sdf.sh

```

1  #### written: 130304
   #### latest edit: 130305
3
   ##### MAKE .SD FILE #####
5  ## Make sdf file and append rank and energy
6
7  ## Instructions for use
   # replace RECEPTOR with protein used e.g. TRIM24dryZA
9  # replace LIGANDS with database used e.g. MaybD/MaybF
   # use in folder containing SDF_toolkit
11 http://cactus.nci.nih.gov/SDF_toolkit/
   # change address to
13 # # .sdf containing all numbered ligands
   # # RECEPTOR_LIGAND_results.csv
15
   # Extract results for SD file
17 sed 's/MaybridgeF//g' /Conway/Docking/OSC/TRIM24dryZA_tightC_MaybridgeD/
   TRIM24dryZA_tightC_MaybD_top100names.txt > /Conway/Docking/OSC/TRIM24dryZA_tightC_MaybridgeD/
   TRIM24dryZA_tightC_MaybD_top100identifiers.txt
   select_sdf -seqfile /Conway/Docking/OSC/TRIM24dryZA_tightC_MaybridgeD/
   TRIM24dryZA_tightC_MaybD_top100identifiers.txt < /Conway/Software/ANGIE/
   Maybridge_Ro3_Fragment_Library_fixed.sdf > /Conway/Docking/OSC/
   TRIM24dryZA_tightC_MaybridgeD/TRIM24dryZA_tightC_MaybD_top100.sdf

```

make_pml.sh

```

##### written: 130226
##### latest edit: 130304

##### MAKE .PML SCRIPT #####
## script to analyse results from docking screen

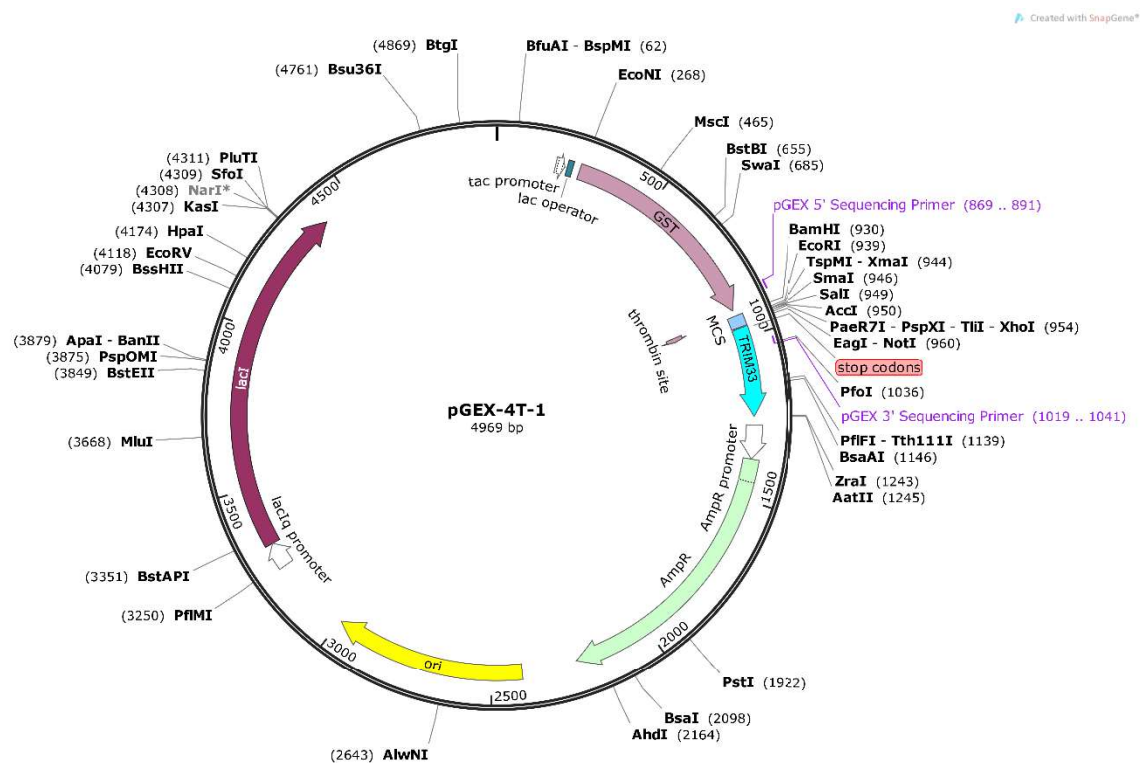
## Instructions for use
# replace RECEPTOR with protein used e.g. TRIM24dryZA
# replace LIGAND with database used e.g. MaybD/MaybF
# replace NN with number of ligands you want to view
# replace MM with number of ligand you want in result file
# change the location your files are in (line 32) - note escape characters
# use in folder containing:
# # vina_screen_get_top_quant.py: change ./LIGAND*/*.pdbqt
# # folders with ./FOLDER/out.pdbqt docking results
# # view_results_start.txt: change RECEPTOR loaded with address and view
# # view_results_end.txt: change RECEPTOR and LIGAND
# # RECEPTOR.pdb
# in pymol run by typing @C://user/folders/address/view_topNN_results_RECEPTOR_LIGAND.pml

## Analyse results
# edited get_top.py to output the affinity as well as file names (included line print p[0],)
./vina_screen_get_top_quant.py MM > RECEPTOR_LIGAND_results.csv
# remove the file pointer to leave names
sed -e '1d' -e 's/\\out.pdbqt/\\/g' RECEPTOR_LIGAND_results.csv > temp.csv
mv temp.csv RECEPTOR_LIGAND_results.csv
# keep top NN to view
head -NN RECEPTOR_LIGAND_results.csv > RECEPTOR_LIGAND_topNN.csv
# names only
cut -d, -f1 RECEPTOR_LIGAND_topNN.csv > RECEPTOR_LIGAND_topNNnames.txt
# parse these to make a .pml script
sed -e 's/^/load C://Users\\NAME\\Documents\\FOLDERS\\/\\/g' RECEPTOR_LIGAND_topNNnames.txt >
    RECEPTOR_LIGAND_topNN_start.txt
sed -e 's/^/\\out.pdbqt, /g' RECEPTOR_LIGAND_topNNnames.txt > RECEPTOR_LIGAND_topNN_end.txt
# merge all of these by line
paste -d '' RECEPTOR_LIGAND_topNN_start.txt RECEPTOR_LIGAND_topNN_end.txt >
    RECEPTOR_LIGAND_topNN_files.txt
# put the start and end on the pymol script
cat view_results_start.txt RECEPTOR_LIGAND_topNN_files.txt view_results_end.txt >
    view_topNN_results_RECEPTOR_LIGAND.pml

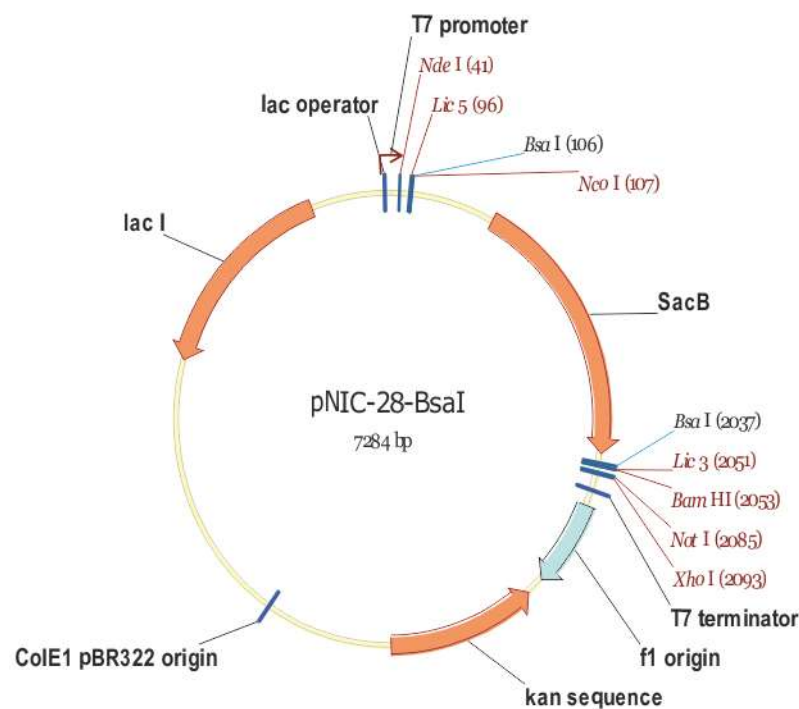
```

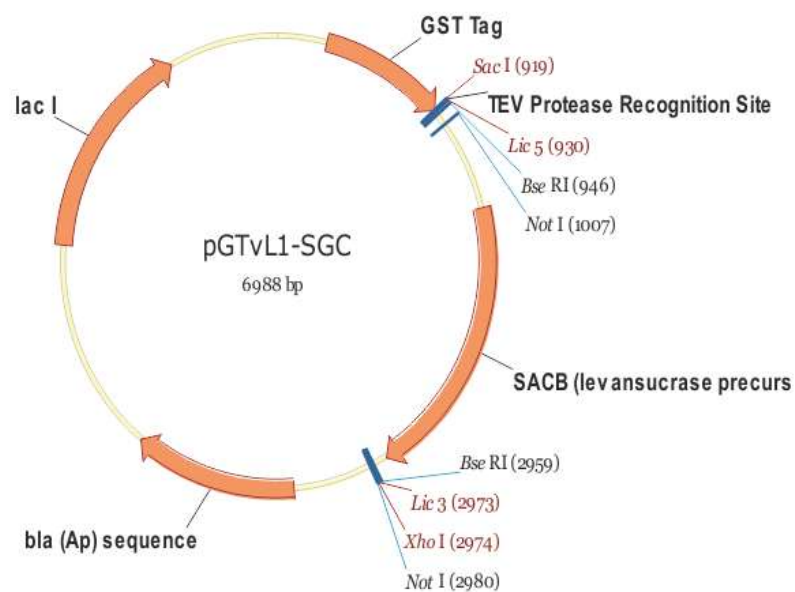
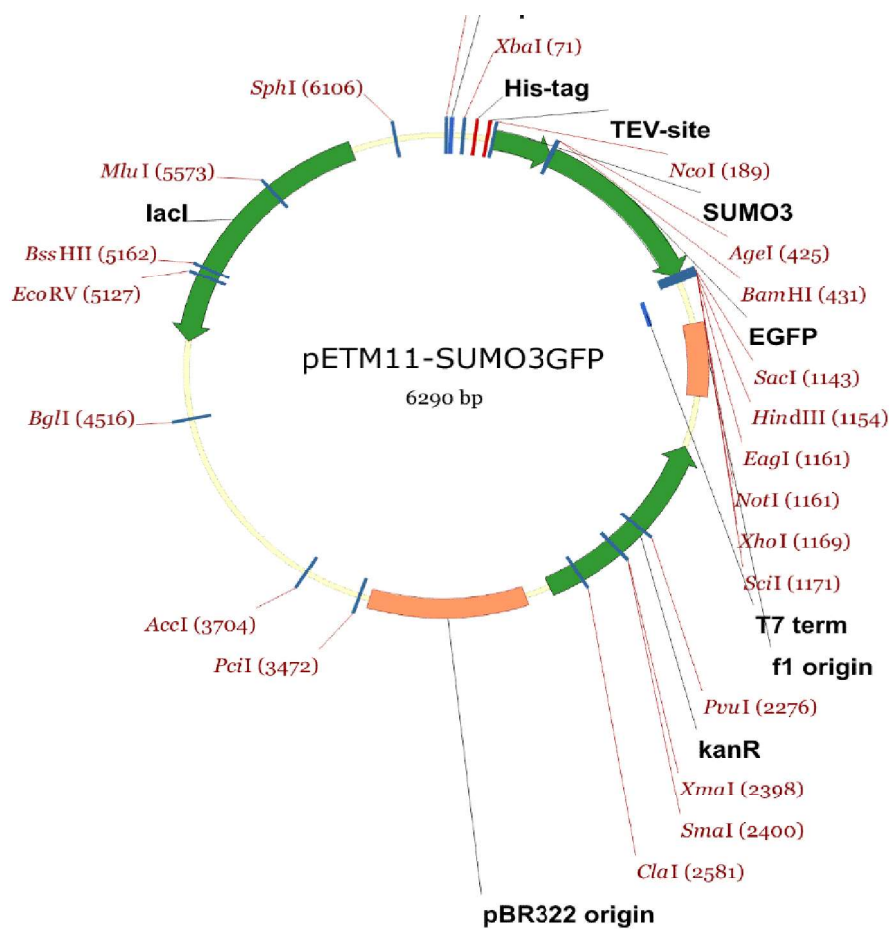
8.2 Vectors

