Engineering and Delivery of Synthetic Chromatin Effectors

by

Stefan Josef Tekel

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2019 by the
Graduate Supervisory Committee

Karmella Haynes, Chair
Michael Caplan
Jeremy Mills
David Brafman

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Synthetic manipulation of chromatin dynamics has applications for medicine, agriculture, and biotechnology. However, progress in this area requires the identification of design rules for engineering chromatin systems. In this thesis, I discuss research that has elucidated the intrinsic properties of histone binding proteins (HBP), and apply this knowledge to engineer novel chromatin binding effectors. Results from the experiments described herein demonstrate that the histone binding domain from chromobox protein homolog 8 (CBX8) is portable and can be customized to alter its endogenous function. First, I developed an assay to identify engineered fusion proteins that bind histone post translational modifications (PTMs) *in vitro* and regulate genes near the same histone PTMs in living cells. This assay will be useful for assaying the function of synthetic histone PTM-binding actuators and probes. Next, I investigated the activity of a novel, dual histone PTM binding domain regulator called Pc_2TF. I characterized Pc_2TF *in vitro* and in cells and show it has enhanced binding and transcriptional activation compared to a single binding domain fusion called Polycomb Transcription Factor (PcTF). These results indicate that valency can be used to tune the activity of synthetic histone-binding transcriptional regulators. Then, I report the delivery of PcTF fused to a cell penetrating peptide (CPP) TAT, called CP-PcTF. I treated 2D U-2 OS bone cancer cells with CP-PcTF, followed by RNA sequencing to identify genes regulated by CP-PcTF. I also showed that 3D spheroids treated with CP-PcTF show delayed growth. This preliminary work demonstrated that an epigenetic effector fused to a CPP can enable entry and regulation of genes in U-2 OS cells through DNA independent interactions. Finally, I described and validated a new screening method that combines the versatility of *in vitro*

transcription and translation (IVTT) expressed protein coupled with the histone tail microarrays. Using Pc_2TF as an example, I demonstrated that this assay is capable of determining binding and specificity of a synthetic HBP. I conclude by outlining future work toward engineering HBPs using techniques such as directed evolution and rational design. In conclusion, this work outlines a foundation to engineer and deliver synthetic chromatin effectors.

DEDICATION

This work is dedicated to my parents, who have always been there when I needed them, and to my cats Selex and Mica. I also dedicate this work to the Washington Capitals, who have proven that we can all overcome adversity and finish on top. We did it boys!

ACKNOWLEDGMENTS

I would like to thank Dr. Debbie Mohler for her support and inspiration for me going to graduate school.

I would like to thank Drs. Matthew Dunn and Andrew Larsen for teaching me how to be successful in molecular biology, graduate school, and most importantly, Starcraft 2.

I would also like to thank Lusheng Song for helping me develop my screening platform and with advice and support during project development.

I would like to thank all of the other graduate students that I have collaborated with on projects: Daniel Vargas, Nick Brookhouser, and Kylie Standage-Beier. I would not have been as successful without all of your help.

I would like to acknowledge Antonio Garcia for support throughout my tenure in the Biological Design Graduate Program, and for always finding funding when needed.

I would like to thank my undergraduate and graduate students for helping me become a better mentor.

Finally, I would like to thank my committee for academic and financial support throughout my PhD, and for the opportunities they have provided that will enable me to be successful in my future endeavors.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

PREFACE

Chapters 1, 2, and 3 presented in this thesis have been previously published as described:

Chapter 1: **Tekel, S. J.**, and Haynes, K. A. (2017) Molecular Structures Guide the Engineering of Chromatin. Nucleic Acids Research. 45 (13), 7555−7570.

Chapter 2: **Tekel, S. J.**, Barrett, C., Vargas, D. and Haynes, K. A. (2018) Design, Construction, and Validation of Histone-Binding Effectors in Vitro and in Cells. Biochemistry, **57**, 4707–4716.

Chapter 3: **Tekel, S. J.**, Vargas, D. A., Song, L., LaBaer, J., Caplan, M. R., and Haynes, K. A. (2018) Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. ACS Synthetic Biology. 7 (3), 842−852.

CHAPTER 1

MOLECULAR STRUCTURES GUIDE THE ENGINEERING OF CHROMATIN

Chromatin Engineering: An Important and Challenging Undertaking

Chromatin is a dynamic nuclear structure that has a central role in eukaryotic development. The mechanics of this ancient, highly conserved system (1-2) are primarily driven by the physical structure and interactions of its components, proteins and nucleic acids. Electrostatic bonds and hydrophobic interactions determine the composition of multipart subunits such as nucleosomes, transcription initiation complexes, and repressive complexes. Because of its impact on tissue development, chromatin has great potential for engineering cell populations. Chromatin proteins exert strong and flexible control over cohorts of genes that determine cell fate and tissue organization. Chromatin states, *i.e.* actively transcribed and silenced, can switch from one to the other. At the same time chromatin-mediated regulation can be very stable, persisting over many cycles of DNA replication and mitosis. The latter property is a mode of epigenetic inheritance, where cellular information that is not encoded in the DNA sequence is passed from mother to daughter cells. The stability of chromatin states allows specific epigenetic programs to scale with tissue development in multicellular organisms.

Early biochemical and protein structure studies of the nucleosome (3) have generated a high resolution model that has persisted over time. A single nucleosome includes a complex of eight histone proteins arranged in a spiral-like disc. Each histone contains a C-terminal globular region composed of helix-turn-helix motifs called the histone fold domain, and an unfolded N-terminal tail (4) (Figure 1.1A). Within each nucleosome, a tetramer of histones H3 and H4 is stacked between two dimers of histones

H2A and H2B (Figure 1.1B). The stacking model can be viewed as a data guided 3D animation, created by D. Berry (5). Roughly 200 bp of DNA is wrapped twice around the histone complex. HistonesH1/H5 interact with the 'linker' DNA at the entry and exit site of the wrapped DNA (6, 7). Nucleosome structures appear repeatedly along each linear chromosome in eukaryotic cells and support higher-order packaging of the entire genome (Figure 1.1B).

In natural systems, the histone octamer is modular and dynamic. There are four natural variants of H2A and H3 with distinct amino acid sequences (reviewed in 8, 9), while histones H4 and H2B are largely invariant (10). Substitutions of H2A and H3 with variants in the octamer complex play critical roles in gene regulation, DNA replication, and chromosome structure (4). Kinetic studies of fluorescently labelled histones have shown that H3 and H4 turnover is very slow. In contrast, exchange of the H2A/H2B dimer occurs within a few minutes to two hours (11). Exchange of histone H1 occurs in under two minutes (12). Addition and erasure of post-translational modifications (PTMs) is another highly dynamic feature of nucleosomes. Histone-modifying enzymes covalently link or remove small molecules at the side chains of specific amino acids within each histone. These modifications take place mostly in the unfolded tails, while a few occur in the globular domain. Over 15 known modifications include lysine acetylation (Kac), lysine methylation (Kme), serine phosphorylation (Sp), sumoylation (su), ubiquitination (ub) and crotonylation (cr) (13) (Figure 1.2). In total, over 50 different amino acid positions are known to be modified. Certain types of PTMs result in transcriptional silencing of a nearby gene, while others enable activation.

Collectively, PTMs make up a rich set of biological information in which single or combinations of molecular tags recruit histone-binding effectors to target genes. By the year 2000, dozens of PTMs had been documented, relationships between histone modifications and gene regulation states began to materialize, and the term 'histone code' was coined by Brian Strahl and David Allis (14–17). Since the first report of the structure of the histone-binding domain from P/CAFin 1999 (18) a plethora of other 3D structures have become available to the scientific community. Investigations of binding pocket specificity support the idea that peptide motifs can distinguish one PTM from another (19–21). These protein structure and interaction data can be used by synthetic biologists to design artificial epigenetic systems and to further confirm or correct aspects of the histone code model. Taverna et al. provide an excellent detailed review of lessons learned from the molecular structures of PTM-binding domains (19).

In spite of its potential usefulness, chromatin is often perceived by biological engineers as an impediment rather than as an enabling tool. Cells are typically engineered by integrating exogenous, recombinant DNA into the chromosomes of the host cell. These transgenes include regulatory components that are carefully designed to operate with predetermined kinetics. However, the transgene often becomes subjected to the surrounding chromatin environment and is mis-regulated (silenced or hyper-activated). The molecular complexity of chromatin may give scientists the impression that chromatin-mediated expression states are impossible to control. Chromatin complexes are often composed of multiple subunits, which have several paralogs in a single organism. For instance, Polycomb Repressive Complex 1 (PRC1) appears as six sub-types that occupy different genomic regions (22). Each of the PRC1 subunits may be one of several

distinct paralogs. Furthermore, the core subunit of chromatin known as the nucleosome contains two copies of four types of histones (H2A, H2B, H3, H4) (4), two of which have multiple variants. Histones H3 and H2A have eight and five known variants, respectively. The variants differ in primary sequence, genome distribution, and expression in different tissues and phases of the cell cycle (23, 24). Compared to simpler biological principles such as Watson–Crick base-pairing, the complex interactions that govern the behavior of chromatin may seem less amenable to bioengineering. Is it worthwhile to attempt to engineer multi-layered systems like chromatin within a complex cellular milieu? Synthetic biologists have demonstrated so far that such work produces valuable new knowledge as well as useful innovations (25, 26).We believe that the current wealth of information produced by decades of research in chromatin epigenetics provides a sufficient platform to support engineering efforts. In this review, we discuss how proteins and nucleic acids that guide epigenetic regulation in nature have been harnessed for custom-built systems. Specifically, we focus on the molecular structures of chromatin proteins and how our understanding of molecular interactions can be leveraged for chromatin engineering. We discuss best practices for chromatin engineering endeavors and present a flexible, standard workflow for efficient, high-throughput engineering of chromatin-derived proteins.