pGEX-4T1



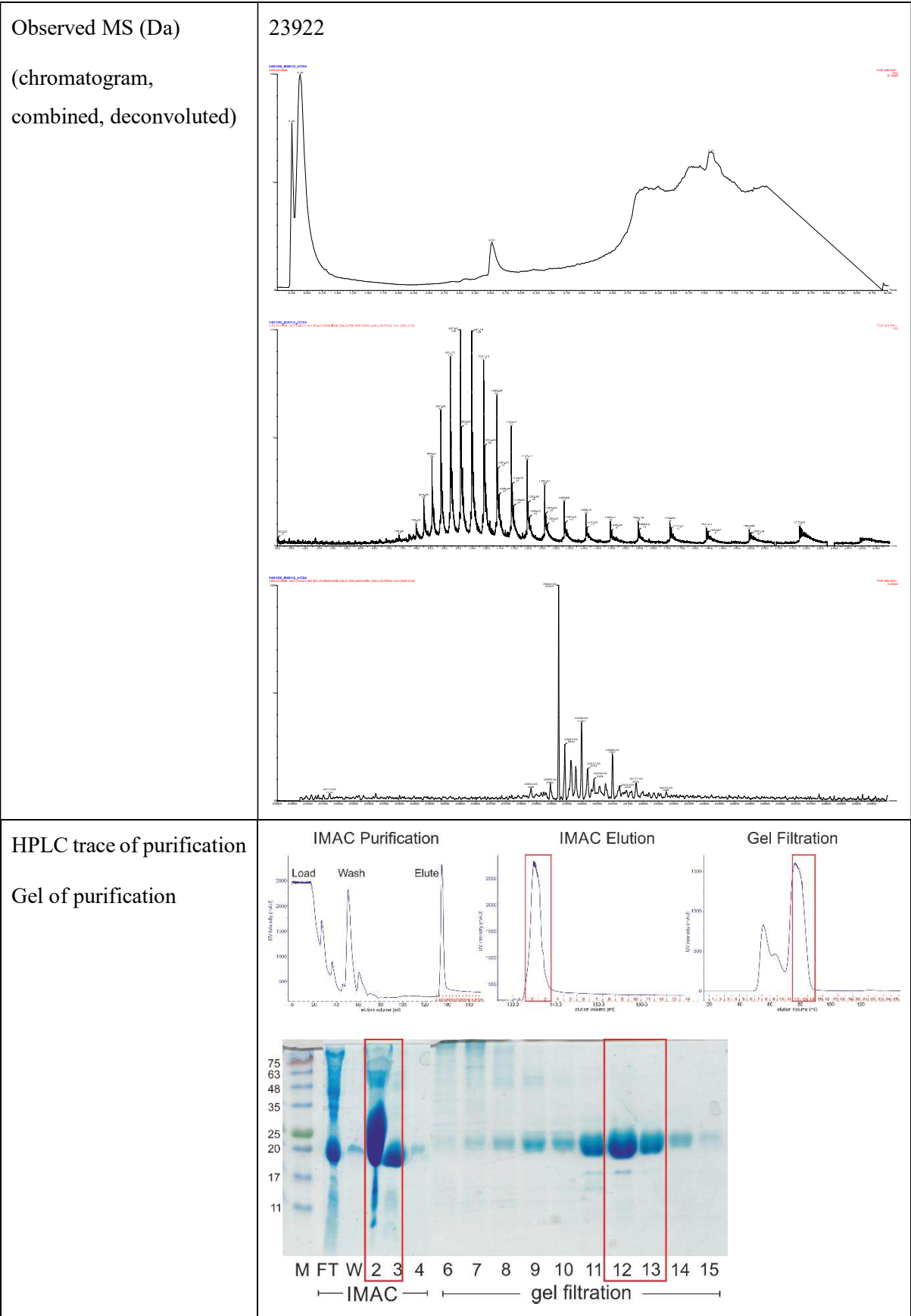
pNIC28-Bsa4

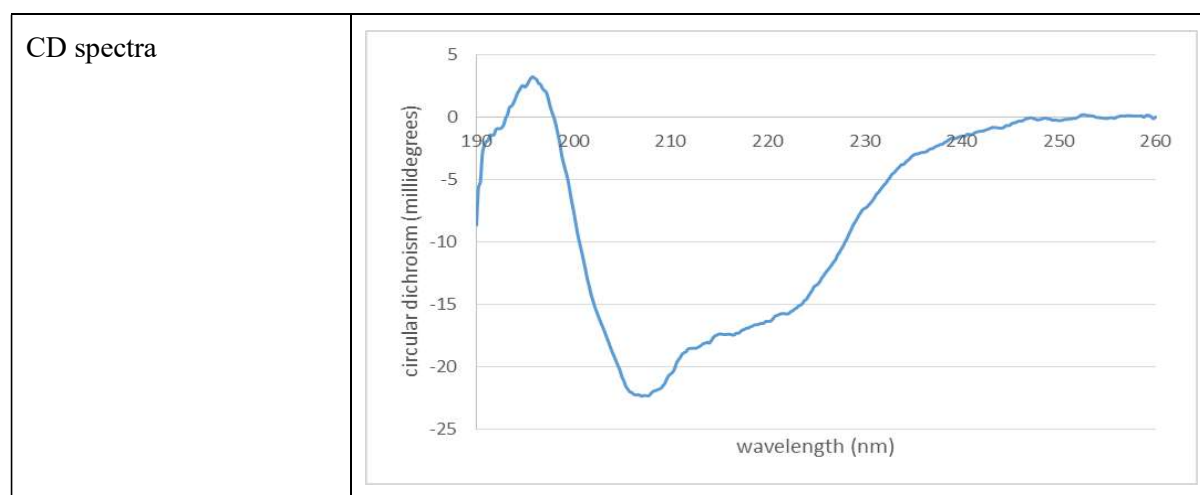


pGTVvL1-SGC**pETM11-SUMO3GFP**

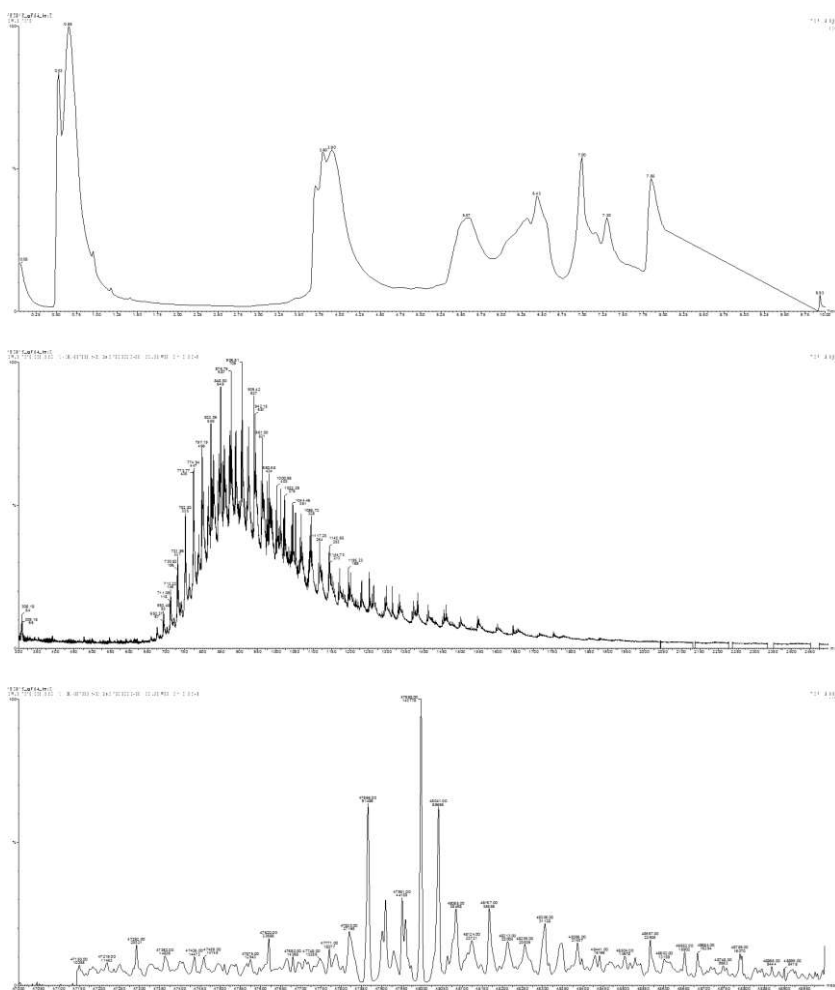
8.3 Protein characterisation data

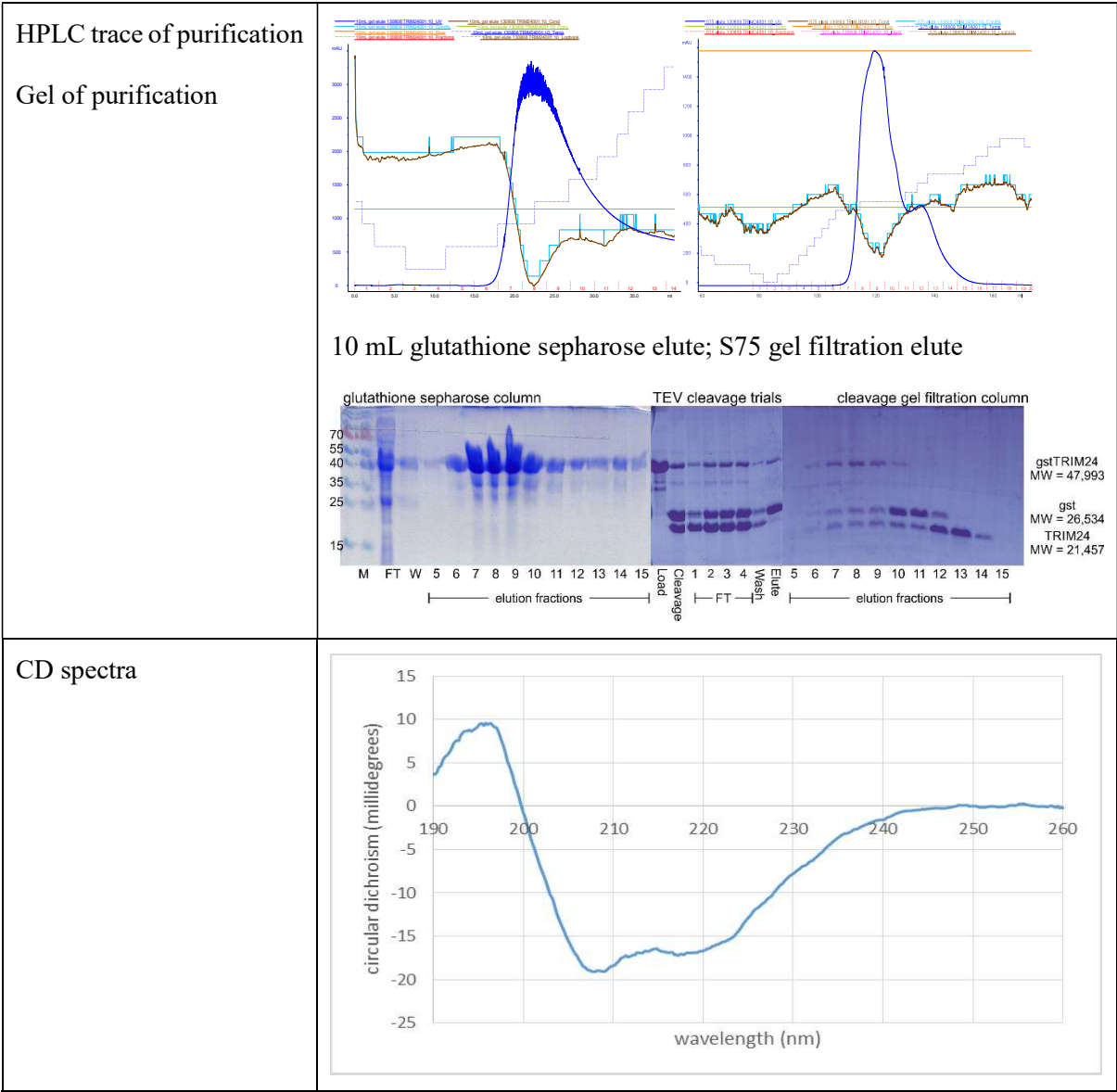
Gene	His₆-TRIM24 WT
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Mutagenesis correction from SGC supplied plasmid
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGAATGAGGACTGGTGTGCAGTTTGTGTC AAAACGGAGGGGAACCTCCTCTGCTGTGAAAAGTGCCCCAAAGTATTC CATCTTTCTTGTGTCATGTGCCACATTGACAAATTTTCCAAGTGGAGA GTGGATTTGCACTTTCTGCCGAGACTTATCTAAACCAGAAGTTGAAT ATGATTGTGATGCTCCCAGTCACAACTCAGAAAAAAGAAAACTGAA GGCCTTGTTAAGTTAACACCTATAGATAAAAGGAAGTGTGAGCGCCT ACTTTTATTTCTTTACTGCCATGAAATGAGCCTGGCTTTTCAAGACC CTGTTCTCTAACTGTGCCTGATTATTACAAAATAATTAATAATCCA ATGGATTTGTCAACCATCAAGAAAAGACTACAAGAAGATTATTCCAT GTACTCAAAACCTGAAGATTTTGTAGCTGATTTTAGATTGATCTTTC AAAAGTGTGCTGAATTCAATGAGCCTGATTCAGAAGTAGCCAATGCT GGTATAAAACTTGAAAATTATTTTGAAGAACTTCTAAAGAACCTCTA TCCAGAATGA
Amino acid sequence	MHHHHHSSGVDLGTENLYFQSMNEDWCAVCQNGGELLCCCEKCPKVF HLSCHVPTLTNFPSEWICTFCRDLSKPEVEYDCDAPSHNSEKKKTE GLVKLTPIDKRKCERLLFLYCHEMSLAFQDPVPLTVPDYKIIKNP MDLSTIKKRLQEDYSMYSKPEDFVADFRLIFQNCAEFNEPDSEVANA GIKLENYFEELLKNLYPE-
Calculated aa; MW (Da); ε (all cystines); pI	206; 23922.2; 25160; 5.24





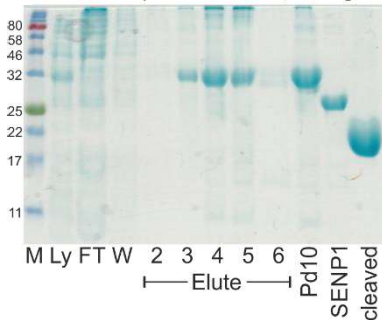
Gene	GST-TRIM24 WT
Vector backbone	pGTVL1-SGC (Ampicillin)
Source	LIC cloning from pNIC28-Bsa4 construct
Affinity label : cleavage enzyme	GST : TEV protease
DNA sequence	AAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCA CAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGC TTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTCGAGAATTGCA TATAGTAAAGACTTTGAACTCTCAAAGTTGATTTTCTTAGCAAGCT ANNNGAAATGCTGAAAATGTTCTGAAGATCGTTTATGTCATAAAACAT ATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGAC GCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTT CCCAAATTAGTTTGTTTTAAAAACGTATTGAAGCTATCCCACAAA TTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAG GGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGAG CTCAGAGAACCTGTACTTCCAATCCATGAATGAGGACTGGTGTGCAG TTTGTCAAACGGAGGGGAACCTCTGCTGTGAAAAGTGCCCCAAA GTATTCCATCTTTCTTGTCATGTGCCACATTGACAAATTTTCCAAG TGGAGAGTGGATTTGCACTTTCTGCCGAGACTTATCTAAACCAGAAG TTGAATATGATTGTGATGCTCCAGTCACAACCTCAGAAAAAAGAAA ACTGAAGGCCTTGTTAAGTTAACACCTATAGATAAAAGGAAGTGTGA GCGCCTACTTTTATTTCTTTACTGCCATGAAATGAGCCTGGCTTTTC AAGACCCTGTTCTCTAACTGTGCCTGATTATTACAAAATAATTAAA AATCCAATGGATTTGTCAACCATCAAGAAAAGACTACAAGAAGATTA

	TTCCATGTACTCAAACCTGAAGATTTTGTAGCTGATTTTAGATTGA TCTTTCAAACTGTGCTGAATTCAATGAGCCTGATTCAGAAGTAGCC AATGCTGGTATAAACTTGAAAATTATTTGAAGAACTTCTAAAGAA CCTCTATCCAGAATGA
Amino acid sequence	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE LGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNM LGGCPKERAEISM LEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKT YLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAI PQ IDKYLKSSKYIAWPLQGWQATFGGDHPPKSSSENLYFQSMNEDWCA VCQNGGELLCCCKPKVFHLSCHVPTLTNFPSGEWICTFCRDLSKPE VEYDCDAPSHNSEKKKTEGLVKLTPIDKRKCERLLLFLYCHEMSLAF QDPVPLTVPDYYKIIKNPMDLSTIKKRLQEDYSMYSKPEDFVADFRL IFQNCAEFNEPDSEVANAGIKLENYFEELLKNLYPE-
Calculated aa; MW (Da); ϵ (all cystines); pI	412; 47993.3; 68270; 5.32
Observed MS (Da) (chromatogram, combined, deconvoluted)	47998 

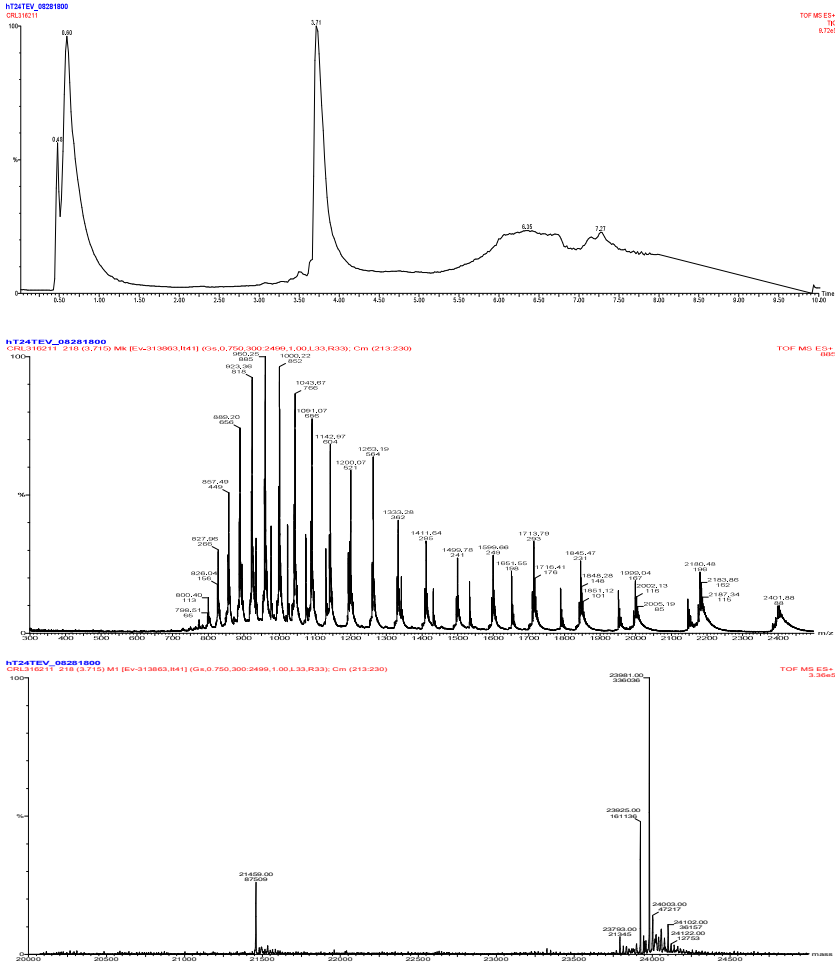
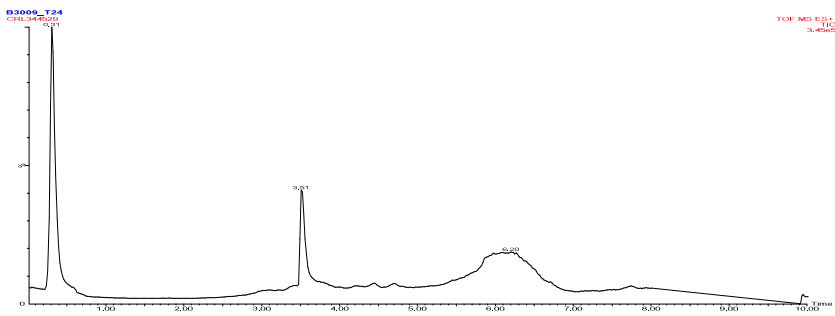


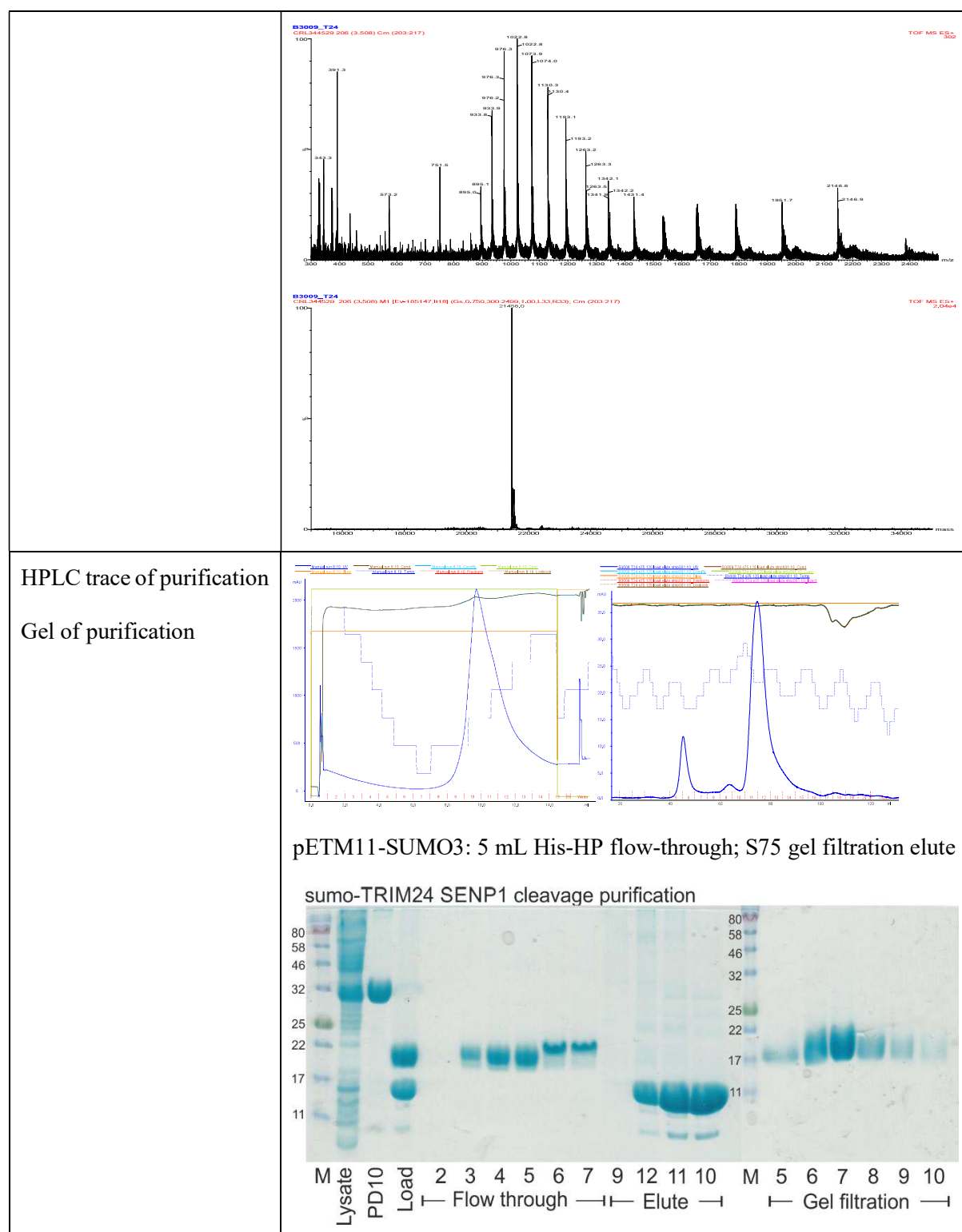
Gene	His ₆ -SUMO-TRIM24 WT
Vector backbone	pETM11-SUMO3 (Kanamycin)
Source	Gibson assembly
Affinity label : cleavage enzyme	His ₆ -SUMO : SENP1
DNA sequence	ATGAAACATCACCATCACCATCACCCATGAGCGATTACGACATCCC CACTACTGAGAATCTTTATTTTCAGGGCGCCATGGGCAACGATCACA TTAACCTGAAAGTGGCCGGTCAAGACGGTAGCGTAGTCCAGTTTAAA ATCAAACGCCACACCCCTCTGTCGAAACTGATGAAAGCCTATTGTGA ACGCCAAGGTCTGTCTATGCGTCAGATCCGTTTTTCGCTTCGATGGAC

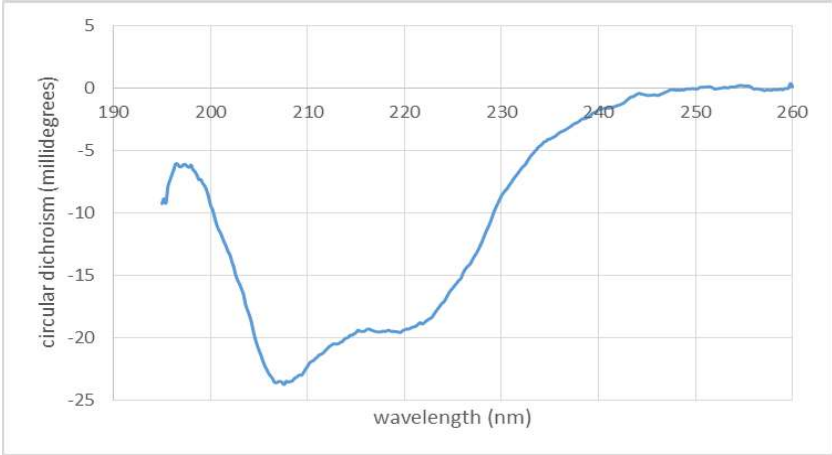
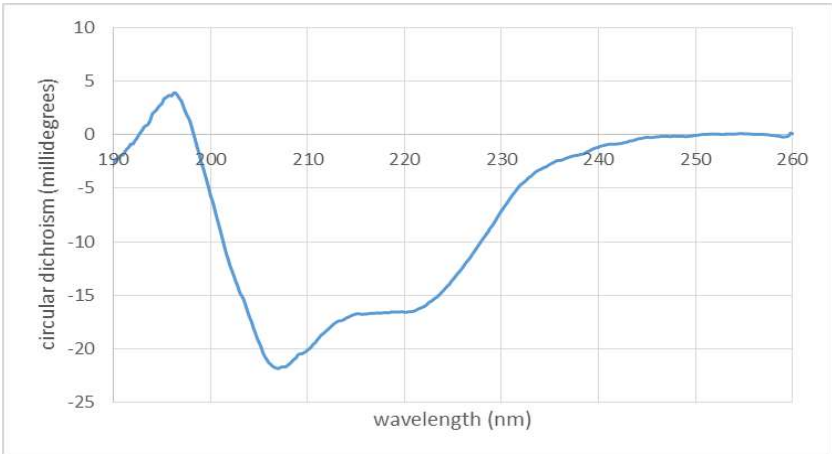
	AGCCGATTAACGAAACCGACACTCCAGCACAGCTGGAAATGGAAGAT GAGGACACCATTGACGTGTTCCAGCAACAGACCGGTGGATCCATGAA TGAGGACTGGTGTGCAGTTTGTCAAAACGGAGGGGAACTCCTCTGCT GTGAAAAGTGCCCCAAAGTATTCCATCTTTCTTGTGCATGTGCCACACA TTGACAAATTTTCCAAGTGGAGAGTGGATTTGCACTTTCTGCCGAGA CTTATCTAAACCAGAAGTTGAATATGATTGTGATGCTCCCAGTCACA ACTCAGAAAAAAGAAAAGTGAAGGCCTTGTTAAGTTAACACCTATA GATAAAGGAAGTGTGAGCGCCTACTTTTATTTCTTTACTGCCATGA AATGAGCCTGGCTTTTCAAGACCCTGTTTCCTCTAACTGTGCCTGATT ATTACAAAATAATTAAAAATCCAATGGATTTGTCAACCATCAAGAAA AGACTACAAGAAGATTATTCCATGTACTCAAACCTGAAGATTTTGT AGCTGATTTTAGATTGATCTTTCAAACCTGTGCTGAATTCATGAGC CTGATTCAGAAGTAGCCAATGCTGGTATAAACTTGAAAATTATTTT GAAGAACTTCTAAAGAACCTCTATCCAGAATGA
Amino acid sequence	MKHHHHHHHPMSDYDIPPTENLYFQGAMGNDHINLKVAGQDGSVVQFK IKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMED EDTIDVFQQQTGGSMNEDWCAVCQNGGELLCCCEKCPKVFHLSCHVPT LTNFPSGEWICTFCRDLSKPEVEYDCDAPSHNSEKKKTEGLVKLTPI DKRK CERLLLFLYCHEMSLAFQDPVPLTVPDYYKIIKNPMDLSTIKK RLQEDYSMYSKPEDFVADFRLIFQNCAEFNEPDSEVANAGIKLENYF EELLKNLYPE-
Calculated aa; MW (Da); E (all cystines); pI	292; 33804.4; 28140; 5.31
Observed MS (Da) (chromatogram, combined, deconvoluted)	33809



Gene			TRIM24 WT
Cleaved sequence	amino acid		SMNEDWCAVCQNGGELLCCCEKCPKVFHLSCHVPTLTNFPSGEWICTF CRDLSKPEVEYDCDAPSHNSEKKKTEGLVKLTPIDKRK CERLLLFLY CHEMSLAFQDPVPLTVPDYYKIIKNPMDLSTIKKRLQEDYSMYSKPE DFVADFRLIFQNCAEFNEPDSEVANAGIKLENYFEELLKNLYPE-

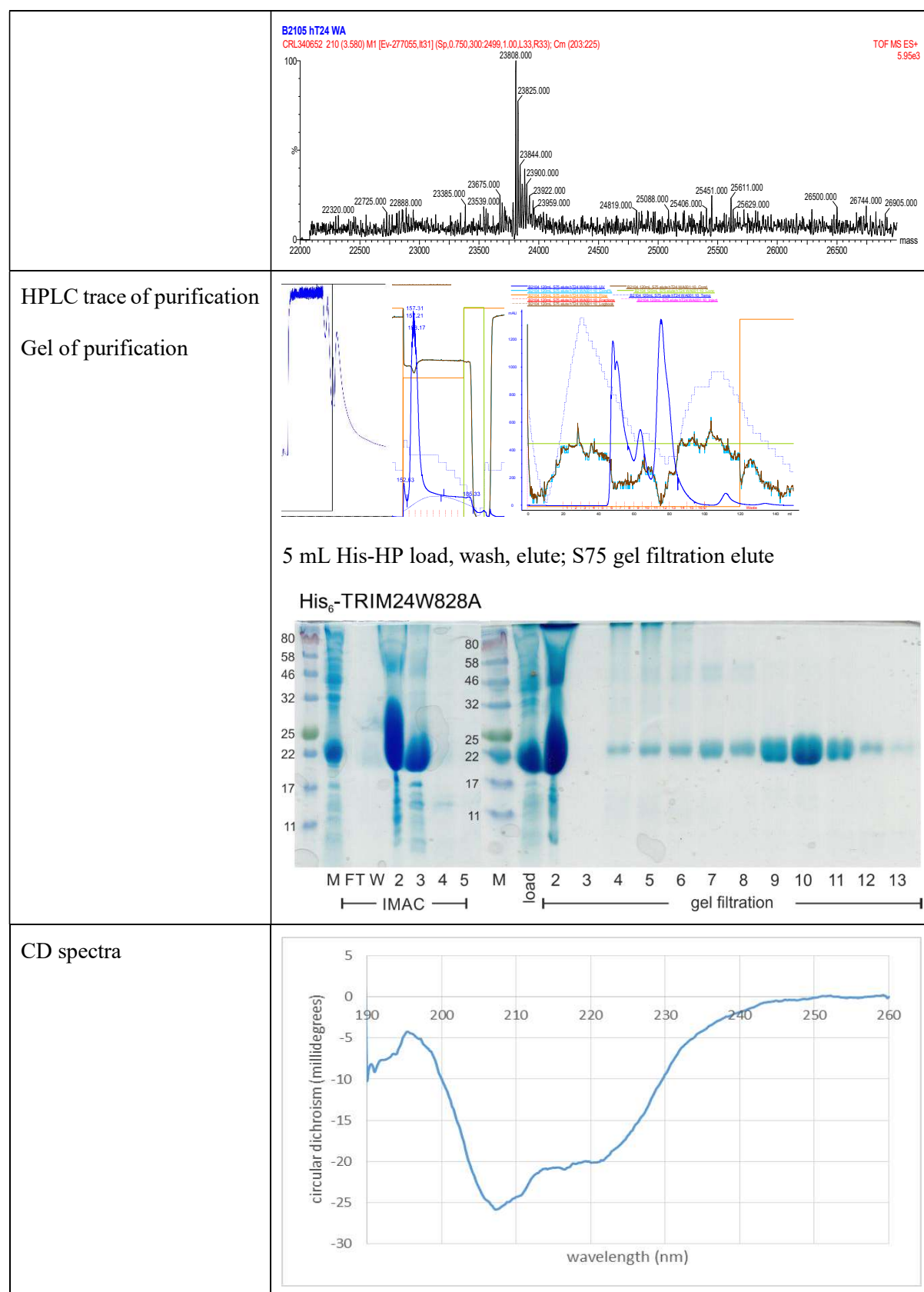
Calculated aa; MW (Da); ε (all cystines); pI	185; 21456.5; 23670; 4.92
Observed MS (Da) (chromatogram, combined, deconvoluted)	Cleavage from pNIC28-Bsa4: minor product 21459  Cleavage from pETM11-SUMO3: 21458 



CD spectra	Cleavage from pNIC28-Bsa4  Cleavage from pETM11-SUMO3 
------------	---

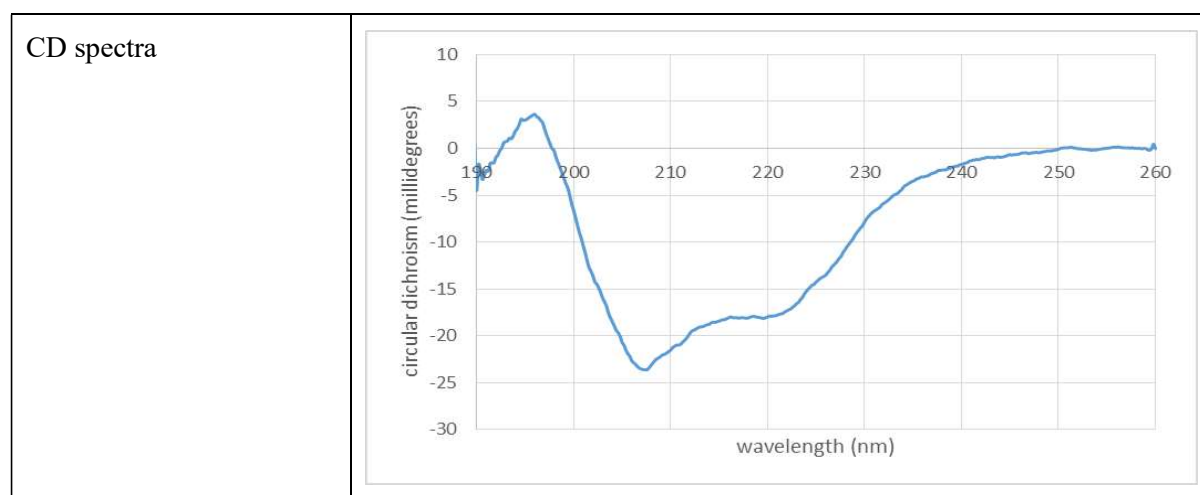
Gene	His₆-TRIM24 W828A
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from WT
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGAATGAGGACGCGTGTGCAGTTTGTC AAAACGGAGGGGAACTCCTCTGCTGTGAAAAGTGCCCCAAAGTATTC CATCTTTCTTGTGCATGTGCCACATTGACAAATTTTCCAAGTGGAGA GTGGATTTGCACTTTCTGCCGAGACTTATCTAAACCAGAAGTTGAAT ATGATTGTGATGCTCCAGTCACAACCTCAGAAAAAAGAAAAGTAA

	GGCCTTGTTAAGTTAACACCTATAGATAAAAGGAAGTGTGAGCGCCT ACTTTTATTTCTTTACTGCCATGAAATGAGCCTGGCTTTTCAAGACC CTGTTTCCTCTAACTGTGCCTGATTATTACAAAATAATTAATAATCCA ATGGATTTGTCAACCATCAAGAAAAGACTACAAGAAGATTATTCCAT GTACTCAAAACCTGAAGATTTTGTAGCTGATTTTAGATTGATCTTTC AAAAGTGTGCTGAATTCAATGAGCCTGATTCAGAAGTAGCCAATGCT GGTATAAACTTGAAAATTATTTTGAAGAACTTCTAAAGAACCTCTA TCCAGAATGA
Amino acid sequence	MHHHHHHSSGVDLGTENLYFQSMNEDACAVCQNGGELLCEKCPKVF HLSCHVPTLTNFPSEWICTFCRDLSKPEVEYDCDAPSHNSEKKKTE GLVKLTPIDKRKCELLLLFLYCHEMSLAFQDPVPLTVPDYKIIKNP MDLSTIKKRLQEDYSMYSKPEDFVADFRLIFQNCAEFNEPDSEVANA GIKLENYFEELLKNLYPE-
Calculated aa; MW (Da); ε (all cystines); pI	206; 23807.0; 19660; 5.24
Observed MS (Da) (chromatogram, combined, deconvoluted)	23808 B2105 hT24 WA CRL340652 TOF MS ES+ TIC 1.15e6 B2105 hT24 WA CRL340652 210 (3.580) Mk [Ev+277055,1131] (Sp,0.750,300,2499,1.00,L33,R33); Cm (203,225) TOF MS ES+ 45.5