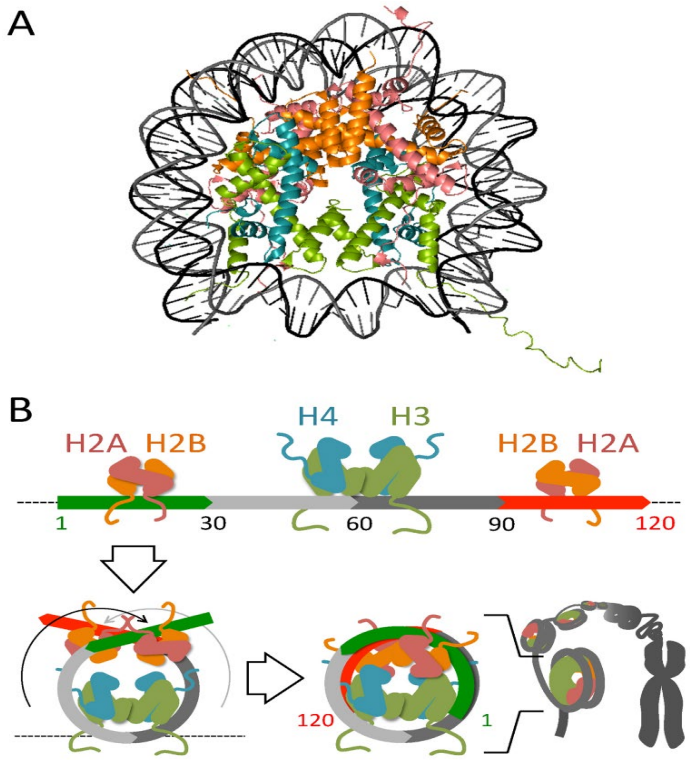


Figure 1.1. The nucleosome, the core subunit of chromatin, is a modular protein complex. (A) 3D model based on K. Luger's 1997 X-ray crystallography data, PDB ID: 1AOI (3). (B) The cartoon abstraction shows the general shape of each histone: the Z-shape is the globular region and the thinner line is the unfolded N-terminal tail. A H3/H4 tetramer (green and blue) and two H2A/H2B dimers (orange and pink) bind across ~120 bp DNA and become stacked to form a solenoid, DNA-wrapped structure (for a data-based animation, see (5)). The lower-right cartoon depicts higher-order packaging of nucleosomes into a metaphase chromosome.

Engineering Nucleosomes, the Core Subunits of Chromatin

Early efforts to engineer nucleosomes used chemical reactions to modify purified histones *in vitro*. Chemical ligation (Figure 1.2B) has been used to attach synthetic, pre-modified peptides into recombinant histone proteins. A detailed

discussion of specific chemical methods is available in (27). Chemical ligation has been used to generate histones bearing acetylation at H3K4, 9, 14, 18, 23, H4K5, 8, 12, 16, tri-methylation at H3K9, and phosphorylation at H3S10 (28–30). Soon after, methylation intermediates (one, two or three methyl groups) were produced on analogs of H3K9, 36, 79 and H4K20 (31). Chemical ligation, and subsequently split inteins,were used to produce ubiquitinated H2B (32,33). Semi-synthetic histones have enabled scientists to build custom nucleosomes and nucleosome arrays to investigate how isolated proteins engage with diverse chromatin templates (27).

Recent advances have allowed scientists to build modified histones via translation of mRNA (Figure 1.2B). Engineered synthetases have been used to charge amber stop codon-binding tRNAswith pre-modified amino acids. During translation, the modified amino acid is incorporated at any position where a codon has been replaced with UAG in a customized (recoded) mRNA sequence. Acetylation at H2AK9, H2BK5 and K20, H3K56, K23 and K27, and H4K16 were all recombinantly expressed in bacteria with this method (34, 35). Kim *et al.* demonstrated successful incorporation of a crotonyl group at position K11 in histone H2B (36). Recently developed synthetases that charge tRNAs with propionyl-lysine, butyryl-lysine (37) and ε-*N*-2- hydroxyisobutyryl-lysine (38) further broaden the spectrum of expressed custom histones. In a landmark mammalian cell study, Elsasser *et al.* used a genetically-integrated recoded mRNA translation system to produce pre-modified histones in E14 mouse embryonic stem cells. Six lysine codons in the genes for histones H3.2 and H3.3 were replaced with amber stop codons. Incorporation of acetyllysines at H3.3 positions 9, 23, 27, 37, 56 and 64 resulted

in the upregulation of 16 genes, including the locus for the noncoding RNA *Xist*, compared to a wild type H3.3 parental line (39). This technology provides a powerful platform to explore the impact of a variety of histone modifications, natural or novel, in the context of live cells.

Fusion proteins have been used to generate and erase PTMs on endogenous histones, and to enhance or repress transcription at genomic sites in live cells. Histone modifying enzymes can be fused in-frame with a DNA binding domain (i.e. dCas9/gRNA, Gal4, LexA, TALE, TetR and Zinc Finger) to generate silencing- or activationassociated PTMs at a gene or non-coding locus (Figure 1.2B). Fusion enzymes have been used to repress gene expression through methylation of histones: H3K27 via Enhancer of Zeste homologue 2 (EZH2) (40, 41), EED (42) and Nuclear Effector (NUE) (43), methylation of H3K9 via Suppressor of Variegation 3–9 homologue 1 (SUV39H1), G9A (44), Kryptonite (KYP) (43), DNA methyltransferase via DMNT3B (42), lysine methyltransferase 1D (EHMT1) (45); and methylation of H4K20 via TgSET8 (43). Gene repression has also been controlled by removal of methyl groups from H3K4 via lysine demethylase 1A (LSD1, KDM1A) (46, 47), and removal of acetyl groups from histones via histone deacetylase 4 (HDAC4) (37), HDAC8, RPD3, silent information regulator 2 (Sir2a), and Sin3a (43). Examples of gene-activating PTMs generated by fusion proteins include acetylation of lysines via the catalytic domain of p300 (48, 49) and P/CAF (KAT2B) (48). In a creative application of light-regulated peptides, Konermann *et al.* controlled histone PTMs H3K9me1, H4K20me3, H3K27me3, H3K9ac and H4K8ac with a TALE-cryptochrome 2 fusion and its light-induced conjugating binding partner CIB1

fused with a variety of catalytic domains (43). Fusion protein-mediated histone modification has been reviewed in detail in other excellent reviews (26, 50).

Fusion proteins have demonstrated great utility for testing simple assumptions about the impact of histone PTMs on gene expression. Further progress in the use of histone modifying fusion proteins to regulate genes requires careful consideration of important contextual aspects: cross signaling between PTMs and interactions of specific PTM modifying enzymes with other enzymes. Cross-signaling between PTMs occurs when one modification leads to the generation or erasure of another modification (51). This event is mediated by enzymes that bind pre-existing histone modifications and then catalyze new PTMs (reviewed in 52– 54). A classic example is H3K9 methylation by the HP1 SUV39H1 complex. H3K9me3 is recognized by HP1, which recruits SUV39H1, an enzyme that methylates K9 at neighboring H3 histones (55). Other examples of crosstalk occur between distinct types of modifications. Phosphorylation of H3S10 induces Gcn5 to acetylate H3K14 in yeast (56). H2B ubiquitination stimulates methylation of H3K4 and K79 via hDot1L (57). Complexes that contain SGF29, NuA or HBO bind H3K4me3 and acetylate histones H3 andH4 (54). In the case ofRpd3S, cross-signaling promotes the conversion of one gene expression state into another. H3K27me3, which is associated with active transcriptional elongation, is targeted by Rpd3S which removes acetyl groups from histones H3 and H4 to promote silencing (58). Cross-signaling can also be inhibitory, where the presence of one PTM blocks the function of another. The activation associated mark H3K4me3 blocks interaction of the PRC2 complex withH3, prevents methylation of H3K27, and prevents gene silencing (59). H3S10p disrupts binding of Dido3 with H3K4me3 (60) and enables progression through mitosis. The second

contextual aspect we discuss here is the corecruitment of PTM-modifying enzymes. KRAB is a strong repressor that has been widely used in early and recent work to silence gene targets (e.g. see (41, 42, 61–64)). The consequences of KRAB fusion-mediated regulation are complicated by interactions with H3K9 methylases (via HP1) as well as histone deacetylases (65). Unintended recruitment of endogenous enzymes can be avoided in some cases by using only the core catalytic domain of the histone modifier (e.g. p300 (48, 49)). In summary, the wealth of information from mechanistic studies such as those cited here should be leveraged to design effective chromatin engineering strategies. Pre-existing PTMs at target loci should be determined so that scientists can predict whether fusion enzyme activity will be blocked or enhanced. Recruitment of PTM-modifying binding partners should either be taken into account or avoided in experiments that aim to investigate mechanisms associated with individual PTMs.

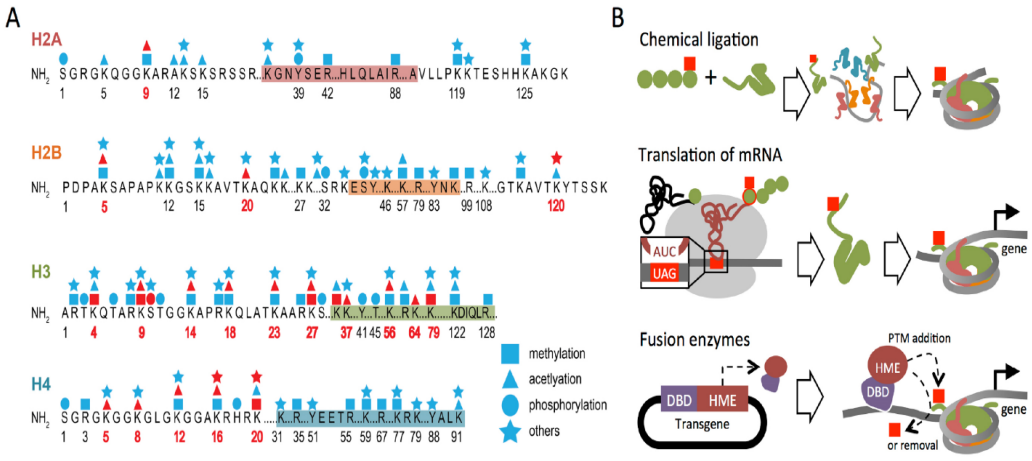


Figure 1.2 Graphical survey of known histone posttranslational modifications and techniques for generating histone PTMs. (A) PTMs that have been artificially generated are shown in red (28–38, 40, 43–46, 49, 63, 150). Positions of methylation, acetylation, phosphorylation and other PTMs are shown for each canonical histone. Other modifications (stars) include citrullination/deimination, ADP-ribosylation,

propionylation, butyrylation, formylation, proline isomerization, hydroxylation, malonylation, glutathionylation, crotonylation and succinylation (166). Globular, folded histone regions are shaded. (B) Illustration of three general methods for generating histone PTMs. Chemical ligation adds pre-modified peptides to recombinant histone proteins. The products are assembled with other histones and DNA to generate nucleosomes in vitro. Translation of mRNA using engineered tRNA synthetases incorporate modified amino acid residues into genetically encoded amber stop codons (UAG) in the open reading frame of a custom sequence/histone. Fusion enzymes are produced in cells by expressing a DNA-binding domain (DBD) in-frame with a histone-modifying enzyme (HME) or catalytic domain. The fusion protein binds a DNA target and generates or erases PTMs at native nucleosomes.