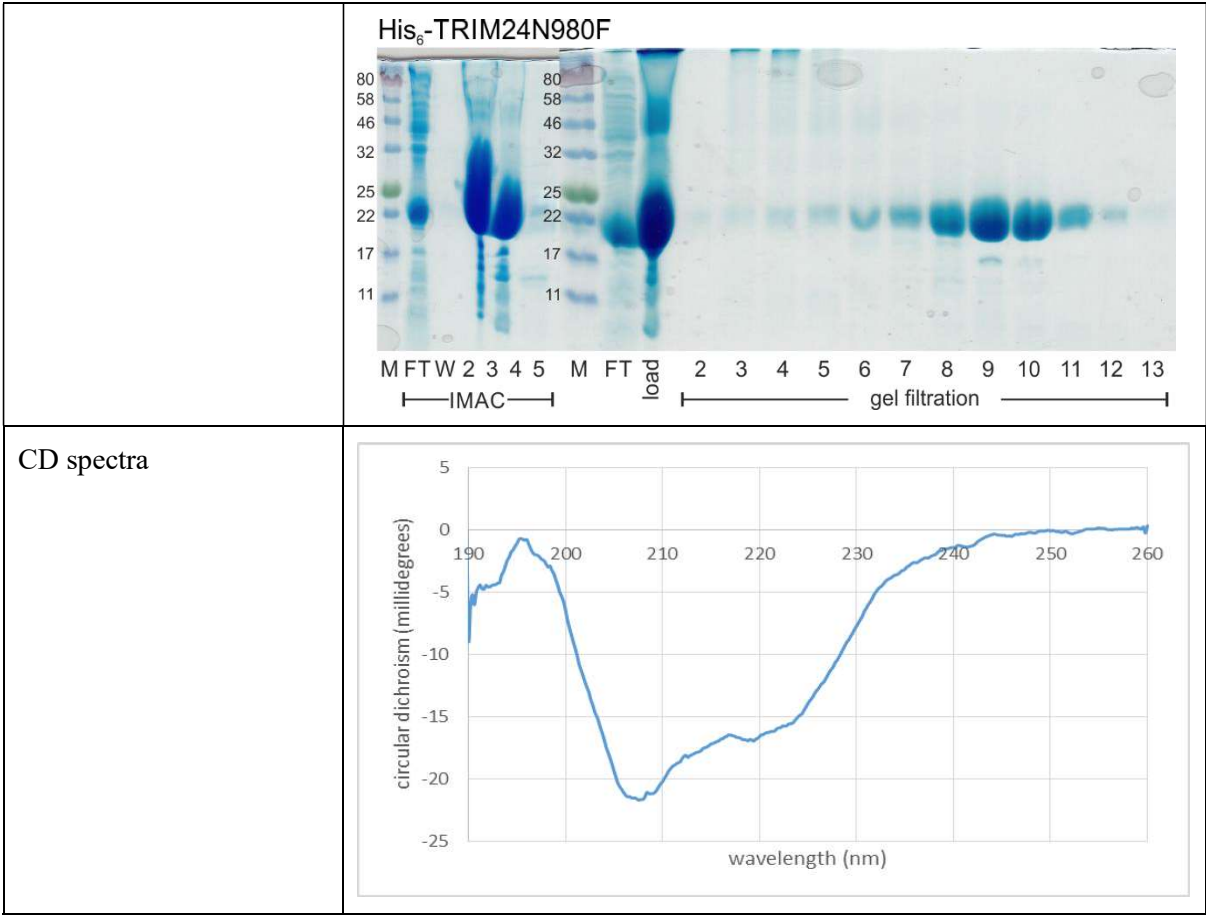
Gene	His₆-TRIM24 N980A
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from WT
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGAATGAGGACTGGTGTGCAGTTTGTG AAAACGGAGGGGAACTCCTCTGCTGTGAAAAGTGCCCCAAAGTATTC CATCTTTCTTGTGATGTGCCACATTGACAAATTTTCCAAGTGGAGA GTGGATTTGCACCTTCTGCCGAGACTTATCTAAACCAGAAGTTGAAT ATGATTGTGATGCTCCCAGTCACAACTCAGAAAAAAGAAAACTGAA GGCCTTGTTAAGTTAACACCTATAGATAAAAGGAAGTGTGAGCGCCT ACTTTTATTTCTTTACTGCCATGAAATGAGCCTGGCTTTTCAAGACC CTGTTCCCTCTAACTGTGCCTGATTATTACAAAATAATTAATAATCCA ATGGATTTGTCAACCATCAAGAAAAGACTACAAGAAGATTATTCCAT GTACTCAAAACCTGAAGATTTTGTAGCTGATTTTAGATTGATCTTTC AAAAGTGTGCTGAATTCGCTGAGCCTGATTCAGAAGTAGCCAATGCT GGTATAAACTTGAAAATTATTTTGAAGAACTTCTAAAGAACCTCTA TCCAGAATGA
Amino acid sequence	MHHHHHHSSGVDLG TENLYFQSMNEDWCAVCQNGGELLCEKCPKVF HLSCHVPTLTNFP SGEWICTFCRDLSKPEVEYDCDAPSHNSEKKKTE GLVKLTPI DKRK CERLLLFLYCHEMSLAFQDPVPLTVPDYKIIKNP MDLSTIKKRLQEDYS MYSKPEDFVADFRLIFQNCAEFAEPDSEVANA GIKLENYFEELLKNLYPE-
Calculated aa; MW (Da); ε (all cystines); pI	206; 23879.2; 25160; 5.24
Observed MS (Da) (chromatogram, combined, deconvoluted)	23881





Gene	His₆-TRIM24 N980F
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from WT
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGAATGAGGACTGGTGTGCAGTTTGTC AAAACGGAGGGGAACCTCCTCTGCTGTGAAAAGTGCCCCAAAGTATTC CATCTTTCTTGTCATGTGCCACATTGACAAATTTTCCAAGTGGAGA GTGGATTTGCACCTTCTGCCGAGACTTATCTAAACCAGAAGTTGAAT ATGATTGTGATGCTCCCAGTCACAACTCAGAAAAAAGAAAACTGAA GGCCTTGTTAAGTTAACACCTATAGATAAAAGGAAGTGTGAGCGCCT ACTTTTATTTCTTTACTGCCATGAAATGAGCCTGGCTTTTCAAGACC CTGTTCCCTCTAACTGTGCCTGATTATTACAAAATAATTAATAATCCA ATGGATTTGTCAACCATCAAGAAAAGACTACAAGAAGATTATTCCAT GTACTCAAAACCTGAAGATTTTGTAGCTGATTTTAGATTGATCTTTC AAAACTGTGCTGAATTCTTTGAGCCTGATTCAGAAGTAGCCAATGCT GGTATAAACTTGAAAATTATTTTGAAGAACTTCTAAAGAACCTCTA TCCAGAATGA
Amino acid sequence	MHHHHHSSGVDLGTENLYFQSMNEDWCAVCQNGGELLCEKCPKVF HLSCHVPTLTNFPSEWICTFCRDLSKPEVEYDCDAPSHNSEKKKTE GLVKLTPIDKRK CERLLLFLYCHEMSLAFQDPVPLTVPDYKIIKNP

	MDLSTIKKRLQEDYSMYSKPEDFVADFRLLIFQNCAEFFEPDSEVANA GIKLENYFEELLKNLYPE-
Calculated aa; MW (Da); E (all cystines); pI	206; 23955.3; 25160; 5.24
Observed MS (Da) (chromatogram, combined, deconvoluted)	23957 B2105 ht24 NF CRL340651 TOF MS ES+ TIC 1.12e6 B2105 ht24 NF CRL340651 212 (3.610) Cm (207.223) TOF MS ES+ 340 B2105 ht24 NF CRL340651 212 (3.610) M1 [Ev-252219.1t29] (Sp,0.750,300;2499,1.00,L33,R33); Cm (207.223) TOF MS ES+ 6.62e3
HPLC trace of purification Gel of purification	5 mL His-HP load, wash, elute; S75 gel filtration elute



Gene	His ₆ -TRIM24 W828A N980F
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from single mutant
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCTTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGAATGAGGACGCGTGTGCAGTTTGTG AAAACGGAGGGGAACTCCTCTGCTGTGAAAAGTGCCCCAAAGTATTC CATCTTTCTTGTCATGTGCCACATTGACAAATTTTCCAAGTGGAGA GTGGATTTGCACTTTCTGCCGAGACTTATCTAAACCAGAAGTTGAAT ATGATTGTGATGCTCCCAGTCACAACTCAGAAAAAAGAAAAGTAA GGCCTTGTTAAGTTAACACCTATAGATAAAAGGAAGTGTGAGCGCCT ACTTTTATTTCTTTACTGCCATGAAATGAGCCTGGCTTTTCAAGACC CTGTTCCCTCTAACTGTGCCTGATTATTACAAAATAATTAATAATCCA ATGGATTTGTCAACCATCAAGAAAAGACTACAAGAAGATTATTCCAT

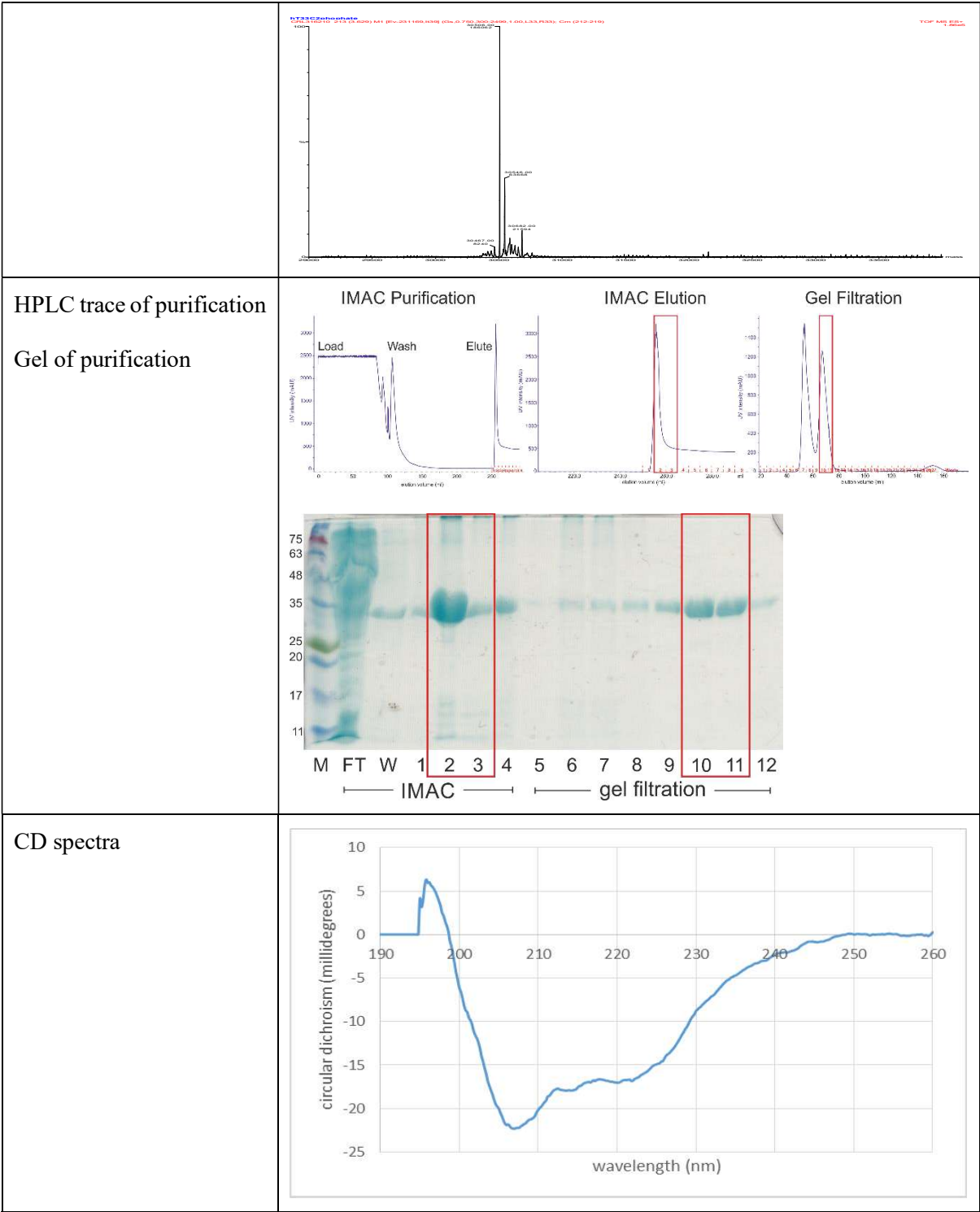
	GTACTCAAAACCTGAAGATTTTGTAGCTGATTTTAGATTGATCTTTC AAACTGTGCTGAATTCTTTGAGCCTGATTCAGAAGTAGCCAATGCT GGTATAAAACTTGAAAATTATTTTGAAGAACTTCTAAAGAACCTCTA TCCAGAATGA
Amino acid sequence	MHHHHHSSGVDLG TENLYFQSMNEDACAVCQNGGELLCCCEKCPKVF HLSCHVPTLTNFPSEWICTFCRDLSKPEVEYDCDAPSHNSEKKKTE GLVKLTPIDKRK CERLLLFLYCHEMSLAFQDPVPLTVPDYKIIKNP MDLSTIKKRLQEDYSMYSKPEDFVADFRLIFQNCAEFFEPDSEVANA GIKLENYFEELLKNLYPE-
Calculated aa; MW (Da); pI (all cystines); pI	206; 23840.1; 19660; 5.24

Gene	GST-TRIM33 WT
Vector backbone	pGEX-4T1 (Ampicillin)
Source	Provided by Prof. Joan Massagué
Affinity label : cleavage enzyme	GST : TEV protease
DNA sequence	TCCAAAGGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGG ATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAA ACTCTCAAAGTTGATTTCTTAGCAAGCTACCTGAAATGCTGAAAATG TTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGT AACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTAT ACATGGACCCAATGTGCCTGGATGCGTTCCCAAATTAGTTTGT AAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATC CAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTG GTGGTGGCGACCATCCTCCAAATCGGATCTGGTTCCGCGTGGATCC ACCATGGACTGGTGTGCTGTCTGCCAAAACGGAGGAGATCTCTTG CTGCGAAAAATGTCCAAAGTCTTTCATCTAAGTTGTCATGTTCCAA CACTACTTAGCTTTCCAAGTGGGGACTGGATATGCACATTTTGTAGA GATATTGAAAGCCAGAAGTTGAATATGATTGTGATAATTGCAACA TAGTAAGAAGGGGAAAACTGCGCAGGGGTTAAGCCCCGTGGACCAAA GGAAATGTGAACGTCTTCTGCTTTACCTCTATTGCCATGAATTAAGT ATTGAATTCCAGGAGCCTGTTCTGCTTCGATACCAAACTACTATAA AATTATAAAGAAACCAATGGATTTATCCACCGTGAAAAAGAAGCTTC

	AGAAAAAACATTCCCAACACTACCAAATCCCGGATGACTTTGTGGCC GATGTCCGTTTGATCTTCAAGAACTGTGAAAGGTTAATGAAATGAT GAAAGTTGTTCAAGTTTATGCAGACACACAAGAGATTAATTTGAAGG CTGATTCAAGTAGCTCAGGCAGGGAAAGCAGTTGCATTGTACTTT GAAGATAAACTCACAGAGATCTACTCAGACAGGACCTTCGCACCTTT GCCAGAGTTTGAGCAGGAAGAGGATGATGGTGAGGTAAGTGAAGACT CTGATGAAGACTTTATACAGCCCCGCAGAAAACGCCTAAAGTCAGAT GAGAGACCAGTACATATAAAGTAA
Amino acid sequence	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE LGLFEPNLPYYIDGDVKLTQSMAIIRYIADKHNMMLGGCPKERAIEISM LEGAVLDIRYGVSR IAYS KDFETLKVDFLSKLP EMLKMFEDRLCHK T YLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAI PQ IDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRGSTMDWCAVCQ NGGDL LCCEKCPKVFHLTCHVPTLLSFPSGDWICTFCRDIGKPEVEY DCDNLQH SKKGKTAQGLSPVDQRKCERLLLYLYCHELSIEFQEPVPA SIPNYYKIIKKPMDLSTVKKKLQKKHSQHYQIPDDFVADVRLIFKNC ERFNEMMKVVQVYADTQEINLKADSEVAQAGKAVALYFEDKLTEIYS DRTFAPLPEFEQEEDDGEVTEDESDEDFIQPRRKRLKSDERP VHIK-
Calculated aa; MW (Da); pI (all cystines); pI	468; 54342.5; 68270; 5.65

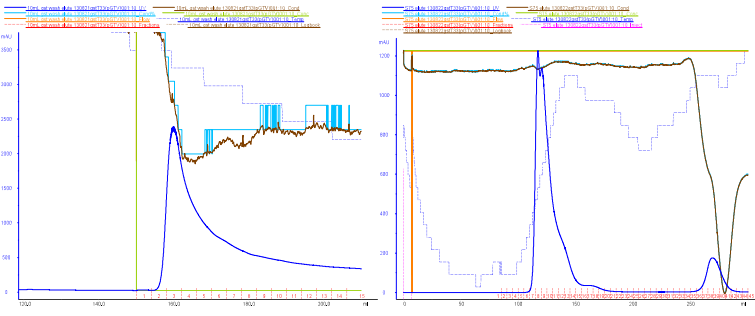
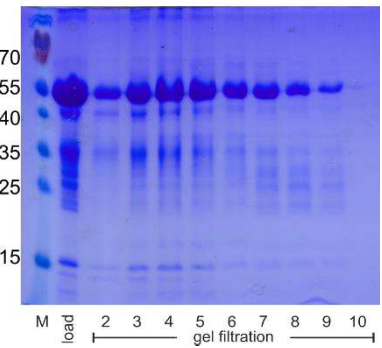
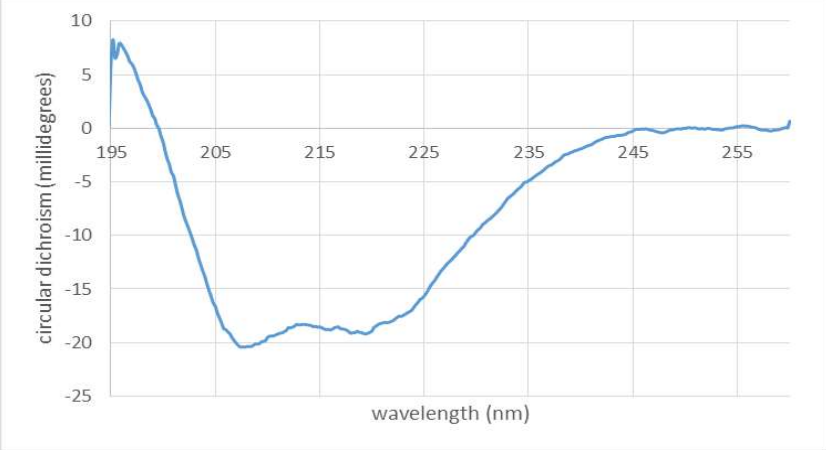
Gene	His₆-TRIM33 WT
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	LIC cloning from pGEX-4T1 construct
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGGACTGGTGTGCTGTCTGCCAAAACG GAGGAGATCTCTTGTGCTGCGAAAAATGTCCAAAGGTCTTTCATCTA ACTTGTCATGTTCCAACACTACTTAGCTTTCCAAGTGGGGACTGGAT ATGCACATTTTGTAGAGATATTGGAAAGCCAGAAGTTGAATATGATT GTGATAATTTGCAACATAGTAAGAAGGGGAAAAC TGCGCAGGGGTTA AGCCCCGTGGACCAAAGGAAATGTGAACGTCTTCTGCTTTACCTCTA TTGCCATGAATTAAGTATTGAATTCCAGGAGCCTGTTCTGCTTCGA

	TACCAAACACTACTATAAAATTATAAAGAAACCAATGGATTTATCCACC GTGAAAAAGAAGCTTCAGAAAAAACATTCCCAACACTACCAAATCCC GGATGACTTTGTGGCCGATGTCCGTTTGATCTTCAAGAACTGTGAAA GGTTTAATGAAATGATGAAAGTTGTTCAAGTTTATGCAGACACACAA GAGATTAATTTGAAGGCTGATTGAGAAGTAGCTCAGGCAGGGAAAGC AGTTGCATTGTACTTTGAAGATAAACTCACAGAGATCTACTCAGACA GGACCTTCGCACCTTTGCCAGAGTTTGAGCAGGAAGAGGATGATGGT GAGGTAACAGAGGACTCTGATGAAGACTTTATACAGCCCCGCAGAAA ACGCCTAAAGTCAGATGAGAGACCAGTACATATAAAGTAA
Amino acid sequence	MHHHHHSSGVDLG TENLYFQ SMDWCAVCQNGDLLCCEKCPKVFHL TCHVPTLLSFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKGKTAQGL SPVDQRKCERLLLYLYCHELSIEFQEPVPASIPNYYKIIKKPMDLST VKKKLQKKHSQHYQIPDDFVADVRLIFKNCERFNEMMKVVQVYADTQ EINLKADSEVAQAGKAVALYFEDKLTEIYSDRTFAPLPEFEQEEDDG EVTEDSDEDFIQPRRKRLKSDERPVIK-
Calculated aa; MW (Da); E (all cystines); pI	263; 30502; 26650; 5.62
Observed MS (Da) (chromatogram, combined, deconvoluted)	30506



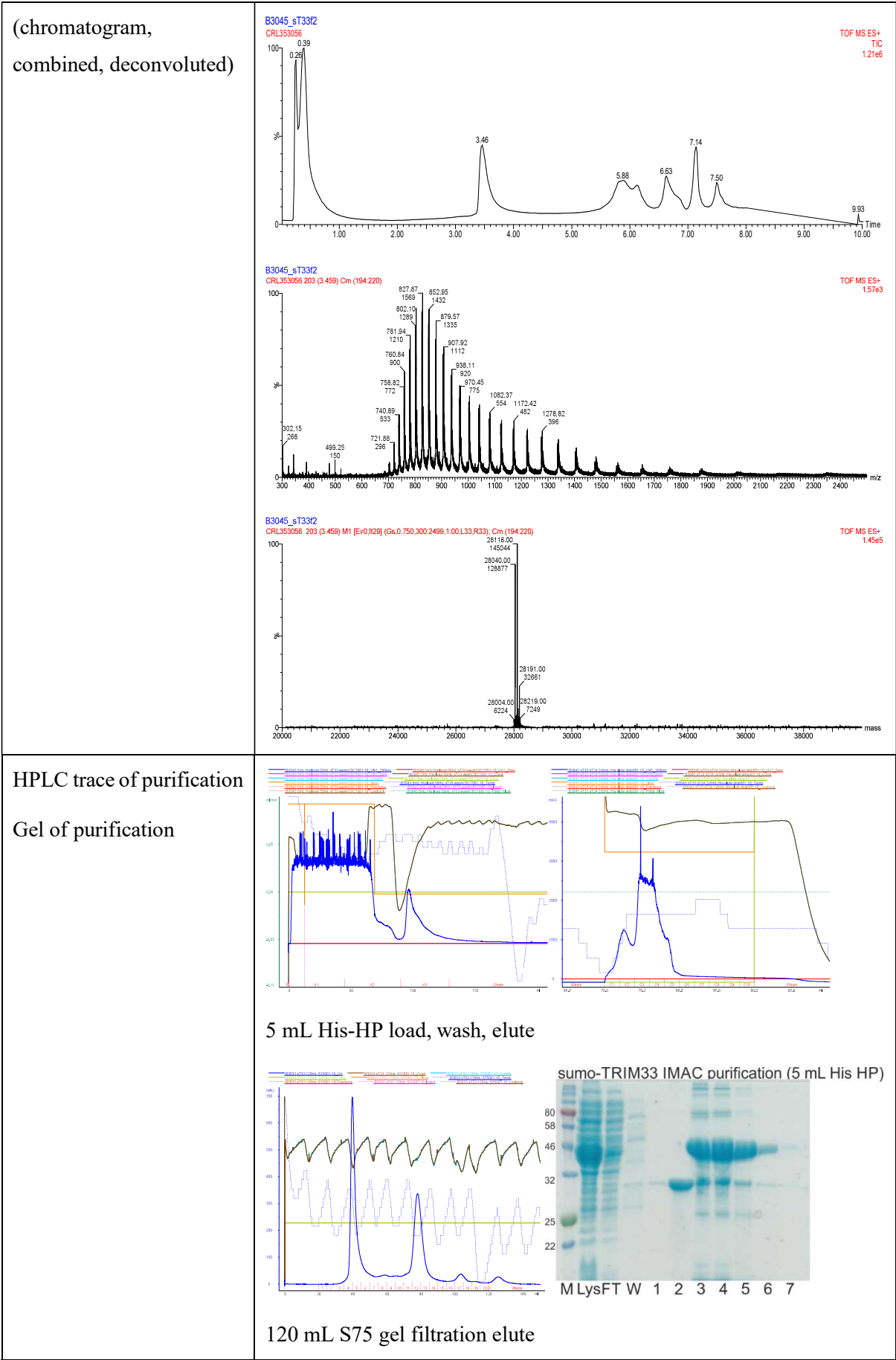
Gene	GST-TRIM33 WT
Vector backbone	pGTVL1-SGC (Ampicillin)
Source	LIC cloning from pGEX-4T1 construct

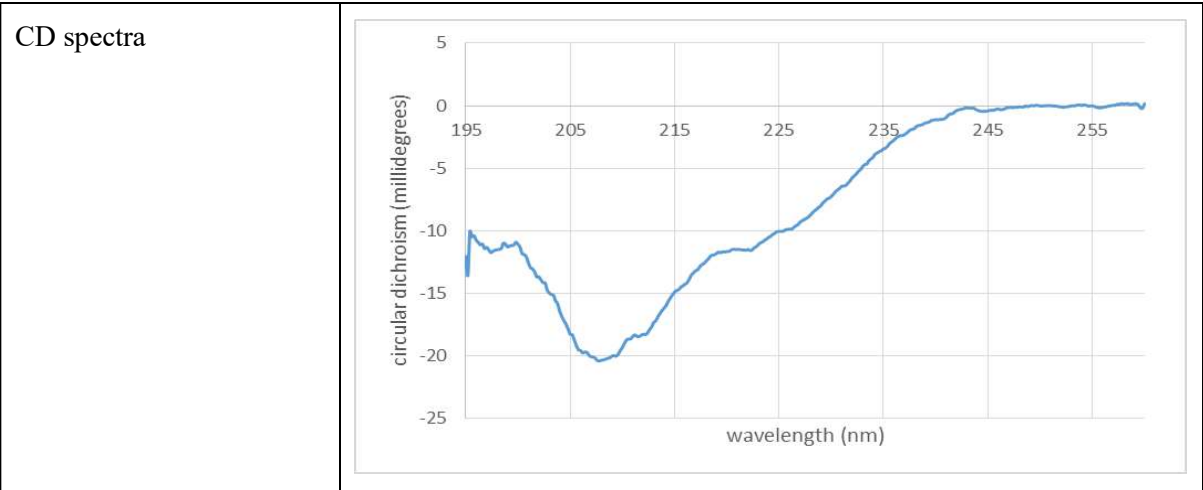
Affinity label : cleavage enzyme	GST : TEV protease
DNA sequence	GAAATGCTGAAAATGTTCTGAAGATCGTTTATGTCATAAAACATATTT AAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTC TTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTTCCCA AAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGA TAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCT GGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGAGCTCA GAGAACCTGTACTTCCAATCCATGGACTGGTGTGCTGTCTGCCAAAA CGGAGGAGATCTCTTGTGCTGCGAAAAATGTCCAAAGGTCTTTCATC TAACTTGTCATGTTCCAACACTACTTAGCTTTCCAAGTGGGGACTGG ATATGCACATTTTGTAGAGATATTGGAAAGCCAGAAGTTGAATATGA TTGTGATAATTTGCAACATAGTAAGAAGGGGAAAACGCGCAGGGGT TAAGCCCCGTGGACCAAAGGAAATGTGAACGTCTTCTGCTTTACCTC TATTGCCATGAATTAAGTATTGAATTCCAGGAGCCTGTTCTGCTTC GATACCAAACACTACTATAAAATTATAAAGAAACCAATGGATTTATCCA CCGTGAAAAAGAAGCTTCAGAAAAAACATCCCCAACACTACCAAATC CCGGATGACTTTGTGGCCGATGTCCGTTTGATCTTCAAGAACTGTGA AAGGTTTAATGAAATGATGAAAGTTGTTCAAGTTTATGCAGACACAC AAGAGATTAATTTGAAGGCTGATTTCAGAAGTAGCTCAGGCAGGGAAA GCAGTTGCATTGTACTTTGAAGATAAACTCACAGAGATCTACTCAGA CAGGACCTTCGCACCTTTGCCAGAGTTTGAGCAGGAAGAGGATGATG GTGAGGTAACCTGAGGACTCTGATGAAGACTTTATACAGCCCCGCAGA AAACGCCTAAAGTCAGATGAGAGACCAGTACATATAAAGTAA
Amino acid sequence	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFE LGLFNPPLPYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAIEISM LEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKT YLNQDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFFKKRIEAI PQ IDKYLKSSKYIAWPLQGWQATFGGGDHPPKSSSENLYFQSMDWCAVC QNGGDLCCCEKCPKV FHLTCHVPTLLSFPSGDWICTFCRDIGKPEVE YDCDNLQHSSKKGKTAQGLSPVDQRKCERLLLYLYCHELSIEFQEPVP ASIPNYYKIIKKPMDLSTVKKKLQKKHSQHYQIPDDFVADVRLIFKN CERFNEMMKVVQVYADTQEINLKADSEVAQAGKAVALYFEDKLTEIY SDRTFAPLPEFEQEEDDGEVTEDESDEDFIQPRRKRLKSDERPVIK-

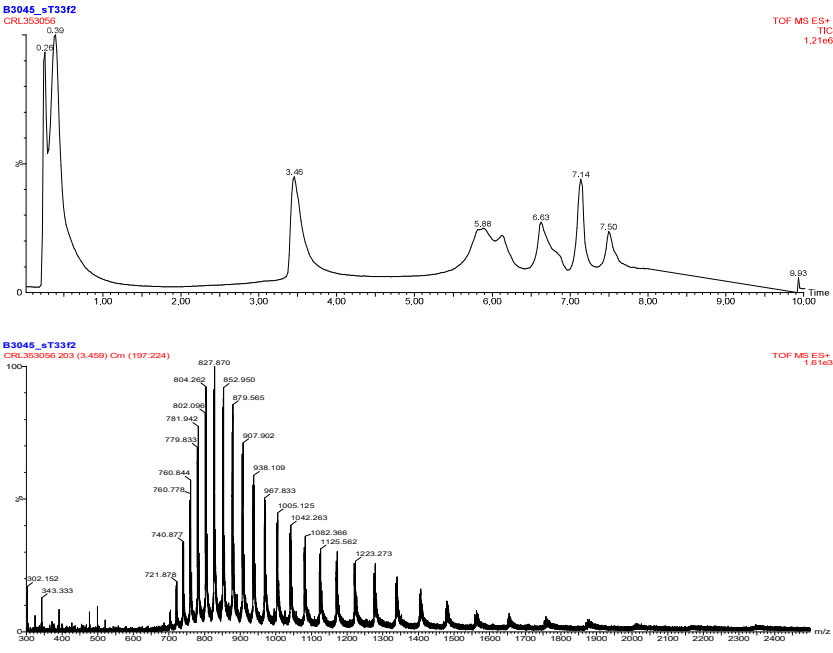
Calculated aa; MW (Da); ε (all cystines); pI	469; 54572.7; 69760; 5.58
HPLC trace of purification Gel of purification	 10 mL glutathione sepharose elute; 120 mL S75 gel filtration elute gst-TRIM33 
CD spectra	

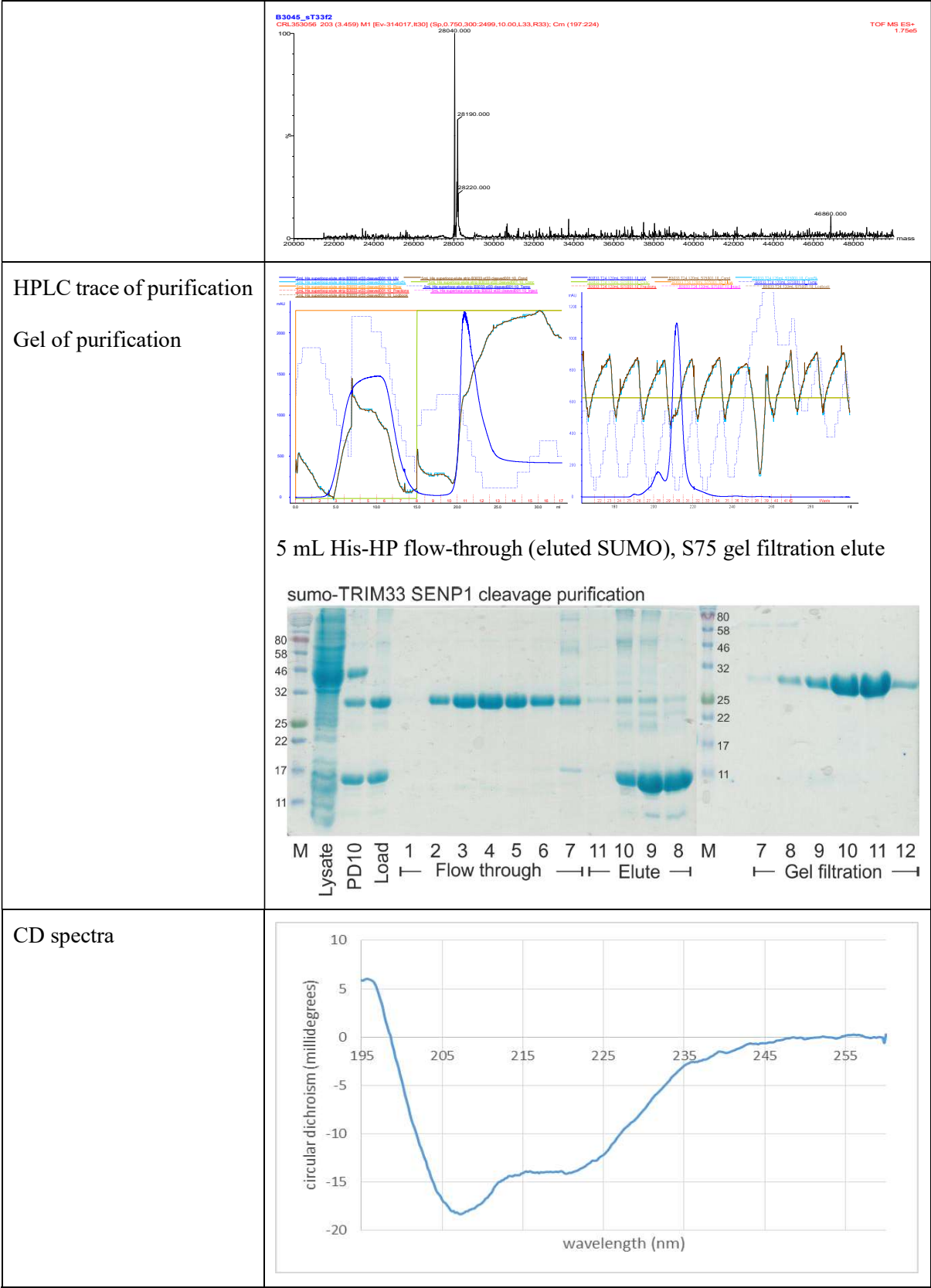
Gene	His₆-SUMO-TRIM33 WT
Vector backbone	pETM11-SUMO3 (Kanamycin)
Source	Gibson assembly
Affinity label : cleavage enzyme	His ₆ -SUMO : SENP1