Engineering Proteins to Recognize Histone Tags

PTM-binding proteins can be used to integrate information from histone marks into engineered systems by physically interacting with histone PTMs. In natural systems, PTM binding proteins act as effectors that regulate gene expression at sites that are enriched for the target histone modification. Here, we highlight four PTM-binding motifs that have been used for chromatin engineering: the chromodomain (CD), bromodomain (BRD), baculovirus inhibitor of apoptosis repeat (BIR) domain, and PHD finger. Several studies have demonstrated that these PTM-binding motifs retain their intrinsic function as isolated peptides, and in some cases after they are incorporated into a fusion protein (Table 1.1). For a detailed discussion of other well-characterized PTM-binding domains, see Park *et al.* (26).

The chromodomain (CD) motif interacts with methylated histone lysine residues. 3D structure analyses and binding assays have demonstrated that the HP1 chromodomain (66–68) and Polycomb chromodomain paralogs (68–70) show preferential binding to their cognate PTM targets *in vitro*. The CD consists of three β strands packed against a C-terminal α-helix, and a hydrophobic pocket that interacts with methyl-lysine (Figure 1.3A) (66,68,69,71,72). Various CD-containing proteins interact with histone H3 (K4, K9, K27, K36 and K79) and histone H4 (K20) methylated once, twice or three times (me1, me2, me3) (73). Structural studies determined how heterochromatin protein 1 chromodomain (HP1 CD) specifically recognizes trimethylated (me3) histone H3K9. K9me3 is buried in a binding pocket comprised of three aromatic residues, while four amino acids preceding K9 interact with the chromodomain (67). Mutational analysis of a Drosophila HP1 CD showed that the hydrophobic binding cage is necessary for HP1 CD ligand affinity (66). Side chain interactions between the H3 tail and residues from HP1 form a zipper-like β-sheet, underscoring the contribution of K9-adjacent histone residues to HP1 CD binding. The Polycomb chromodomain (PCD) is a different type of CD that preferentially binds H3K27me3 and has been shown to cross react with H3K9me3 *in vitro* (71). Although they share high levels of sequence similarity, five mammalian PCDs (CBX2, 4, 6, 7, 8) show significant differences in binding preferences (70). Negatively charged and hydrophobic surfaces distinguish two classes of mammalian PCDs that have high and low affinity, respectively, for their histone ligands (74). Like HP1 CD, the CBX PCDs have a conserved binding motif where a β-strand from the histone tail forms a β sheet with the CD. PCDs share a unique binding site for A25, which fits into a hydrophobic pocket that will not tolerate any other amino acids. In summary, the

structural studies of chromodomains suggest two general requirements for CD binding: a hydrophobic pocket and a CD-histone beta sheet (Figure 1.3). Studies in live cells have demonstrated that CD peptides retain intrinsic PTM-recognition activity within fusion proteins. In Drosophila, fusion proteins containing beta-galactosidase and either the HP1 CD (75) or the Polycomb CD (76) show binding distributions on chromosomes that are similar to the corresponding natural proteins. The Polycomb CD from CBX8 has been used to build a synthetic activator that stimulates gene expression at H3K27me3-enriched genes in human cells (77, 78), potentially counteracting oncogenic, repressed chromatin states.

Bromodomains (BRDs) bind acetylated histones tails. These motifs are typically 110 residues in length and appear in histone acetyltransferases and nucleosome remodeling complexes (79–81). Isolated BRD domains from tail peptides in 3D structure studies and in affinity assays (79, 80, 82, 83). Although there is a high degree of sequence variation among the 61 known human BRD proteins, they share a conserved tertiary structure consisting of a lefthanded bundle of four α helices linked by variable length loops that surround the histone binding site (Figure 1.3B). For instance, the BRD region of BRD2 contains a deep hydrophobic pocket contributes to the affinity for acetylated lysines (82, 83). Specificity is determined by the primary sequences of the variable loops (80) and by a hydrogen bond between the oxygen of the acetyl carbonyl on the peptide and the amide nitrogen of a conserved asparagine residue (Asn407) (79). In some cases, BRDs recognize PTMs other than acetylation (80). Natural bromodomains from BRD9, CECR2 and TAF1 have been shown to bind to butyryland crotonyllysine (84). Flynn *et al.* discovered a gain-offunction mutation in BRD1 that conferred affinity for butyryl,

with no loss of the intrinsic acetyl binding (84). Studies of a BRD fusion suggests that BRD can be used to engineer proteins that bind acetyl-histones. BRD domains confer histone-binding activity to the NUT protein in cells where an oncogenic chromosome rearrangement fuses the N-terminal BRD domains of BRD4 to NUT (85). A potential use for BRD domains is a fusion that binds acetyl marks and represses target genes. Aberrant histone acetyltransferase activity has recently become the focus of novel anti-inflammatory therapeutics (reviewed in (86)).

Table 1.1 PTM-binding motifs that retain their intrinsic histone specificity in isolation or as a sub-domain within a synthetic or oncogenic fusion protein

Motif	**Protein (length, a.a.)**	**Species**	**Target PTM(s)**	**Experiments with isolated peptides *in vitro***	**Peptide (a.a.)**	**Synthetic (S) and oncogenic (O) fusions studied in cells**	**Peptide (a.a)**
CD	HP1 (206)	*D. mel*	H3K9me2, 3	X-ray crystallography (47, 51) FP (70)	17-76	(S) CD-beta-gal (72)	1-95
CD	Pc (390)	*D. mel*	H3K27me3	X-ray crystallography (70-71), FP* (70)	15-77, 23-77, 1-98*	(S) CD-beta-gal (73)	1-117
CD	CBX8 (389)	*H. sap M. musc**	H3K27me3	FP* (52)	1-62	(S) CD-VP64 (7, 74)	1-62
BRD	BRD4 (1362)	*H. sap*	H3Kac, H4Kac	X-ray crystallography, SPOT array (55)	44-168, 333-460	(O) BRD4-NUT (75)	1-719
BRD	BPTF (3046)	*H. sap*	H4K12ac, H4K16ac, H4K20ac	X-ray crystallography, SPOT array, peptide pulldown, FP, ITC (76)	2781-2890	(S) FLAG-PHD-BRD (76)	2722-2890
BIR	BIRC5 (142)	*H. sap*	H3T3p	NMR (62)	1-120	n/r	

PHD	BPTF (3046)	*H. sap*	H3K4me3	X-ray crystallography, SPOT array, peptide pulldown, FP, ITC (76)	2722-2781	(S) FLAG-PHD-BRD (76)	2722-2890
PHD	Dido3 (2256) Dido1* (614)	*M. musc*	H3K4me3	NMR, X-Ray crystallography, fluorescence microscopy, peptide pulldown, microarray, tryptophan fluorescence, (68)	266-325	(S) HA-DIDO (68)	1-528*

Representative examples from published reports are included in the table. CD = chromodomain, BRD = bromodomain, FP = fluorescence polarization assay, ITC = isothermal titration calorimetry, n/r = none reported. Lengths of proteins are from Uniprot.

The baculovirus inhibitor of apoptosis repeat (BIR) domain is an approximately 70 amino acid domain containing a zinc coordinated by histidine and cysteine residues, and recognizes phosphorylated histones (87) (reviewed in 88). BIR domain binding has been determined by X-ray crystallography of full-length BIRC5 bound to H3T3p (87, 89, 90). The BIRC5 (Survivin) protein contains a BIR domain that binds to phosphorylated histone H3 threonine 3 (H3T3p) by surrounding the first four residues of the peptide with 10 negatively charged residues. The residues from BIRC5 that make direct contact with the H3 tail are conserved across vertebrates (89, 91). Structural analysis shows that the BIR domain contains an accessible, preformed binding site for H3T3p recognition, which suggests a rigid scaffold for binding. This BIR-H3 interaction pattern is highly conserved in other regulatory proteins that contain BIR domains (87, 90). Studies revealed a consensus binding motif that includes an N-terminal phosphorylated A-X-S/T-R/K (90)

(Figure 1.3C). Mutational analysis targeting the histone-interacting residues of BIR resulted in reduced binding and reduced centromere localization or poor microtubule depolymerase (MCAK) recruitment in cells (87). Kelley *et al.* showed that BIR interacts with its cognate histone PTM in the absence of other subunits from the chromosomal passenger complex (89). This observation suggests an intrinsic histone recognition activity that might enable a BIR fusion protein to bind H3T3p. No BIR fusion proteins have been reported, so this is yet to be determined.

Plant homeodomain fingers (PHD) are a class of diverse motifs consisting of two 50–80 amino acid domains that contain a zinc binding site. PHDs generally interact with histone H3 trimethylated at lysine 4 (92, 93). Isolated PHD fingers from several proteins, such as Yng1 (94), ING2 (95), TAF3 (96), NURF(97, 98), ING3 (99), JARID1A (100) and DIDO (Death Inducer Obliterator) (60) have shown similar interactions with H3K4me3 in different crystal structure studies. The PHD finger of L3MBTL1 recognizes its target via cavity insertion (101). The PHD domain of DIDO (Figure 1.3D), binds H3K4me3 with a noncanonical aromatic cage that contains a histidine residue. When aligned against other PHD fingers known to bind to H3K4me3, all proteins shared a conserved tryptophan residue (W291 in full length Dido3) in the aromatic cage, as well as other residues known to coordinate zinc ions (60). Using structural information from methyllysine binding, Li *et al.* engineered a plant homeodomain (PHD) finger, which normally binds trimethylated lysines (H3K4me3), to preferentially bind me2 and me1. A single mutation (Y > E) in the aromatic cage responsible for substrate recognition changed the mode of interaction from cavity insertion to surface groove recognition. PHD motifs can also recognize acetyllysines, as observed for the protein DPF3b which

interacts with H3K14ac (102). PHD motifs retain their intrinsic activity in fusion proteins. The oncogenic fusion of nucleoporin 98 (NUC98) with the PHD domain from JARID1A shows broad colocalization with H3K4me3 and accumulates at the H3K4me3-enriched *HOXA9* gene in human cells (100). In a study using mouse cells, an HA-tagged N terminal peptide containing the PHD domain from Dido1 (528 aa) was sufficient to rescue the epigenetic activity of the full length Dido3 isoform (2256 aa) (K. van Wely, personal communication) and (60)). A mouse model has been generated to express a DIDO PHD-RFP fusion (60,103), but neither the subcellular localization nor histone interaction of this fusion protein have been reported yet.

Allosteric regulation of PTM-binding proteins by small molecules suggests that chromatin proteins can be designed to switch between binding and nonbinding states at will. Tightly-regulated chromatin systems enable insights into epigenetic dynamics, as seen in other work where a small molecule (doxycycline) was used to control docking of TetR fusions at reporter genes (42,64). Gelato *et al.* described a H3 binding protein, UHRF1 that is allosterically regulated to allow or to block binding to modified histones. UHRF1 has two binding states that target either unmodifiedH3 tails via a homeodomain or H3K3me3 via a tandem tudor domain (TTD). In the apo state, a polybasic region on the C terminus sterically occupies a binding groove in the TTD, disrupts interaction with H3K9me3, and allows the homeodomain to bind to unmodified H3 tails. When UHRF1 is bound by phosphatidylinositol phosphate at the polybasic region, a global conformational change is induced, allowing the tudor domain to bind to H3K9me3. This mechanism demonstrates basic allosteric modulation of a histone reader (104). In summary, mechanistic studies of PTM binding protein folds have provided a useful, relatively

untapped resource for fusion protein design. The recognition of specific histone marks could enable broad co-regulation of sets of genes (Figure 1.3E). Alternatively, DNA and histone recognition could be coupled to conditionally control a single gene target based upon the epigenetic state of the locus or allele (Figure 1.3F).

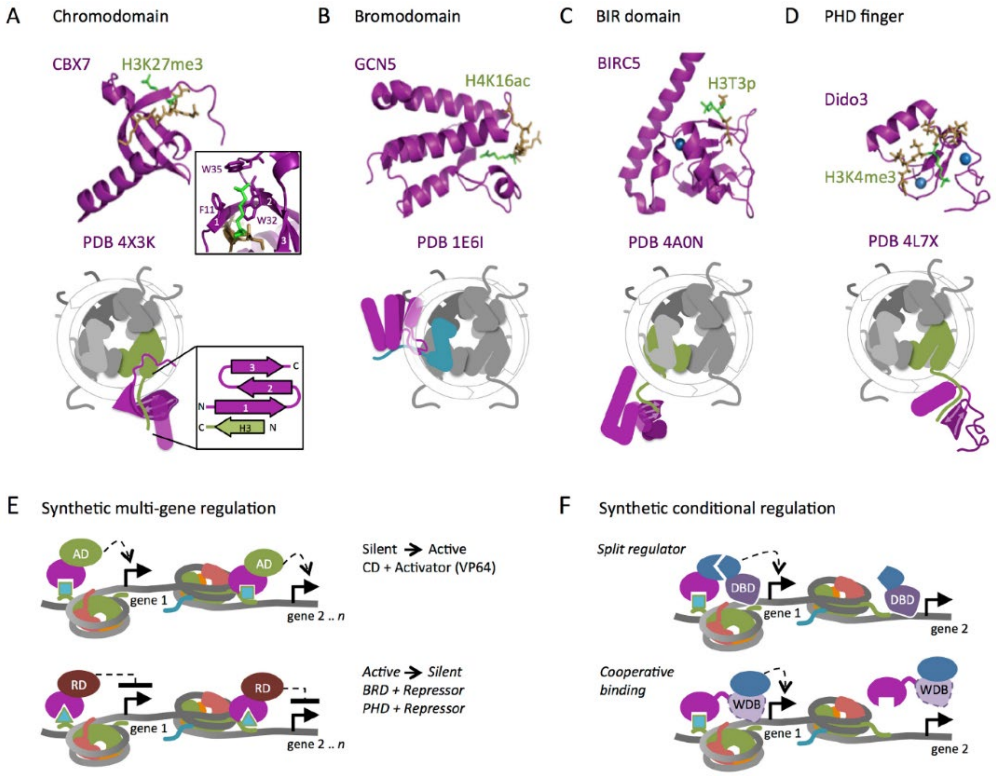


Figure 1.3 Structural features of representative histone binding domains and their applications in synthetic systems. (A) The Chromodomain of CBX7 (116), (B) bromodomain of GCN5 (79), (C) BIRC5 (Survivin) subunit of the Chromosomal Passenger Complex bound to H3T3p (90) and (D) the PHD finger of Dido3 (60). Green = histone sidechain and PTM, blue sphere = coordinating metal ion. Cartoons below each structure depict the binding domain–PTM interaction in the context of a whole nucleosome. (E, F) Cartoons illustrate uses of histone-binding domains to design synthetic transcriptional regulators. Hypothetical applications have italic captions. (E)

Synthetic effectors co-regulate all genes (gene 1, 2, ..., n) near a target PTM. This application could be used to convert silencing PTMs into gene activation and activating PTMs into gene repression. The CD from CBX8 has been used to build a synthetic reader that targets the H3K27me3 gene-silencing mark and stimulates gene expression through a C-terminal VP64 domain (77, 78). (F) Synthetic regulators could be designed to integrate DNA sequence and PTM information at a specific target gene or allele. (top) Assembly of a split transcriptional activator requires the binding of two fusions. (bottom) Stable binding of weak DNA binding domain (WDB) requires the presence of a specific PTM. AD = transcriptional activator domain, RD = transcriptional repressor domain.

Multivalent Histone-Binding Domains Integrate Multiple PTM Signals

Natural multivalent PTM-binding proteins and complexes can interpretmultiple epigenetic marks at once and increase binding affinity for a target histone (26). Structures of these multivalent proteins have provided insights into how multivalency can be exploited for synthetic systems. Recognition of combinatorial PTMs could provide greater specificity. For instance, Su *et al.* reported a K_d of 45 nM for the dual spin/Ssty repeat domains of Spindlin1 for histone H3 trimethylated at K4 and dimethylated at R8 (H3K4me3- R8me2) (105). The K_d was lower (139 nM) for an off-target peptide H3K4me3R8me1. Multivalent binding might compensate for the intrinsic low affinity of interactions between single PTM-binding proteins and single targets. In the same investigation, Su *et al.* observed reduced binding affinity (22 μM) of Spindlin for H3R8me2 alone (105). The HP1 protein achieves enhanced affinity through self dimerization. In a study of HP1 *in vitro*, mutations that disrupted HP1 self-dimerization

showed reduced affinity for H3K9me3 (106). The dimerization motif is also required for HP1 activity in cells (107,108).

Research of natural and synthetic multivalent domains suggests principles for designing proteins that recognize more than one PTM. Spacing and orientation of the PTM binding motifs within the protein appears to be critical for optimal function. The double bromodomain module from TAFII250 recognizes H4K5acK12ac (K_d of 1.4 μM). Interestingly, the distance between the tandem bromodomains matches the distance between the acetylated residues (109). In a study of the tandem PHD finger and bromodomain of BPTF, Ruthenburg *et al.* observed that the insertion of two amino acids (QS) into the rigid alpha helical linker region to rotate the domains 200° out of phase disrupted binding to a doubly-modified nucleosome (94). Mutations in the linker that added flexibility also impaired bivalent binding, underscoring the importance of domain orientation in multivalent interactions.

The arrangement of nucleosomes and histone PTMs at genomic target regions influences the function of multivalent PTM-binding proteins. The combined contribution of the PHD finger domain in Rco1 and the chromodomain in Eaf3 direct the histone deacetylase complex Rpd3S to active loci that are enriched H3K36me (58, 110). Hu *et al.* proposed that di-nucleosome recognition reinforces Rpd3S binding, thus allowing Rpd3S to tolerate fluctuations in H3K36me levels (111). Optimal binding was achieved *in vitro* when the adjacent nucleosomes were 30–40 bp apart (112). BPTF is another example of a protein that shows a preference for specific PTM placement. In experiments with reconstituted, customized nucleosomes, the divalent PHD bromodomain from BPTF

showed greater interaction with histones when the target PTMs were placed on a single nucleosome, compared to PTMs that were distributed across two nucleosomes (94).

Disrupting Chromatin: Synthetic Antagonists of Chromatin Complexes

Chromatin complexes are stabilized by several intramolecular interactions between proteins and histone tails as well as between subunits within the complex. Molecular antagonists have been designed to disrupt chromatin complexes and alter gene expression states by targeting a core subunit that recruits other binding partners to the gene target. Polycomb Repressive Complex 1 types 2 and 4 (PRC1.2, PRC1.4) generally include a histone PTM-binding chromobox (CBX) paralog, RING1A/B, RING1 and YY1- binding protein (RYBP), and polycomb group ring finger 2 (PCGF2, Mel-18) or polycomb group ring finger 4 (PCGF4, BMI1) (22) (Figure 1.4A). Histone deacetylases (HDACs) support PRC function by removing acetyl groups from H3K27 to allow methylation, which is recognized by CBX. These proteins work in concert to support chromatin compaction and gene silencing, often at tumor suppressors in cancer cells. To perturb CBX7 activity, inhibitors have been designed to bind H3K27me (113) or CBX7 (114–116). Informed by the solved structure of CBX7 in complex with a native peptide or a synthetic inhibitor, Ren *et al.* performed an *in vitro* screen of a library of compounds to identify CBX7 inhibitors. These compounds (see Figure 1.4B for an example) led to de-repression (transcriptional activation) of CBX7 target genes in human prostate cancer cell lines (72). An *in silico* screen identified a structurally unique compound that demonstrated improved cellular activity. When tested in mouse ES cells, highly specific de-repression of CBX7 target genes was observed, compared to almost no impact on non-target genes (72). Simhadri *et al.* started with trimethyl-lysine peptides and structural data

to eventually develop a peptide to directly bind the CBX7 CD and inhibit its function (114) (Figure 1.4C). In later work, Stuckey *et al.* used a molecular dynamics platform to develop a compound that mimics natural peptide binding to CBX7 (115). Structure-guided work has also led to the discovery of inhibitors of the paralog CBX6 (117). Lastly, molecular structures have enabled the discovery of inhibitors of PTM-modifying enzymes. Whitehead *et al.* identified class 1 HDAC inhibitors that are selective for the HDAC8 isoform (118) (Figure 1.4D). This work demonstrates that enzyme inhibitors can be highly specific while maintaining biologically-relevant affinities.