DNA sequence	ATGAAACATCACCATCACCATCACCCCATGAGCGATTACGACATCCC CACTACTGAGAATCTTTATTTTCAGGGCGCCATGGGCAACGATCACA TTAACCTGAAAGTGGCCGGTCAAGACGGTAGCGTAGTCCAGTTTAAA ATCAAACGCCACACCCCTCTGTGAAACTGATGAAAGCCTATTGTGA ACGCCAAGGTCTGTCTATGCGTCAGATCCGTTTTTCGCTTCGATGGAC AGCCGATTAACGAAACCGACACTCCAGCACAGCTGGAAATGGAAGAT GAGGACACCATTGACGTGTTCCAGCAACAGACCGGTGGATCCATGGA CTGGTGTGCTGTCTGCCAAAACGGAGGAGATCTCTTGTGCTGCGAAA AATGTCCAAAGGTCTTTCATCTAACTTGTGTCATGTTCCAACACTACTT AGCTTTCCAAGTGGGGACTGGATATGCACATTTTGTAGAGATATTGG AAAGCCAGAAGTTGAATATGATTGTGATAATTTGCAACATAGTAAGA AGGGGAAAACCTGCGCAGGGGTAAAGCCCCGTGGACCAAAGGAAATGT GAACGTCTTCTGCTTTACCTCTATTGCCATGAATTAAGTATTGAATT CCAGGAGCCTGTTCCCTGCTTCGATACCAAACCTACTATAAAATTATAA AGAAACCAATGGATTTATCCACCGTGAAAAAGAAGCTTCAGAAAAAA CATTTCCCAACACTACCAAATCCCGGATGACTTTGTGGCCGATGTCCG TTTGATCTTCAAGAACTGTGAAAGGTTTAAATGAAATGATGAAAGTTG TTCAAGTTTATGCAGACACACAAGAGATTAATTTGAAGGCTGATTCA GAAGTAGCTCAGGCAGGGAAAGCAGTTGCATTGTACTTTGAAGATAA ACTCACAGAGATCTACTCAGACAGGACCTTCGCACCTTTGCCAGAGT TTGAGCAGGAAGAGGATGATGGTGAGGTAAGTCTGAGGACTCTGATGAA GACTTTATACAGCCCCGCAGAAAACGCCTAAAGTCAGATGAGAGACC AGTACATATAAAGTAAAAGCTTGCGGCCGCACTCGAGCACCACCACC ACCACCAC
Amino acid sequence	MKHHHHHHHPMSDYDIPTTENLYFQGAMGNDHINLKVAGQDGSVVQFK IKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMED EDTIDVFQQQTGGSMDWCAVCQNGGDLCCCEKCPKVFHLTCHVPTLL SFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKGKTAQGLSPVDQRKC ERLLLYLYCHELSIEFQEPVPASIPNYYKIIKKPMDLSTVKKKLQKK HSQHYQIPDDFVADVRLIFKNCERFNEMMKVVQVYADTQEINLKADS EVAQAGKAVALYFEDKLTEIYSDRTFAPLPEFEQEEDDGEVTEDSDE DFIQPRRKRLKSDERPVIK-
Calculated aa; MW (Da); ϵ (all cystines); pI	349; 40383.8; 29630; 5.61
Observed MS (Da)	Cleavage on column: 28040 and 28116 observed

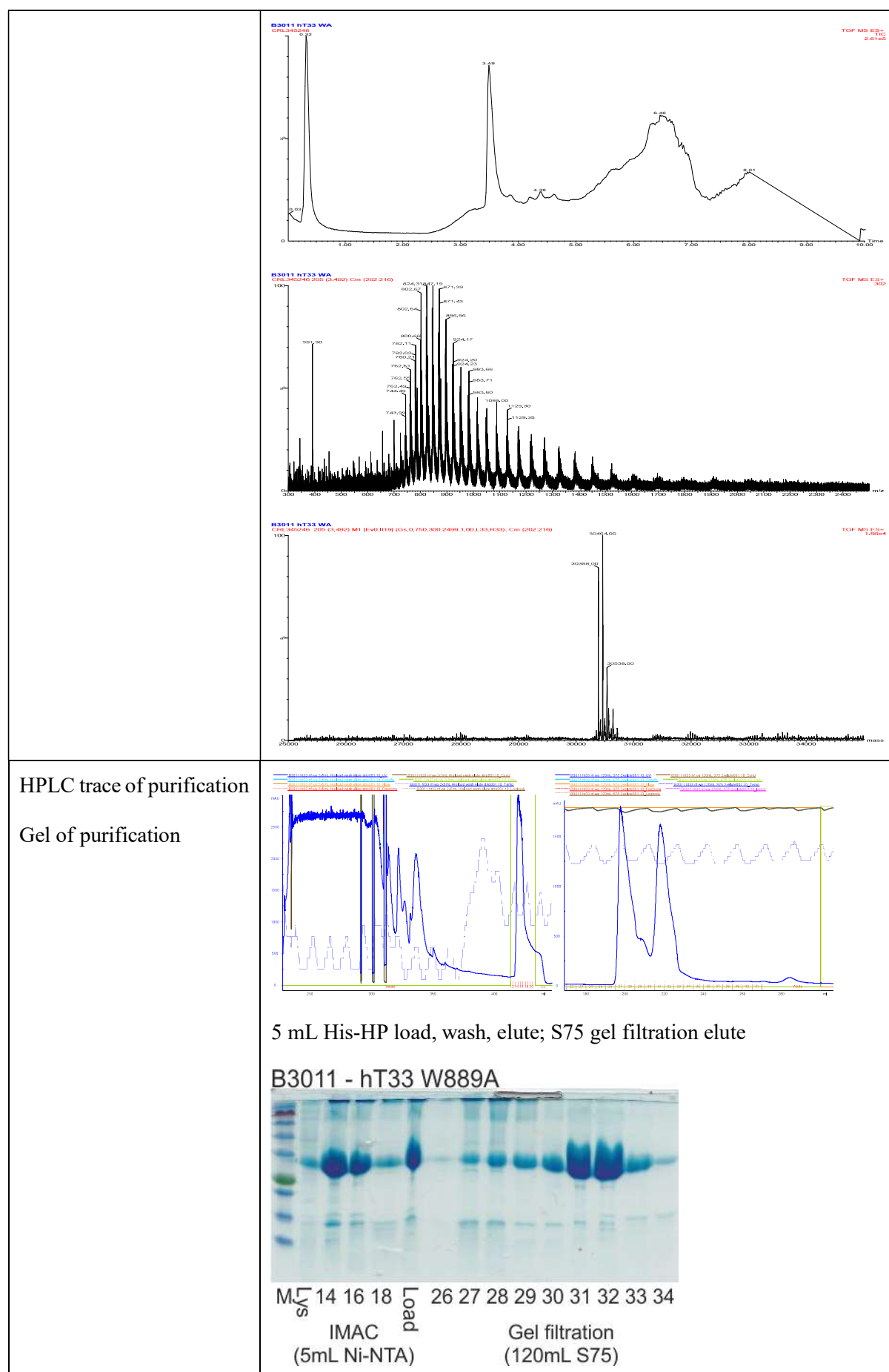


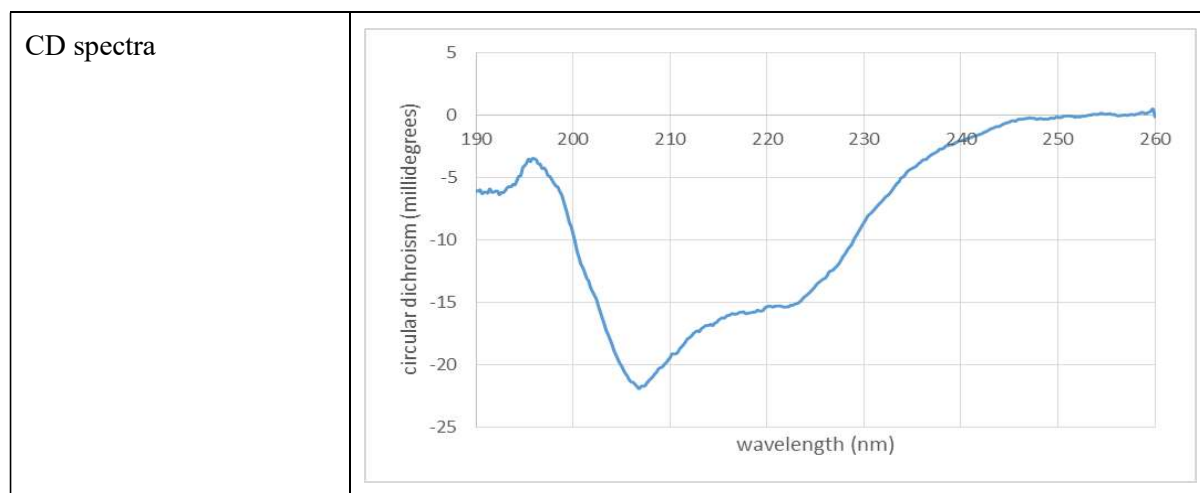


Gene	TRIM33 WT
Cleaved amino acid sequence	SMDWCAVCQNGDLLCCEKCPKVFHLTCHVPTLLSFPSGDWICTFCR DIGKPEVEYDCDNLQHSKKGKTAQGLSPVDQRKCERLLLYLYCHELS IEFQEPVPASIPNYYKIIKKPMDLSTVKKKLQKKHSQHYQIPDDFVA DVRLI FKN CERFNEMMKVVQVYADTQEINLKADSEVAQAGKAVALYF EDKLTEIYSDRTFAPLPEFEQEEDDGEVTEDSDEDFIQPRRKRLKSD ERPVIHK-
Calculated aa; MW (Da); E (all cystines); pI	242; 28035.9; 25160; 5.35
Observed MS (Da) (chromatogram, combined, deconvoluted)	Cleaved from pETM11-SUMO3: 28040 Da 



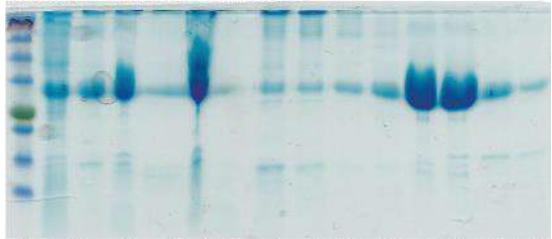
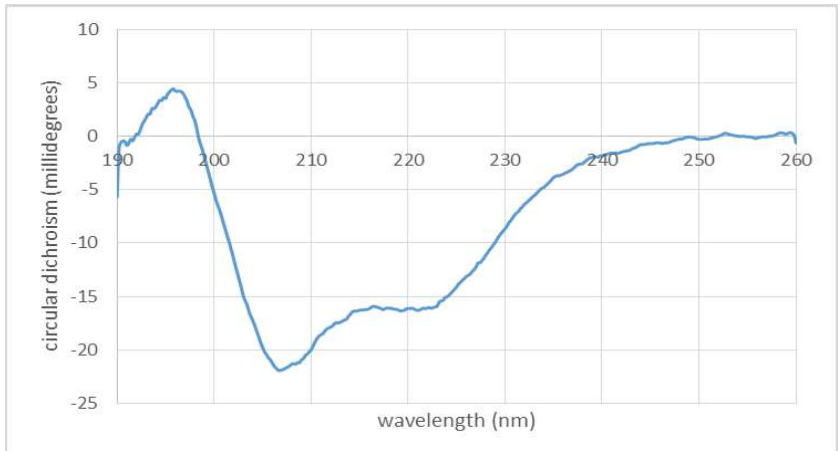
Gene	His₆-TRIM33 W889A
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from WT
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGGACGCGTGTGCTGTCTGCCAAAACG GAGGAGATCTCTTGTGCTGCGAAAAATGTCCAAAGGTCTTTCATCTA ACTTGTCATGTTCCAACACTACTTAGCTTTCCAAGTGGGGACTGGAT ATGCACATTTTGTAGAGATATTGGAAAGCCAGAAGTTGAATATGATT GTGATAATTTGCAACATAGTAAGAAGGGGAAAAC TGCGCAGGGGTTA AGCCCCGTGGACCAAAGGAAATGTGAACGTCTTCTGCTTTACCTCTA TTGCCATGAATTAAGTATTGAATTCCAGGAGCCTGTTCTGCTTCGA TACCAACTACTATAAAATTATAAAGAAACCAATGGATTTATCCACC GTGAAAAAGAAGCTTCAGAAAAAACATTCCCAACACTACCAAATCCC GGATGACTTTGTGGCCGATGTCCGTTTGATCTTCAAGAACTGTGAAA GGTTTAATGAAATGATGAAAGTTGTTCAAGTTTATGCAGACACACAA GAGATTAATTTGAAGGCTGATTCAGAAGTAGCTCAGGCAGGGAAAGC AGTTGCATTGTACTTTGAAGATAAACTCACAGAGATCTACTCAGACA GGACCTTCGCACCTTTGCCAGAGTTTGAGCAGGAAGAGGATGATGGT GAGGTAAC TGAGGACTCTGATGAAGACTTTATACAGCCCCGCAGAAA ACGCCTAAAGTCAGATGAGAGACCAGTACATATAAAGTAA
Amino acid sequence	MHHHHHSSGVDLGTENLYFQSM DACAVCQNGDLLCCEKCPKVFHL TCHVPTLLSFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKGKTAQGL SPVDQRKCERLLLYLYCHELSIEFQEPVPASIPNYYKIIKKPMDLST VKKKLQKKHSQHYQIPDDFVADVRLIFKNCERFNEMMKVVQVYADTQ EINLKADSEVAQAGKAVALYFEDKLTEIYSDRTFAPLPEFEQEEDDG EVTEDSDEDFIQPRRKRLKSDERP VHIK-
Calculated aa; MW (Da); ε (all cystines); pI	263; 30386.4; 21150; 5.62
Observed MS (Da) (chromatogram, combined, deconvoluted)	30388 and 30464





Gene	His₆-TRIM33 N980F
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from WT
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGGACTGGTGTGCTGTCTGCCAAAACG GAGGAGATCTCTTGTGCTGCGAAAAATGTCCAAAGGTCTTTCATCTA ACTTGTCATGTTCCAACACTACTTAGCTTTCCTCAAGTGGGGACTGGAT ATGCACATTTTGTAGAGATATTGGAAAGCCAGAAGTTGAATATGATT GTGATAATTTGCAACATAGTAAGAAGGGGAAAACCTGCGCAGGGGTTA AGCCCCGTGGACCAAAGGAAATGTGAACGTCTTCTGCTTTACCTCTA TTGCCATGAATTAAGTATTGAATTCCAGGAGCCTGTTCTGCTTCGA TACCAACTACTATAAAATTATAAAGAAACCAATGGATTTATCCACC GTGAAAAAGAAGCTTCAGAAAAACATTCCCAACACTACCAATCCC GGATGACTTTGTGGCCGATGTCCGTTTGATCTTCAAGAACTGTGAAA GGTTTTTTGAAATGATGAAAGTTGTTCAAGTTTATGCAGACACACAA GAGATTAATTTGAAGGCTGATTGAGAAGTAGCTCAGGCAGGGAAAGC AGTTGCATTGTACTTTGAAGATAAACTCACAGAGATCTACTCAGACA GGACCTTCGCACCTTTGCCAGAGTTTGAGCAGGAAGAGGATGATGGT GAGGTAAGTGAAGACTCTGATGAAGACTTTATACAGCCCCGCAGAAA ACGCCTAAAGTCAGATGAGAGACCAGTACATATAAAGTAA