BRD2 interacts with acetylated histones, as well as several non-histone proteins including glioma tumor suppressor candidate region gene 1 (GLTSCR1), the histone arginine demethylase JMJD6, and histone methyltransferase NSD3to stimulate gene expression (119) (Figure 1.4E). BRD containing proteins stimulate MYC oncogene expression and promote the growth of cancer cells (119). Therefore, BRD is an important therapeutic target (81). Bromodomain inhibitors such as JQ1 (120) and I-BET (121) have been identified in high throughput screens to identify small compounds that fit the histone recognition pocket of BRD2 (1- 473), BRD3 (1-434) and BRD4 (1-477) (121).

A noteworthy example of structure-guided design come from James *et al.*, who developed strong inhibitors that bind within the methyl-lysine recognition cavity in malignant brain tumor (MBT) repeat domains (122), and others who developed inhibitors of the H3K27 methyltransferase EZH2 (123, 124). Horton *et al.* determined the structures of a diverse set of demethylase inhibitors to elucidate the molecular mechanism of binding. Their work provides a general platform for designing epigenetic inhibitors. Structures of the demethylase complexed with various small molecules indicated the

potential diversity of inhibitors. The complexes showed a similar binding interfaces but differed in atomic interactions, such as metal coordinating ligands (125).

Since the first reports of the role of Polycomb proteins in oncogenesis and metastasis, a multitude of inhibitors have been developed as epigenetic drugs. Inhibitors of PTM-generating enzymes and their impact on drug discovery is discussed in several excellent reviews (126–131). Structure-focused work has illuminated the molecular interactions between inhibitors and PTM-binding proteins or histone deacetylase enzymes (HDACs). Stuckey *et al.* demonstrated that a PRC1 inhibitor could be designed to inhibit human prostate cancer cell (PC3) proliferation (115). The structure of the inhibitor is similar to the natural ligand H3K27me3 and is highly specific for two chromodomains, CBX4 and CBX7. Inhibitors that bind and interfere with bromodomains (BRDs) have been used to control gene expression in disease-relevant cell culture model systems. Small molecule-mediated inhibition of BRD was used to down-regulate MYC transcription in leukemia and myeloma cells (132,133) and to suppress inflammation in bone marrow-derived macrophages (121). Recently, *in vivo* work has shown that a new BRD-specific inhibitor, MS402, preferentially binds to the first bromodomain motif within BET proteins and is therefore more specific than the broadacting inhibitor JQ1 (Figure 1.4F). Cheung *et al.* demonstrated the therapeutic potential of MS402 by preventing and reducing T-cell transfer-induced colitis in mice (134). These studies underscore the clinical potential for synthetic, epigenetic regulation of genes that have therapeutic properties.

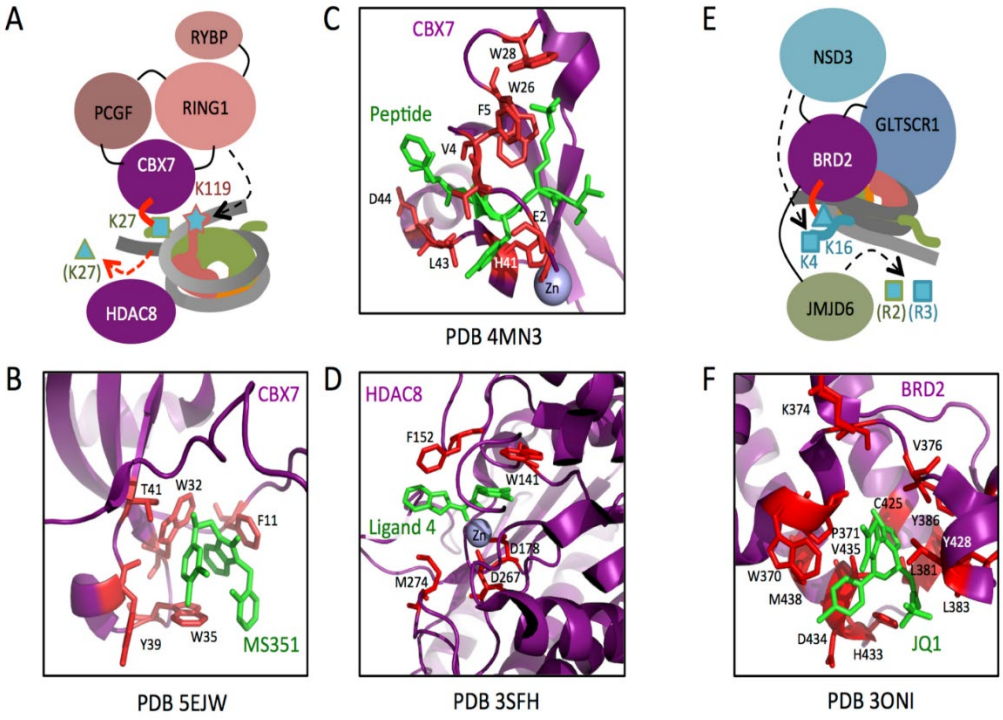


Figure 1.4 Artificial disruption of multi-protein chromatin complexes. Cartoons depict the context in which inhibitors interact with specific subunits. Shapes are scaled to reflect relative protein sizes (kiloDaltons). Histone PTM symbols are the same as in Figure 1.2. Arcs = protein interactions (referenced in the text), dashed arrows = histone modifying activity, red lines = inhibitor-mediated disruption. (A) The PRC1.2/1.4 type silencing complex can be disrupted by inhibiting the binding of CBX7 CD to H3K27me3 or by inhibiting the removal of acetyl groups from H3K27 by HDAC8. Examples of CBX7 CD inhibitors include MS351 (116) (B) and a synthetic peptide (114) (C). (D) The inhibitor Ligand 4 (118) disrupts HDAC8 activity. (E) BRD2 recruits several proteins, including histone modifiers, to stimulate gene activation. Inhibitors such as JQ1 (120) (F) have been designed to disrupt the central histone-binding activity. (B–D, F) Boxes show sites within chromatin proteins where low molecular weight inhibitors (green) interact with specific residues (red).

Building Chromatin: Scaffolding Chromatin Complexes with RNA and DNA

In natural systems, histone-modifying proteins and other chromatin components are spatially arranged along DNA strands. Several studies have demonstrated that specific arrangements of these proteins are non-random, highly regulated, and play a role in epigenetic regulation of gene expression. Chromosome looping, sub-nuclear compartmentalization and gene expression can be altered by manipulating non-coding DNA elements such as the locus control region (LCR), CTCF-bound insulators, and lamina-associated domains (LADs) (reviewed in Park *et al.* (26). Here, we discuss the two classes of epigenetic scaffold elements that have been rigorously analysed within synthetic systems: Polycomb Response Elements and long non-coding RNA.

Polycomb Response Elements (PREs) are *cis*-regulatory DNA sequences that recruit chromatin complexes to clusters of genes. Robust epigenetic control through PRE activity is required to stabilize distinct transcription profiles within subpopulations of cells (135). Drosophila PREs contain clusters of short motifs that interact with several chromatin proteins including Pleiohomeotic (PHO), Zeste protein, GAGA factor (GAF), Dsp1 and others (135). A remarkable characteristic of PREs is 'epigenetic memory', the ability to confer a stable silenced or active state at nearbygenes after removal of the inducer of either state. Earlywork in Drosophila demonstrated that a PRE called *Fab-7* could be used to switch the expression state at an adjacent reporter gene from active to silenced and vice versa (136–138). These artificially-induced expression states persisted over several rounds of mitosis andmeiosis.While Drosophila PREs have been well documented (139), discovery of human PREs has been elusive (140) until recently. In 2010, Woo *et al.* identified a noncoding PRE-like DNA element from the human *HOXD*

gene region (141). When this element, D11.12, was placed upstream of a *luciferase* reporter, the transgene became enriched for repressive histone PTMs and silencing associated proteins and *luciferase* expression was reduced to <5% of the active state. Since this discovery, other human PREs have been identified (142). Experiments have shown that PRE function can be ported fromone metazoan species into another, demonstrating that epigenetic memory is an inherent property of PRE DNA fragments. For instance, mammalian PRE candidates have been validated based on their activity in Drosophila. PREs that were identified in studies of human T cells (143) or the mouse genome (144, 145) showed Polycomb protein binding and repression of a reporter gene in transgenic flies.

Investigations of PREs provide some guidance on how these elements can be used to control synthetic genetic systems. In Drosophila, artificially-induced states mediated by PREs can persist over timescales of metazoan tissue development, a characteristic that is critical for the practical use of synthetic gene circuits in multicellular organisms. An important outstanding question for PRE-based genetic engineering is to what extent the behavior of a PRE can be customized to generate distinct gene expression patterns. Mounting evidence from studies in Drosophila show that PREs respond to inducers either early or late in development, and are sensitive or insensitive to their genomic location (reviewed in (140)). Multiple short motifs of varying copy number and order can be found in PREs from Drosophila (146, 147) and mammalian genomes (135, 141, 142) (Figure 1.5A). Therefore, it is tempting to surmise an underlying protein-scaffolding code. However, current models for PRE function are incomplete. For instance, no human orthologs have been identified for two key Drosophila PRE-recognition proteins, GAF

and Zeste (Figure 1.5A). The Drosophila protein Dsp1 and the human ortholog HMGB2 (hHMG2) have been linked to PRE activity, but there is conflicting evidence for the DNA sequence motif that is recognized by the proteins (135). Finally, although the number of motifs can be identified within a PRE, the stoichiometry of proteins per PRE is not precisely defined. If a deterministic code is eventually identified, artificial PREs could be designed to control the magnitude and dynamics of epigenetic expression in synthetic genetic constructs.

Reminiscent of PREs, synthetic DNA constructs can be designed to include combinations of shortDNAmotifs that are recognized by transcriptional regulators. Artificial operators have been constructed *de novo* from protein-binding motifs found in well-studied bacterial operons (e.g. Tet, lac, trp) (148) as well as eukaryotic loci (e.g. yeast Gal4 UAS, Zinc finger recognition sites) (149). These motifs can be arrayed within a synthetic DNA fragment and placed upstream of a target gene. Chromatin fusion proteins that recognize the motifs assemble at the scaffold based on the number and arrangement of motifs. Keung *et al.* used two orthogonal Zinc finger (ZF) adapters to co-recruit pairs of gene-regulating proteins to a single promoter (150) (Figure 1.5A). Although this study was limited to observing synergy or antagonism between single chromatin proteins and the ZF-VP16 activator, the synthetic ZF platform has potential for more general use, i.e. to control spatial arrangements of pairs or larger combinations of chromatin proteins.