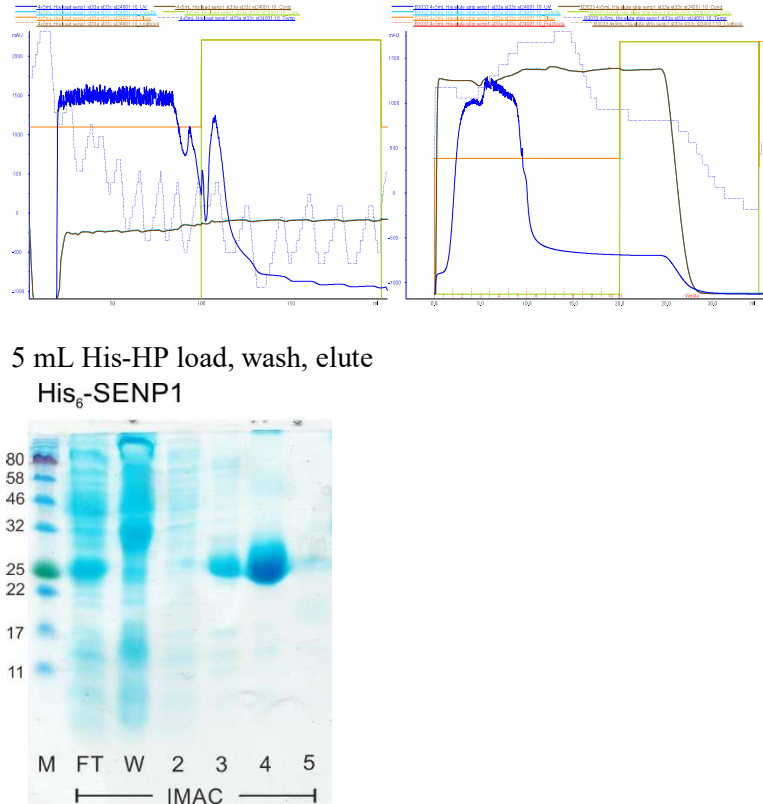
Amino acid sequence	MHHHHHSSGVDLG TENLYFQ SMDWCAVCQNGDLLCCEKCPKV FHL TCHVPTLLSFP SGDWICTFCRDIGKPEVEYDCDNLQHSKKGKTAQGL SPVDQRKCERLLL YLYCHELSIEFQEPVPASIPNYYKIIKKPMDLST VKKKLQKKHSQHYQIPDDFVADVRLIFKNCERFFEMMKVVQVYADTQ EINLKADSEVAQAGKAVALYFEDKLTEIYSDRTFAPLPEFEQEEDDG EVTEDSDEDFIQPRRKRLKSDERPVHIK-
Calculated aa; MW (Da); ε (all cystines); pI	263; 30534.7; 26650; 5.62
Observed MS (Da) (chromatogram, combined, deconvoluted)	30537 Da

	5 mL His-HP load, wash, elute; S75 gel filtration elute B3011 - hT33 N1039F  M 4 6 8 Load 5 6 7 8 9 10 11 12 13 IMAC (5mL Ni-NTA) Gel filtration (120mL S75)
CD spectra	

Gene	His₆-TRIM33 W889A N980F
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Site-directed mutagenesis from single mutant
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGCACCATCATCATCATCTTCTTCTGGTGTAGATCTGGGTACCGA GAACCTGTACTTCCAATCCATGGACGCGTGTGCTGTCTGCCAAAACG GAGGAGATCTCTTGTGCTGCGAAAAATGTCCAAAGGTCTTTCATCTA ACTTGTCATGTTCCAACACTACTTAGCTTTCCAAGTGGGGACTGGAT ATGCACATTTTGTAGAGATATTGGAAAGCCAGAAGTTGAATATGATT GTGATAATTTGCAACATAGTAAGAAGGGGAAAAGTGCAGGGGGTTA AGCCCCGTGGACCAAGGAAATGTGAACGTCTTCTGCTTTACCTCTA TTGCCATGAATTAAGTATTGAATTCCAGGAGCCTGTTTCTGCTTCGA TACCAAACTACTATAAAATTATAAAGAAACCAATGGATTTATCCACC

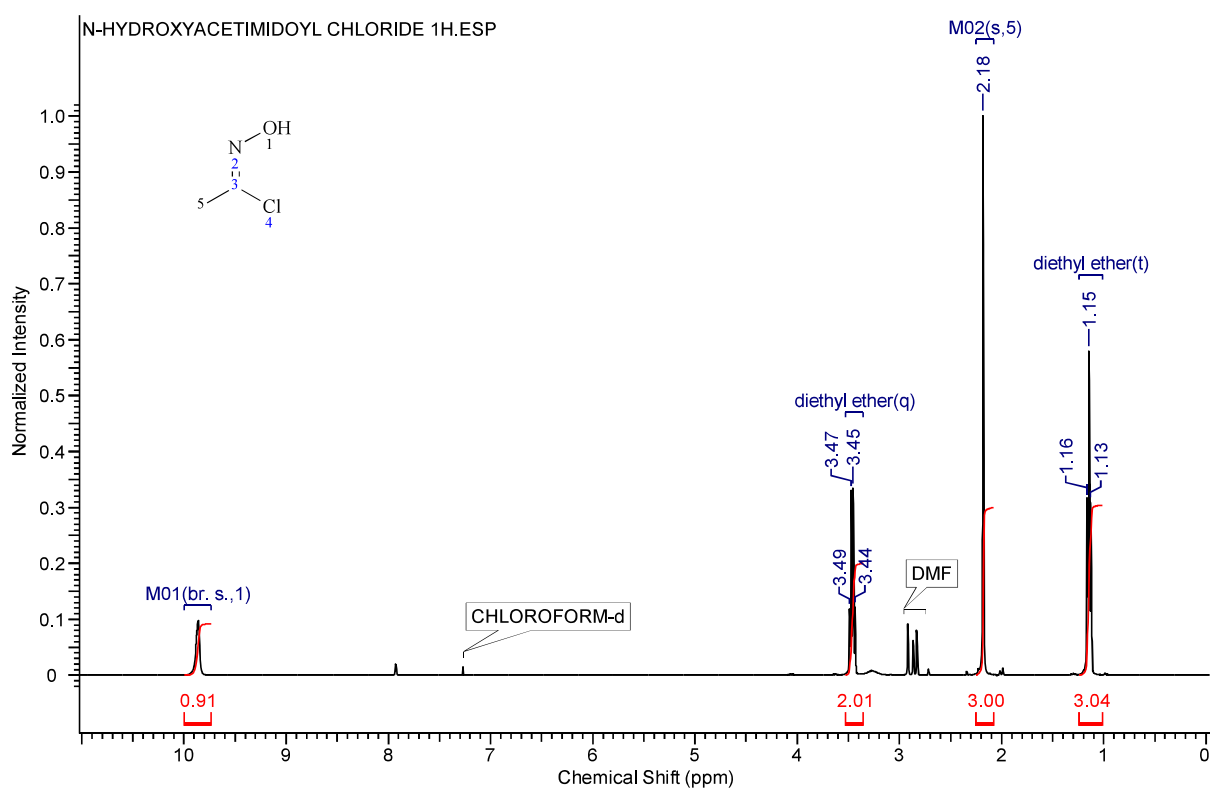
	GTGAAAAGAAGCTTCAGAAAAACATTCCCAACACTACCAAATCCC GGATGACTTTGTGGCCGATGTCCGTTTGATCTTCAAGAACTGTGAAA GGTTTTTTGAAATGATGAAAGTTGTTCAAGTTTATGCAGACACACAA GAGATTAATTTGAAGGCTGATTCAGAAAGTAGCTCAGGCAGGGAAAGC AGTTGCATTGTACTTTGAAGATAAACTCACAGAGATCTACTCAGACA GGACCTTCGCACCTTTGCCAGAGTTTGAGCAGGAAGAGGATGATGGT GAGGTAAGTGAAGACTCTGATGAAGACTTTATACAGCCCCGCAGAAA ACGCCTAAAGTCAGATGAGAGACCAGTACATATAAAGTAA
Amino acid sequence	MHHHHHSSGVDLGTENLYFQSM DACAVCQNGDLLCCEKCPKVFHL TCHVPTLLSFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKGKTAQGL SPVDQRK CERLLLYLYCHELSIEFQEPVPASIPNYYKIIKKPMDLST VKKKLQKKHSQHYQIPDDFVADVRLIFKNCERFFEMMKVVQVYADTQ EINLKADSEVAQAGKAVALYFEDKLTEIYSDRTFAPLPEFEQEEDDG EVTEDSDEDFIQPRRKRLKSDERP VHIK-
Calculated aa; MW (Da); ε (all cystines); pI	263; 30419.5; 5.62; 21150

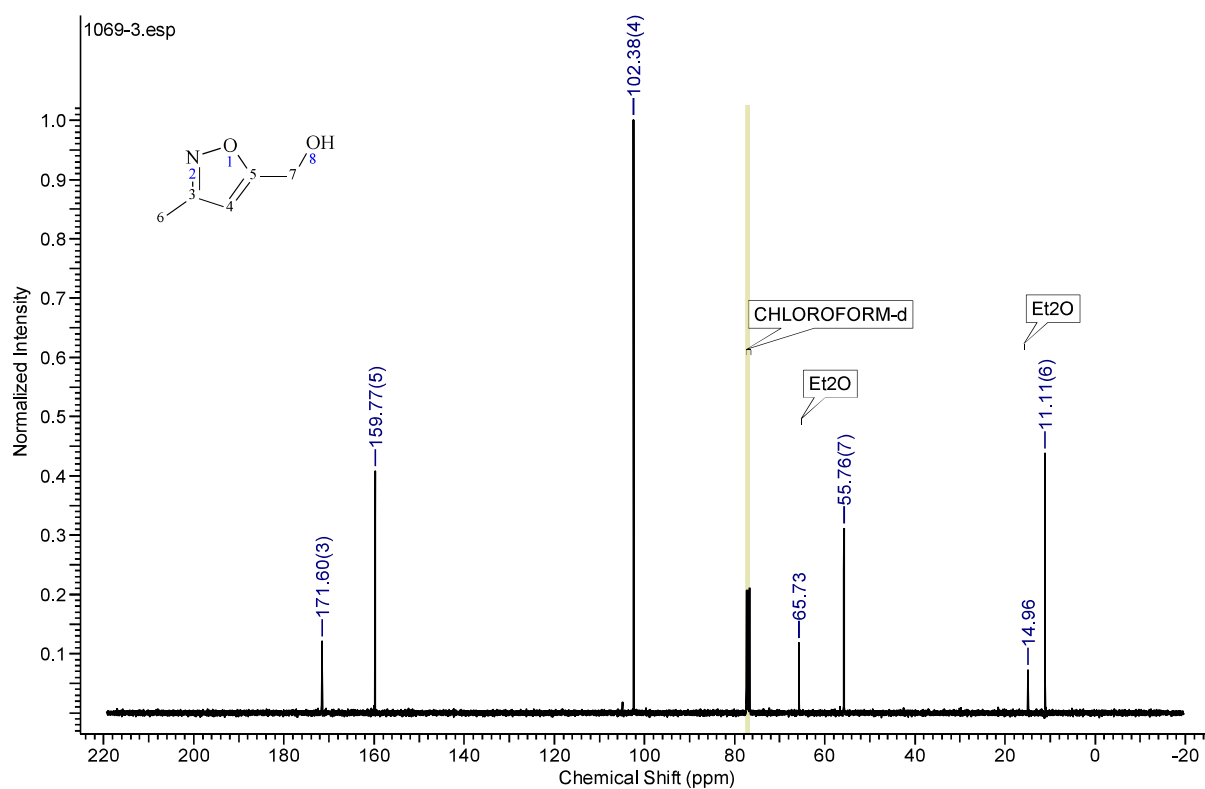
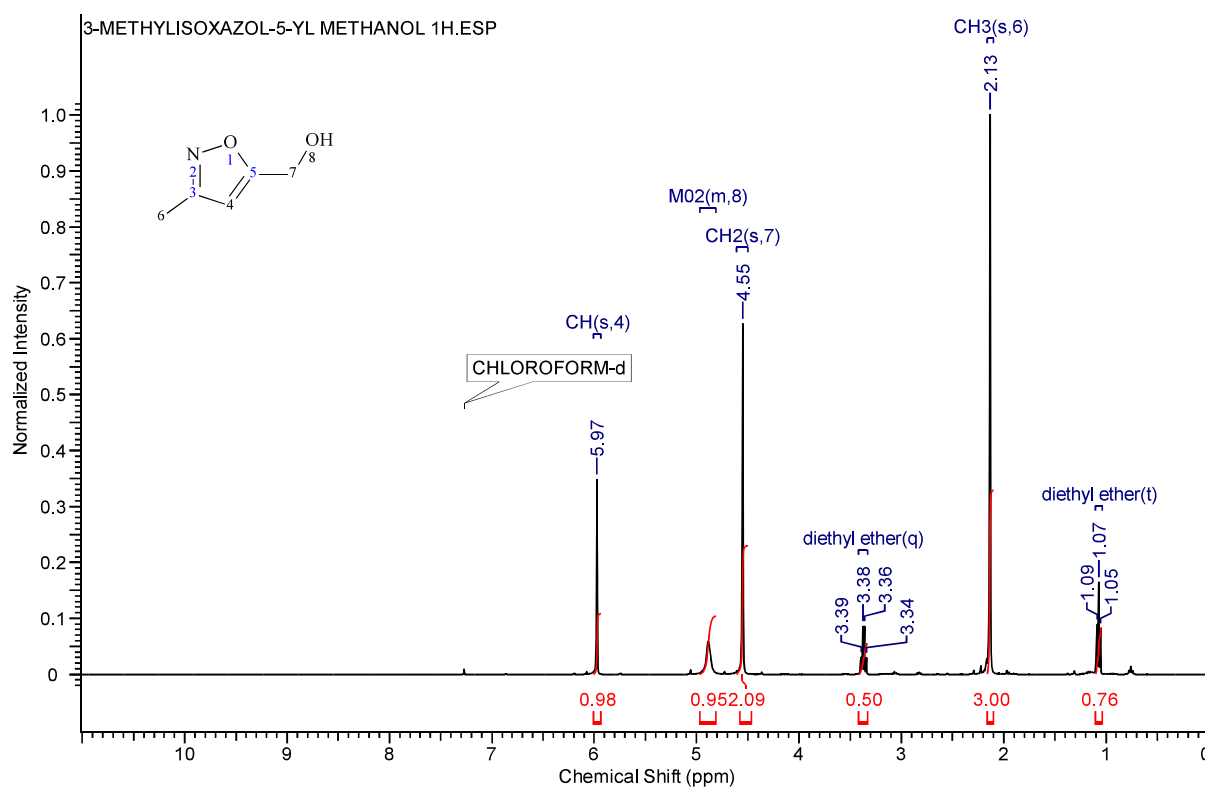
Gene	His₆-SEN1
Vector backbone	pNIC28-Bsa4 (Kanamycin)
Source	Provided by SGC
Affinity label : cleavage enzyme	His ₆ : TEV protease
DNA sequence	ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCC GCGCGGCAGCCATATGGAATTTCTGAAATTACAGAGGAAATGGAGA AAGAAATAAAGAATGTATTTTCGTAATGGGAATCAGGATGAAGTTCTC AGTGAAGCATTTTCGCCTGACCATTACACGCAAAGATATTCAAACCTCT AAACCATCTGAATTGGCTCAATGATGAGATCATCAATTTCTACATGA ATATGCTGATGGAGCGAAGTAAAGAGAAGGGCTTGCCAAGTGTGCAT GCATTTAATACCTTTTTCTTCACTAAATTAAAAACGGCTGGTTATCA GGCAGTGAAACGTTGGACAAAGAAAGTAGATGTATTTTCTGTTGACA TTCTTTTGGTGCCCATTCACCTGGGAGTACACTGGTGTCTAGCTGTT GTGGACTTTAGAAAGAAGAATATTACCTATTACGACTCCATGGGTGG GATAACAATGAAGCCTGCAGAATACTCTTGCAATACCTAAAGCAAG AAAGCATTGACAAGAAAAGGAAAGAGTTTGACACCAATGGCTGGCAG