Long noncoding RNAs (lncRNAs) also act as modular scaffolds for multi-protein complexes. The underlying design principles of these elements are just beginning to emerge. Engreitz *et al.* provide a superb review of lncRNA structure and function in the

context of chromatin and epigenetics (151). Here, we highlight examples of well defined modular lncRNAs may eventually inform synthetic RNA scaffold design. The lncRNA known as HOTAIR (HOX transcript antisense RNA) has two modules with secondary structures that each interact with distinct histone modifying complexes to maintain transcriptional repression (152). The 5' domain of HOTAIR binds the PRC2 complex that generates H3K27me3, while the 3' domain binds the LSD1/CoREST/REST complex that stimulates H3K4 demethylation (153). The 5' domain is a distinct module with intrinsic PRC2 binding activity. Structural dissection of HOTAIR revealed that nucleotides 1–300 are sufficient for interaction with PRC2 (153) and that a core motif of 89 nucleotides is required for binding (154). However, recent work has shown that the role of PRC2 is dispensable for HOTAIR-mediated silencing (155). The lncRNA Firre (functional intergenic repeating RNA element) con- tains twelve tandem repeats of an RNA motif that binds the nuclear matrix protein HNRNPU, also known as SAFA (scaffold attachment factor A) (156). The structure of Firre demonstrates that tandem replicate RNA modules can operate as independent units, allowing several proteins to cooccupy a single RNA scaffold. Xist is an example of a lncRNA that interacts with a with a DNA-binding protein as well as other chromatin proteins. SAFA acts as a DNA-binding module that tethers the Xist RNA–protein complex to DNA sequences (Figure 1.5B). The RNA-binding arginine–glycine–glycine (RGG) domain of SAFA is required for interaction with Xist (157). Chromatin proteins may recognize a variety of RNA sequences and secondary structure conformations. Mounting evidence is beginning to shed light on the mechanistic details of protein-lncRNA binding, but there is not sufficient information to engineer lncRNA-derived systems (158).

Scientists must first identify core RNA subunits with intrinsic, portable activity beforewe can use lncRNAs for synthetic systems. For an immediate solution, researchers have turned to protein-RNA modules from well-understood viral systems to design lncRNA-inspired gene regulators. Zalatan *et al.* used the viral RNA stem-loop hairpins *MS2*, *PP7* and *com* to build a scaffold that was recognized by the proteins MCP, PCP and Com respectively (Figure 1.5B). They targeted the scaffold to a specific gene by using a CRISPR/dCas system where the DNA-binding guide RNA (gRNA) included a long RNA extension with different combinations of the viral hairpins. The group demonstrated epigenetic repression of a target gene by recruiting KRABCom fusion to a *com* RNA hairpin (159). The RNA hairpin adapters are also functional in tandem (Figure 1.5B), which suggests that this system could be used to recruit combinations of chromatin proteins to a single locus.

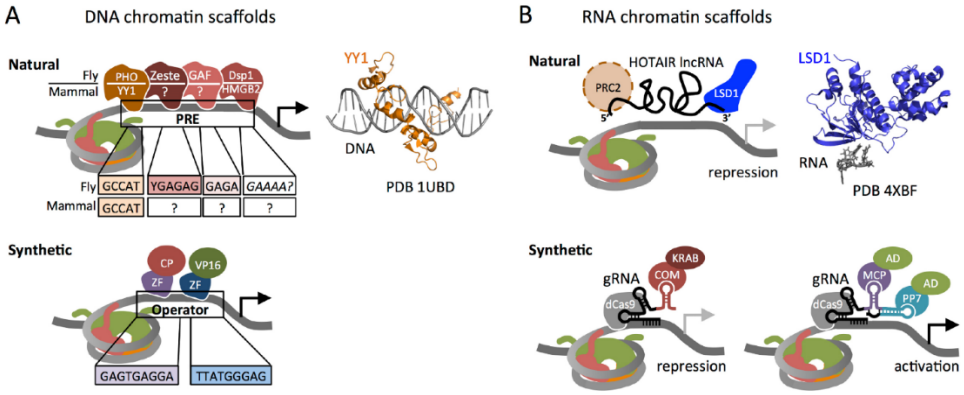


Figure 1.5 Scaffolding of chromatin proteins on DNA or RNA. (A) Natural Polycomb response elements (top) include combinations of short motifs that are recognized by non-histone chromatin proteins. Known PRE-binding proteins and DNA core motifs are shown for Drosophila and mammals. Crystal structure data (168) (top right) shows that YY1 (human PHO homologue) interacts with DNA via a zinc finger motif. A synthetic operator (bottom) designed by Keung et al. is shown as an example of a synthetic DNA

chromatin scaffold. Nucleotides are named according to the IUPAC code. CP = chromatin protein, ZF = zinc finger. (B) HOTAIR (top) is an example of a natural RNA chromatin scaffold. Its 3' and 5' regions act as modules for PRC2 and LSD1/CoREST/REST recruitment. The role of PRC2 (dashed circle) may be dispensable for HOTAIR function (154). LSD1 residues 171–317, 571–654 and 769–836 and RNA (UUAGG) are shown from the crystal structure model (169). Two CRISPR/dCas9-based RNA scaffolds (bottom) are shown as examples of synthetic RNA scaffold systems (158). AD = activation domain.

Conclusions and Future Outlook: Best Practices for Chromatin Engineering

We have described recent advances in synthetic, chromatin derived systems and the fundamental research discoveries that have preceded and enabled the development of these technologies. In this section, we will discuss important next steps and opportunities for continued advancement of chromatin engineering.

Deeper understanding of the consequences of artificial post-translational modifications (PTMs) on chromatin is critical to advance synthetic PTM technology. Exploratory studies should determine the impact of customizable parameters such as the relative position of the PTM nucleation site to the target gene, rate of PTM production, and the interactions between PTMs and regulators at single target sites. Recent studies have demonstrated the value of systematically adjusting design parameters. For instance, Hilton *et al.* placed the active core domain of the p300 acetyltransferase at different distances from the target gene and observed significant activation of genes at distal enhancers and regulatory regions up to 46 kb from the transcriptional start site (49). Bintu *et al.* determined the kinetics of four distinct silencing regulators (EED, KRAB, HDAC4,

DMNT3B) in single CHO-K1 cells to identify conditions that generally enable silencing of a predetermined degree and duration (42). Amabile *et al.* recruited combinations fusion proteins to a single gene to compare silencing induced by one, two or three repressors (KRAB, DMNT3a, DNMT3L). Two or three repressors generated long-term silencing at endogenous genes in human HEK293T, K562 and B-lymphoblastoid cells, human primary T lymphocytes, and mouse NIH-3T3 cells (64). In a study of KRAB and Sss1 co-recruitment, synergy was dampened by competition of each effector for shared cofactors (41). This work revealed an important caveat in using components from overlapping pathways. Keung *et al.* recently reported an impressive study in which 223 PTM-generating fusion proteins enhanced, antagonized, or did not affect the function of a VP16 activator at a target reporter gene in yeast (150). Pioneering efforts such as the studies highlighted here provide a glimpse of the exciting work yet to be done. In addition to customizable parameters, uncontrollable events such as cross-talk between PTMs, enzyme co-recruitment and the impact of these processes on gene expression states (discussed previously in this review) must also be considered and measured to maximize the value of the results from synthetic studies.

Customization of non-enzymatic chromatin proteins represents an expansive design space that has barely been explored. In order to accelerate the pace of discovery, scientists should develop and share workflows to efficiently parse large libraries of synthetic protein candidates. As an example, we present a workflow that is under development in our lab. We begin with a large library of fusion protein variants and carry out protein–histone peptide interaction tests *in vitro* (ELISA) to identify functional candidates (Figure 1.6). Candidates that show preference for the target histone PTM in

the first-pass test are used to build synthetic gene regulators. We then expose a PTM-bearing reporter gene to the synthetic regulator to validate its function in live cells. This workflow allows us to identify modular, reusable PTM binding domains to aid in the design, construction, and application of synthetic, chromatin-derived proteins. In addition, we use peptide arrays to determine cross-reactions with various histone PTMs and to calculate affinities for each target. So far, our workflow has demonstrated the interaction of a Polycomb Chromodomain fusion (PCDmCherry) with H3K27me3 *in vitro* (not published), which corroborates previous studies of this fusion protein in live cells (77, 78). We expect ongoing work to identify new variants with interesting properties, such as enhanced affinity for H3K27me3.

A challenge for practical and reliable use of PTM-binding domains in living cells (e.g. as illustrated in Figure 1.3E and F) is the broad and varying distribution of histone modifications throughout the genome over time and at different stages of cell development. Therefore, it is important to carefully consider the histone PTM target from a systems viewpoint rather than as a single gene, as is the case for DNA-binding regulators. For instance H3K27me3, the target of the PCD motif, appears at thousands of genes and many non-coding regions in human cells. The real utility of regulating a cohort of this size may not be immediately obvious. However, a single PTM might not be sufficient to support artificial regulation of the entire set of genes. We have observed that a regulator fused to the PCD from CBX8 affects only a subset of all H3K27me3-positive genes (77, 78) and that many of these genes are near bivalent H3K27me3/H3K4me3-marked promoters (78). The key lesson from this example is that simple assumptions about epigenetic targets should be tested by using integrated transcriptomic and

epigenomic analyses. The identification of predictive epigenetic signatures at target genes will enable practical use of PTM-binding proteins for synthetic systems.

Eventually, new chromatin engineering tools may become available for human and animal health applications. The safety and efficacy of these tools will need to be determined in the context of a complex epigenome. Therefore, it is critical to determine the global impact of the synthetic component on genome-wide expression levels in order to identify off-target or broad effects. Evidence from HDAC inhibition experiments in pancreas cells (160–164) suggest that generating activation-associated PTMs at key genes could transdifferentiate alpha cells into beta cells, boost insulin production in pancreatic tissue, and cure diabetes. The RNAseq data from Bramswig et al. show that dozens of genes that are not involved in the insulin production pathway are also affected (162). Nyer *et al.* used RNA-seq and ChIPseq to determine that a broad-acting H3K27me3-binding synthetic regulator reactivated silenced tumor suppressors as well as hundreds of other genes (78). The data from the pancreas and cancer cell studies should be further explored to determine the long-term impact of broad changes in gene expression on cell phenotype. Even in cases where the synthetic chromatin protein binds a single unique DNA target, chromatin modifications nucleated at one site can spread into neighboring loci (discussed in (49)). To verify target specificity of a Cas9-p300 histone acetyltransferase fusion protein, Hilton et al. performed RNA-seq to identify cases where only the target gene was activated and no other genes were affected (49). It is imperative that synthetic biologists and chromatin engineers include genome-wide analysis, bioinformatics, and gene network analysis in their arsenal of research techniques.

Synthetic epigenetic research is still in its infancy. The plethora of structural and biochemical data that describe histone-modifying enzymes, histone PTM-binding proteins, and modular nucleic acid scaffolds present an opportunity to expand the bioengineering toolbox. Histone PTMs represent a rich set of biological information that can be exploited for gene targeting and cellular regulation. Genomic profiling can be coupled with chromatin protein engineering to identify and manipulate epigenetically-regulated target sites. While some epigenetic fusion constructs have been successfully used to activate or repress target genes, the number of utilized domains represents a small fraction of well-characterized PTM binding and modifying proteins (80,149). So far, most of the reported synthetic epigenetic systems are limited to specific regulation of a single genetic locus using PTM modifying enzymes that are fused to DNA-binding domains. Histone PTM-binding proteins fused with effector domains, which have been explored less, could enable co-regulation of many genes at once. The efforts summarized in this review represent important, initial advances into a vast exploration space of potential chromatin protein designs and applications. Chromatin is a central mechanism for precise and reliable control of genes that drive multicellular development. Maturation of the synthetic epigenetics field will produce new technologies and discoveries that will have significant impacts on genetic research, agricultural science and biomedical engineering.

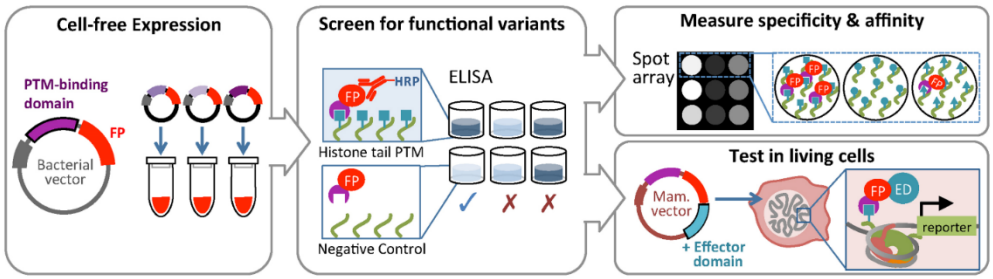


Figure 1.6. Workflow to discover modular, reusable chromatin-derived peptides that bind histone PTM's. FP = fluorescent protein, HRP = horseradish peroxidase-conjugated antibody, Mam. vector = mammalian expression vector, ED = effector domain.

References

1. Turner,B.M. (2005) Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat. Struct. Mol. Biol.*, **12**, 110–112.

2. Schwaiger,M., Schonauer,A., Rendeiro,A.F., Pribitzer,C., Schauer,A., Gilles,A.F., Schinko,J.B., Renfer,E., Fredman,D. and Technau,U. (2014) Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.*, **24**, 639–650.

3. 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 A resolution. *Nature*, **389**, 251–260.

4. Talbert,P.B. and Henikoff,S. (2010) Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.*, **11**, 264–275.

5. Berry,D. (2003) Molecular Visualizations of DNA. http://www.wehi.edu.au/wehi-tv/molecular-visualisations-dna.

6. Lyubitelev,A. V, Nikitin,D. V, Shaytan,A.K., Studitsky,V.M. and Kirpichnikov,M.P. (2016) Structure and functions of linker histones. *Biochem.*, **81**, 213–223.

7. Zhu,P. and Li,G. (2016) Structural insights of nucleosome and the 30-nm chromatin fiber. *Curr. Opin. Struct. Biol.*, **36**, 106–115.

8. Sarma,K. and Reinberg,D. (2005) Histone variants meet their match. *Nat. Rev. Mol. Cell Biol.*, **6**, 139–149.

9. Talbert,P.B. and Henikoff,S. (2016) Histone variants on the move: substrates for chromatin dynamics. *Nat. Rev. Mol. Cell Biol.*, **18**, 115–126.

10. Campos,E.I. and Reinberg,D. (2009) Histones: annotating chromatin. *Annu. Rev. Genet.*, **43**, 559–599.

11. Kimura,H. and Cook,P.R. (2001) Kinetics of core histones in living human cells. *J. Cell Biol.*, **153**, 1341–1354.

12. Lever,M.a, Th'ng,J.P., Sun,X. and Hendzel,M.J. (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, **408**, 873–876.

13. Tan,M., Luo,H., Lee,S., Jin,F., Yang,J.S., Montellier,E., Buchou,T., Cheng,Z., Rousseaux,S., Rajagopal,N. *et al.* (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016–1028.

14. Strahl,B.D. and Allis,C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41–45.

15. Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.

16. Bannister,A.J. and Kouzarides,T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.

17. Huang,H., Lin,S., Garcia,B.A. and Zhao,Y. (2015) Quantitative proteomic analysis of histone modifications. *Chem. Rev.*, **115**, 2376–2418.

18. 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.

19. Taverna,S.D., Li,H., Ruthenburg,A.J., Allis,C.D. and Patel,D.J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.*, **14**, 1025–1040.

20. Yun,M., Wu,J., Workman,J.L. and Li,B. (2011) Readers of histone modifications. *Cell Res.*, **21**, 564–578.

21. Musselman,C.A., Lalonde,M.-E., Cˆot'e,J. and Kutateladze,T.G. (2012) Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.*, **19**, 1218–1227.

22. Gao,Z., Zhang,J., Bonasio,R., Strino,F., Sawai,A., Parisi,F., Kluger,Y. and Reinberg,D. (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell*, **45**, 344–356.

23. Kamakaka,R.T. (2005) Histone variants: deviants? *Genes Dev.*, **19**, 295–316.

24. Taguchi,H., Xie,Y., Horikoshi,N., Maehara,K., Harada,A., Nogami,J., Sato,K., Arimura,Y., Osakabe,A., Kujirai,T. *et al.* (2017) Crystal structure and characterization of novel human histone H3 variants, H3.6, H3.7, and H3.8. *Biochemistry*, **56**, 2184–2196.

25. Haynes,K.A. and Silver,P.A. (2009) Eukaryotic systems broaden the scope of synthetic biology. *J. Cell Biol.*, **187**, 589–596.

26. Park,M., Keung,A.J. and Khalil,A.S. (2016) The epigenome: the next substrate for engineering. *Genome Biol.*, **17**, 183.

27. Pick,H., Kilic,S. and Fierz,B. (2014) Engineering chromatin states: chemical and synthetic biology approaches to investigate histone modification function. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1839**, 644–656.

28. He,S., Bauman,D., Davis,J.S., Loyola,A., Nishioka,K., Gronlund,J.L., Reinberg,D., Meng,F., Kelleher,N. and McCafferty,D.G. (2003) Facile synthesis of site-specifically acetylated and methylated histone proteins: Reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12033–12038.

29. Shogren-Knaak,M.A., Fry,C.J. and Peterson,C.L. (2003) A native peptide ligation strategy for deciphering nucleosomal histone modifications. *J. Biol. Chem.*, **278**, 15744 15748.

30. Shogren-Knaak,M. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, **311**, 844–847.

31. 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**, 1003–1012.

32. McGinty,R.K., Kim,J., Chatterjee,C., Roeder,R.G. and Muir,T.W. (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature*, **453**, 812–816.

33. David,Y., Vila-Perell ’o,M., Verma,S. and Muir,T.W. (2015) Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nat. Chem.*, **7**, 394–402.

34. 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**, 153–163.

35. Kallappagoudar,S., Dammer,E.B., Duong,D.M., Seyfried,N.T. and Lucchesi,J.C. (2013) Expression, purification and proteomic analysis of recombinant histone H4 acetylated at lysine 16. *Proteomics*, **13**, 1687–1691.

36. Kim,C.H., Kang,M., Kim,H.J., Chatterjee,A. and Schultz,P.G. (2012) Site-specific incorporation of _-N-crotonyllysine into histones. *Angew. Chem. Int. Ed.*, **51**, 7246–7249.

37. Wilkins,B.J., Hahn,L.E., Heitm· · uller,S., Frauendorf,H., Valerius,O., Braus,G.H. and Neumann,H. (2015) Genetically encoding lysine modifications on histone H4. *ACS Chem. Biol.*, **10**, 939–944.

38. Xiao,H., Xuan,W., Shao,S., Liu,T. and Schultz,P.G. (2015) Genetic incorporation of _-*N*-2-hydroxyisobutyryl-lysine into recombinant histones. *ACS Chem. Biol.*, **10**, 1599–1603.

39. Elasser,S.J., Ernst,R.J., Walker,O.S. and Chin,J.W. (2016) Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat. Methods*, **13**, 158–164.

40. Hansen,K.H., Bracken,A.P., Pasini,D., Dietrich,N., Gehani,S.S., Monrad,A., Rappsilber,J., Lerdrup,M. and Helin,K. (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.*, **10**, 1291–1300.

41. Ma,A.-N., Wang,H., Guo,R., Wang,Y.-X., Li,W., Cui,J., Wang,G., Hoffman,A.R. and Hu,J.-F. (2014) Targeted gene suppression by inducing de novo DNA methylation in the gene promoter. *Epigenetics and Chromatin*, **7**, 20.

42. Bintu,L., Yong,J., Antebi,Y.E., McCue,K., Kazuki,Y., Uno,N., Oshimura,M. and Elowitz,M.B. (2016) Dynamics of epigenetic regulation at the single-cell level. *Science*, **351**, 720–724.

43. Konermann,S., Brigham,M.D., Trevino,A., Hsu,P.D., Heidenreich,M., Le,Cong, Platt,R.J., Scott,D.a, Church,G.M. and Zhang,F. (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, **500**, 472–476.

44. Snowden,A.W., Gregory,P.D., Case,C.C. and Pabo,C.O. (2002) Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Curr. Biol.*, **12**, 2159–2166.

45. Kungulovski,G., Nunna,S., Thomas,M., Zanger,U.M., Reinhardt,R. and Jeltsch,A. (2015) Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenet. Chromatin*, **8**, 12.

46. Mendenhall,E.M.,Williamson,K.E., Reyon,D., Zou,J.Y., Ram,O., Joung,J.K. and Bernstein,B.E. (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.*, **31**, 1133–1136.

47. Kearns,N.a, Pham,H., Tabak,B., Genga,R.M., Silverstein,N.J., Garber,M. and Maehr,R. (2015) Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat. Methods*, **12**, 401–403.