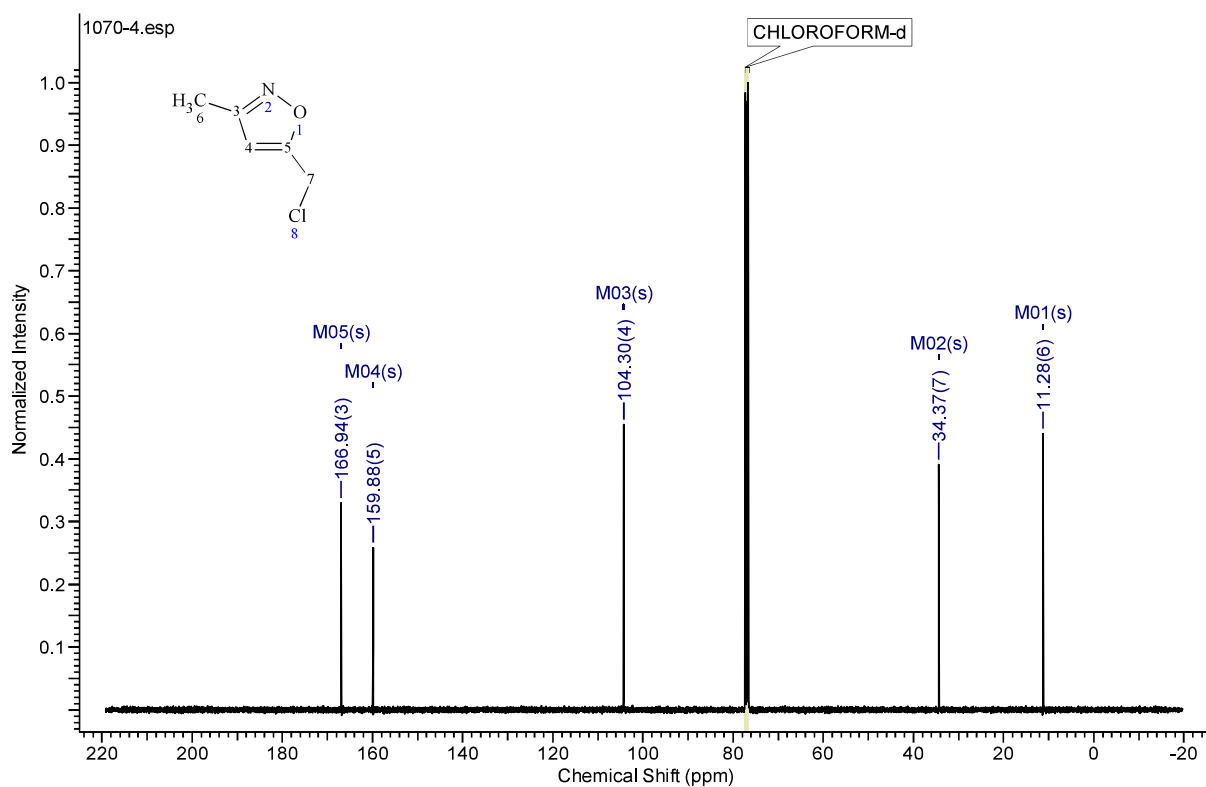
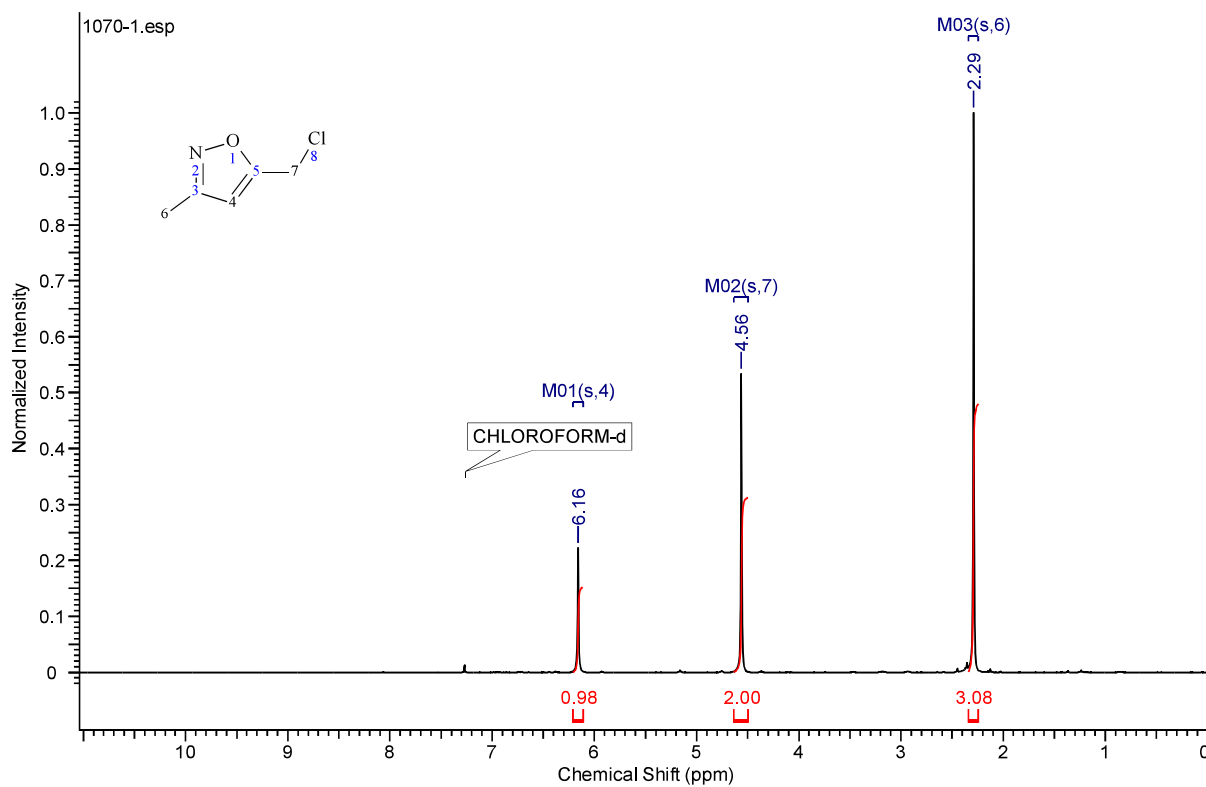
	CTTTTCAGCAAGAAAAGCCAGGAGATTCCTCAGCAGATGAATGGAAG TGACTGTGGGATGTTTGCCTGCAAATATGCTGACTGTATTACCAAAG ACAGACCAATCAACTTCACACAGCAACACATGCCATACTTCCGGAAG CGGATGGTCTGGGAGATCCTCCACCGAAAACTCTTGTGA
Amino acid sequence	MGSSHHHHHHSSGLVPRGSHMEFPEITEEMEKEIKNVFRNGNQDEVL SEAFRLTITRKDIQTLNHLNWLNDEIINFYMNMLMERSKEKGLPSVH AFNTFFFTKLKTAGYQAVKRWTKKVDVFSVDILLVPIHLGVHWCLAV VDFRKKNITYYDSMGGINNEACRILLQYLKQESIDKKRKEFDTNGWQ LFSKKSQEI PQQMNGSDCGMFACKYADCITKDRPINFTQQHMPYFRK RMVWEILHRKLL-
Calculated aa; MW (Da); ϵ (all cystines); pI	247; 29186.6; 38180; 9.10
HPLC trace of purification Gel of purification	 The figure displays the purification of His₆-SENP1. The top section shows two HPLC traces: a blue trace for absorbance at 280 nm and a black trace for refractive index. The traces show a major peak at approximately 15 minutes, indicating successful purification. The bottom section is an SDS-PAGE gel with lanes labeled M, FT, W, 2, 3, 4, and 5. Molecular weight markers are indicated on the left (80, 58, 46, 32, 25, 22, 17, 11 kDa). Lane M contains the marker. Lanes FT, W, 2, 3, 4, and 5 show the protein bands. A prominent band is visible in lane 4, corresponding to the expected molecular weight of His₆-SENP1. The gel is labeled 'IMAC' at the bottom, indicating ion-exchange chromatography. 5 mL His-HP load, wash, elute His₆-SENP1

8.4 Compound NMR spectra

N-Hydroxyacetimidoyl chloride (23)



(3-Methylisoxazol-5-yl)methanol (24)

5-(Chloromethyl)-3-methylisoxazole (25: Cl-DMI)

8.5 AlphaScreen: biotinylated peptide signal response

Table 8.1 AlphaScreen emission signal (a.u. fluorescence) for TRIM24 with varyingly modified peptides in two-fold dilution series. Colour scale from low signal (red) to high signal (green). FAC – final assay concentration. Where multiple rows have the same FAC protein, these may use protein from independent expression and purifications. Assay parameters used in screening are highlighted with a bold outline.

Peptide	FAC protein (nM)	FAC peptide (nM)											
		0.025	0.244	0.488	0.977	1.95	3.91	7.81	15.6	31.3	62.5	125	250
B1: K9Me3	10	193	181	215	247	259	372	423	397	391	377	364	275
	10	203	195	203	285	324	395	606	662	504	528	522	427
	10	247	299	324	364	598	852	1205	1197	1240	1201	1234	708
	30	136	192	469	1361	2739	2829	3407	3697	3879	3426	2006	840
	30	355	346	373	1337	4189	9674	13067	13556	14267	14393	13767	8839
	50	187	248	770	4061	9444	19881	26300	26305	28107	26946	21899	10613
	50	329	365	913	8655	26042	42868	48862	48416	46849	49693	49986	37922
	90	395	7264	17982	30984	47347	53532	56837	55982	54965	59037	48676	34408
	250	3516	33418	57576	82866	115264	142811	142481	141060	132660	131037	126014	107497
	270	4663	38742	53918	73065	87222	91536	90857	87251	84840	82622	81788	70851
B2: K18Ac	10	281	669	1318	4380	15265	38498	60511	54207	53443	42766	36135	19371
	10	3352	70380	102026	133748	162287	186960	199667	200236	180712	176096	172273	152382
	10	146	1001	2764	9995	26301	41928	50016	45139	40744	33723	24452	9286
	30	304	4852	10375	22366	44313	72778	95144	87883	79610	73643	62998	49482
	30	148	8118	22155	43154	67627	84521	88568	80740	76976	68118	53125	33928
	50	275	15643	23712	38399	73390	95608	113964	102004	99031	89102	83611	57700
	50	193	11287	25940	49612	88905	126728	158537	161065	144602	135108	117589	85098
	90	419	36852	56219	85885	109515	127484	130319	118765	113405	107687	96940	79589
	250	187	468	982	3885	17875	34947	62976	65450	58386	46108	31637	13536
	270	5436	77081	96464	109143	128969	130954	130919	130589	120848	117066	116434	102626
	10	13219	92269	124822	153209	175749	188456	186824	183715	165613	163524	148827	111615
	10	161	626	3904	11936	24437	29855	28199	21984	18616	9933	4095	822
	30	159	7862	22538	45551	70751	84730	73674	67817	58272	48047	30467	13496
	50	256	8810	26813	61357	99656	138835	156832	144288	113134	101718	74112	43763

Peptide	FAC protein (nM)	FAC peptide (nM)											
		0.025	0.244	0.488	0.977	1.95	3.91	7.81	15.6	31.3	62.5	125	250
	90	370	38356	63635	87301	112665	115070	110448	100687	90168	81291	61261	41277
	250	158	190	286	3333	16351	31161	36958	37515	27860	16239	9979	1420
	270	4635	87208	105206	117754	128840	125659	120591	119728	109838	107269	94327	72557
	10	283	2746	7568	21783	47745	67838	79468	61006	54050	39362	26643	13605
B4: K9Me3K18Ac	10	155	9945	20932	40079	60868	73199	65838	54502	41642	26360	14380	4969
	30	307	16491	34481	50978	76672	101376	110448	99720	85530	71831	54835	35328
	30	178	35295	52679	78895	100793	116671	108497	97777	81454	65077	39231	20257
	50	313	35558	59604	79339	102369	115867	126323	106586	103176	87506	74513	47485
B4: K9Me3K23Ac	90	418	59885	89316	110296	133542	137139	130667	121890	112989	106023	79946	51665
	270	3793	89427	109622	128802	140778	139862	140587	136321	121591	120899	112725	96054
	10	4027	106749	138547	171374	186542	195264	189605	186511	165554	163474	141856	107890
	10	112	1351	6978	20174	35858	44132	36844	31478	22336	13585	5610	1434
B5: K9Me3K23Ac	30	140	15949	35466	57806	14513	92594	79435	67590	59656	45459	27704	10477
	50	185	20416	41596	83758	120009	144162	151617	136817	115563	99878	68524	36492
	90	721	48921	80001	99469	120832	118364	112015	100491	94794	84711	61192	38498
	250	149	317	1114	9056	28302	40463	53920	40243	28225	17313	6752	1983
B5: K9Me3K14Ac	270	2980	94417	115002	123571	132013	126834	124459	117987	110640	105792	93505	69594
	10	3109	6246	10917	24778	64258	117379	171465	187757	167588	168398	163659	155687
	10	135	124	162	156	208	236	228	213	261	218	234	187
	30	143	144	134	271	543	1857	2100	1365	1275	1285	1100	1263
B6: (5-24)K14Ac	50	170	236	290	411	2073	7519	17825	16033	18152	17597	15051	15547
	90	430	1987	4934	16728	38567	64335	73498	63896	64659	67075	64532	64374
	250	173	163	173	178	182	273	344	335	310	301	320	276
	270	1835	12822	29900	52949	88315	117808	122082	115729	110745	111211	108167	109121
B7: K14Ac	10	151	163	170	361	865	1737	718	5306	4717	4939	4254	3912
	30	107	136	315	2720	15859	33375	51096	53318	55920	53334	50041	44628
	90	430	6269	13855	28518	55571	80071	104503	102881	102250	99952	96431	93339
	270	1088	29291	45799	66734	91385	112216	125357	122104	117087	117254	115247	111071

Table 8.2 AlphaScreen emission signal (a.u. fluorescence) for TRIM33 with varyingly modified peptides in two-fold dilution series. Colour scale from low signal (red) to high signal (green). FAC – final assay concentration. Where multiple rows have the same FAC protein, these may use protein from independent expression and purifications. Assay parameters used in screening are highlighted with a bold outline

Peptide	FAC protein (nM)	FAC peptide (nM)											
		0.025	0.244	0.488	0.977	1.95	3.91	7.81	15.6	31.3	62.5	125	250
B1: K9Me3	50	145	227	279	289	330	378	441	490	493	521	493	250
	50	238	241	265	323	362	438	453	510	505	547	466	475
	50	423	443	487	519	664	768	1140	2120	1772	1705	1658	414
	50	412	552	684	799	1022	1144	1639	3160	2576	3012	2736	1547
	100	132	228	257	373	317	393	544	745	620	620	570	1741
	200	190	335	372	530	622	695	1586	2670	2199	2188	2066	457
	250	229	209	258	513	520	682	1047	1450	1562	1514	1520	1470
	250	486	566	629	1108	2321	4746	14658	31994	34401	46872	43418	1262
	250	328	684	952	1232	2255	5280	22671	46003	49071	57504	58680	43802
	400	173	397	986	793	1119	1862	6079	12703	12360	12657	10708	45413
	500	259	331	608	1015	1528	2620	6200	8412	8321	9932	8724	8883
	500	372	817	1626	5687	12487	26921	65865	81484	90078	99061	109545	23348
B2: K18Ac	500	399	635	1693	1808	7302	25902	71192	93866	91570	97778	96466	120363
	50	439	376	436	459	592	947	1903	3613	5827	6978	7645	128464
	50	202	220	177	215	268	349	401	552	534	696	679	8236
	50	161	205	210	231	326	356	486	605	695	783	812	693
	100	158	188	254	387	619	746	1074	1962	2867	3295	3331	875
	200	200	399	1098	2506	3601	5655	10597	18369	24864	31174	31656	3953
	250	579	526	561	623	6980	10641	47592	72466	88180	121674	115705	34131
	250	204	207	324	882	3233	6696	11949	17554	25625	28847	31321	98161
	400	147	2125	4555	10181	18969	30766	46279	67553	72131	80719	78366	45316
	500	456	826	2891	8153	28950	63801	98253	136812	142757	133744	141478	81165
	500	259	516	2018	6421	14679	34344	50821	67777	67745	72704	74032	131523
	50	198	202	208	213	256	279	372	435	363	386	385	73609
B3: K23Ac	50	165	200	255	280	324	372	417	431	416	482	427	352
	100	152	237	277	356	364	444	526	525	542	547	613	403
	200	192	338	366	490	569	681	825	1167	922	1241	945	460
	200												879

Peptide	FAC protein (nM)	FAC peptide (nM)											
		0.025	0.244	0.488	0.977	1.95	3.91	7.81	15.6	31.3	62.5	125	250
	250	220	199	258	287	469	463	526	539	595	623	601	584
	400	167	305	473	564	678	859	1631	2179	1855	1517	1801	1653
	500	304	220	303	421	567	697	984	1282	1128	1233	1231	1328
	50	373	599	515	917	6802	31545	52140	79255	84342	84246	84813	74087
B4: K9Me3K18Ac	50	177	363	829	3429	8115	12950	18835	22979	22169	22231	21634	17525
	100	148	1588	5470	15896	34585	60602	82604	82230	78669	79123	73642	66277
	200	175	6665	18318	39736	72922	110640	141361	153478	142216	147502	136976	129681
	250	524	1371	5614	23207	59038	107880	143885	158316	156008	153633	138926	141003
	400	1436	19468	35123	64118	105766	138387	162335	156233	149761	149652	143211	138989
	500	573	5348	21572	46959	93639	130383	162236	173964	172749	158273	166931	152346
	50	169	245	212	216	284	288	379	389	438	461	461	465
	50	2556	225	262	220	271	365	404	599	503	577	545	509
B5: K9Me3K23Ac	100	127	244	289	359	370	459	622	993	1064	984	1001	1066
	200	181	419	590	839	1113	1489	3576	6490	6310	6627	6031	6652
	250	195	233	248	789	1450	1742	2649	3576	3008	3868	3691	6692
	400	155	1311	2487	4343	7455	11103	15140	20127	18417	20672	20284	20432
	500	294	703	2315	5308	7816	14114	21528	20769	23687	25678	24663	20182
	50	183	182	182	201	212	240	260	270	257	257	213	257
	50	186	146	206	222	229	289	304	283	273	335	323	283
	100	144	168	153	214	250	280	249	329	293	293	280	305
B6: (5-24)K14Ac	200	168	142	188	250	280	345	354	465	399	357	383	315
	250	189	174	135	198	227	338	344	354	325	399	374	370
	400	154	149	163	231	291	333	372	371	319	430	380	311
	500	263	121	145	184	278	297	327	375	390	386	405	370
B7: K14Ac	50	147	162	192	213	248	283	329	362	383	399	407	396
	100	108	205	180	236	287	301	332	447	504	422	502	421
	200	152	172	274	320	345	415	475	794	1085	984	879	875
	400	138	224	274	341	398	454	532	1164	2223	1785	1625	1915

8.6 References

(1) Xi, Q., Wang, Z., Zaromytidou, A.-I., Zhang, X. H.-F.-F., Chow-Tsang, L.-F., Liu, J. X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., Studer, L., Mark, W., Patel, D. J., Massagué, J., and Massague, J. (2011) A Poised Chromatin Platform for TGF-beta Access to Master Regulators. *Cell* 147, 1511–1524.