48. Kwaks,T.H.J., Sewalt,R.G.A.B., van Blokland,R., Siersma,T.J., Kasiem,M., Kelder,A. and Otte,A.P. (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *J. Biotechnol.*, **115**, 35 46.

49. Hilton,I.B., D'Ippolito,A.M., Vockley,C.M., Thakore,P.I., Crawford,G.E., Reddy,T.E. and Gersbach,C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.*, **33**, 510–517.

50. Cano-Rodriguez,D. and Rots,M.G. (2016) Epigenetic editing: on the verge of reprogramming gene expression at will. *Curr. Genet. Med. Rep.*, **4**, 170–179.

51. Suganuma,T. and Workman,J.L. (2011) Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.*, **80**, 473–499.

52. Suganuma,T. and Workman,J.L. (2008) Crosstalk among histone modifications. *Cell*, **135**, 604–607.

53. Smith,E. and Shilatifard,A. (2010) The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol. Cell*, **40**, 689–701.

54. Zhang,T., Cooper,S. and Brockdorff,N. (2015) The interplay of histone modifications—writers that read. *EMBO Rep.*, **16**, 1467–1481.

55. Bannister,A.J., Zegerman,P., Partridge,J.F., Miska,E.a, Thomas,J.O., Allshire,R.C. and Kouzarides,T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, **410**, 120–124.

56. VanDemark,A.P., Kasten,M.M., Ferris,E., Heroux,A., Hill,C.P. and Cairns,B.R. (2007) Autoregulation of the Rsc4 tandem bromodomain by Gcn5 acetylation. *Mol. Cell*, **27**, 817–828.

57. McGinty,R.K., Kim,J., Chatterjee,C., Roeder,R.G. and Muir,T.W. (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature*, **453**, 812–816.

58. Li,B., Gogol,M., Carey,M., Lee,D., Seidel,C. and Workman,J.L. (2007) Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science*, **316**, 1050–1054.

59. Schmitges,F.W., Prusty,A.B., Faty,M., Stutzer,A., Lingaraju,G.M., Aiwazian,J., Sack,R., Hess,D., Li,L., Zhou,S. *et al.* (2011) Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell*, **42**, 330–341.

60. Gatchalian,J., F· · utterer,A., Rothbart,S.B., Tong,Q., Rincon-Arano,H., Sanchez de Diego,A., Groudine,M., Strahl,B.D., Martınez-A,C., Van Wely,K.H.M. *et al.* (2013) Dido3 PHD modulates cell differentiation and division. *Cell Rep.*, **4**, 148–158.

61. Pengue,G., Calabr'o,V., Bartoli,P.C., Pagliuca,A. and Lania,L. (1994) Repression of transcriptional activity at a distance by the evolutionary conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res.*, **22**, 2908–2914.

62. Pengue,G. and Lania,L. (1996) Kr· · uppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 1015–1020.

63. Thakore,P.I., D'Ippolito,A.M., Song,L., Safi,A., Shivakumar,N.K., Kabadi,A.M., Reddy,T.E., Crawford,G.E. and Gersbach,C.A. (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods*, **12**, 1143–1149.

64. Amabile,A., Migliara,A., Capasso,P., Biffi,M., Cittaro,D., Naldini,L. and Lombardo,A. (2016) Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell*, **167**, 219–232.

65. Lupo,A., Cesaro,E., Montano,G., Zurlo,D., Izzo,P. and Costanzo,P. (2013) KRAB-zinc finger proteins: a repressor family displaying multiple biological functions. *Curr. Genomics*, **14**, 268–278.

66. Jacobs,S.A. (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*, **295**, 2080–2083.

67. Nielsen,P.R., Nietlispach,D.,Mott,H.R., Callaghan,J., Bannister,A., Kouzarides,T., Murzin,A.G., Murzina,N.V. and Laue,E.D. (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature*, **416**, 103–107.

68. Fischle,W., Wang,Y., Jacobs,S.A., Kim,Y., Allis,C.D. and Khorasanizadeh,S. (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes Dev.*, **17**, 1870–1881.

69. Min,J. (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.*, **17**, 1823–1828.

70. Bernstein,E., Duncan,E.M., Masui,O., Gil,J., Heard,E. and Allis,C.D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.*, **26**, 2560–2569.

71. Eissenberg,J.C. (2012) Structural biology of the chromodomain: Form and function. *Gene*, **496**, 69–78.

72. Ren,C., Morohashi,K., Plotnikov,A.N., Jakoncic,J., Smith,S.G., Li,J., Zeng,L., Rodriguez,Y., Stojanoff,V., Walsh,M. *et al.* (2015) Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.*, **22**, 161–168.

73. Black,J.C., Van Rechem,C. and Whetstine,J.R. (2012) Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell*, **48**, 491–507.

74. Kaustov,L., Ouyang,H., Amaya,M., Lemak,A., Nady,N., Duan,S., Wasney,G.A., Li,Z., Vedadi,M., Schapira,M. *et al.* (2011) Recognition and specificity determinants of the human Cbx chromodomains. *J. Biol. Chem.*, **286**, 521–529.

75. Platero,J.S., Hartnett,T. and Eissenberg,J.C. (1995) Functional analysis of the chromo domain of HP1. *EMBO J.*, **14**, 3977–3986.

76. Messmer,S., Franke,A. and Paro,R. (1992) Analysis of the functional role of the polycomb chrom domain in Drosophila melanogaster. *Genes Dev.*, **6**, 1241–1254.

77. Haynes,K.A. and Silver,P.A. (2011) Synthetic reversal of epigenetic silencing. *J. Biol. Chem.*, **286**, 27176–27182.

78. Nyer,D.B., Daer,R.M., Vargas,D., Hom,C. and Haynes,K.A. (2017) Regulation of cancer epigenomes with a histone-binding synthetic transcription factor. *npj Genomic Med.*, **2**, 1.

79. Owen,D.J. (2000) The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J.*, **19**, 6141–6149.

80. Filippakopoulos,P., Picaud,S., Mangos,M., Keates,T., Lambert,J.-P., Barsyte-Lovejoy,D., Felletar,I., Volkmer,R., M· · uller,S., Pawson,T. *et al.* (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, **149**, 214–231.

81. Fujisawa,T. and Filippakopoulos,P. (2017) Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.*, **18**, 246–262.

82. Huang,H., Zhang,J., Shen,W., Wang,X., Wu,J., Wu,J. and Shi,Y. (2007) Solution structure of the second bromodomain of Brd2 and its specific interaction with acetylated histone tails. *BMC Struct. Biol.*, **7**, 57.

83. Umehara,T., Nakamura,Y., Jang,M.K., Nakano,K., Tanaka,A., Ozato,K., Padmanabhan,B. and Yokoyama,S. (2010) Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J. Biol. Chem.*, **285**, 7610–7618.

84. Flynn,E.M., Huang,O.W., Poy,F., Oppikofer,M., Bellon,S.F., Tang,Y. and Cochran,A.G. (2015) A subset of human bromodomains recognizes butyryllysine and crotonyllysine histone peptide modifications. *Structure*, **23**, 1801–1814.

85. Wang,R. and You,J. (2015) Mechanistic analysis of the role of bromodomain-containing protein 4 (BRD4) in BRD4-NUT oncoprotein-induced transcriptional activation. *J. Biol. Chem.*, **290**, 2744–2758.

86. Dekker,F.J., van den Bosch,T. and Martin,N.I. (2014) Small molecule inhibitors of histone acetyltransferases and deacetylases are potential drugs for inflammatory diseases. *Drug Discov. Today*, **19**, 654–660.

87. Niedzialkowska,E., Wang,F., Porebski,P.J., Minor,W., Higgins,J.M.G. and Stukenberg,P.T. (2012) Molecular basis for phosphospecific recognition of histone H3 tails by Survivin paralogues at inner centromeres. *Mol. Biol. Cell*, **23**, 1457–1466.

88. Rossetto,D., Avvakumov,N. and Cˆot'e,J. (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, **7**, 1098–1108.

89. Kelly,A.E., Ghenoiu,C., Xue,J.Z., Zierhut,C., Kimura,H. and Funabiki,H. (2010) Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase aurora B. *Science*, **330**, 235–239.

90. Jeyaprakash,A.A., Basquin,C., Jayachandran,U. and Conti,E. (2011) Structural basis for the recognition of phosphorylated histone H3 by the surviving subunit of the chromosomal passenger complex. *Structure*, **19**, 1625–1634.

91. Wang,F., Dai,J., Daum,J.R., Niedzialkowska,E., Banerjee,B., Stukenberg,P.T., Gorbsky,G.J. and Higgins,J.M.G. (2010) Histone H3 Thr-3 phosphorylation by haspin positions aurora B at centromeres in mitosis. *Science*, **330**, 231–235.

92. Sanchez,R. and Zhou,M.-M. (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.*, **36**, 364–372.

93. Musselman,C.A. and Kutateladze,T.G. (2011) Handpicking

epigenetic marks with PHD fingers. *Nucleic Acids Res.*, **39**, 9061–9071.

94. 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. *et al.* (2011) Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell*, **145**, 692–706.

95. Pena,P. V., Davrazou,F., Shi,X., Walter,K.L., Verkhusha,V. V, Gozani,O., Zhao,R. and Kutateladze,T.G. (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature*, **442**, 100–103.

96. van Ingen,H., van Schaik,F.M.A., Wienk,H., Ballering,J., Rehmann,H., Dechesne,A.C., Kruijzer,J.A.W., Liskamp,R.M.J., Timmers,H.T.M. and Boelens,R. (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure*, **16**, 1245–1256.

97. Li,H., Ilin,S., Wang,W., 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.

98. Wysocka,J., Swigut,T., Xiao,H., Milne,T.A., Kwon,S.Y., Landry,J., Kauer,M., Tackett,A.J., Chait,B.T., Badenhorst,P. *et al.* (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, **442**, 86–90.

99. Kim,S., Natesan,S., Cornilescu,G., Carlson,S., Tonelli,M., McClurg,U.L., Binda,O., Robson,C.N., Markley,J.L., Balaz,S. *et al.* (2016) Mechanism of histone H3K4me3 recognition by the plant homeodomain of inhibitor of growth 3. *J. Biol. Chem.*, **291**, 18326–18341.

100. Wang,G.G., Song,J., Wang,Z., Dormann,H.L., Casadio,F., Li,H., Luo,J., Patel,D.J. and Allis,C.D. (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature*, **459**, 847–851.

101. Li,H., Fischle,W., Wang,W., Duncan,E.M., Liang,L., Murakami-Ishibe,S., Allis,C.D. and Patel,D.J. (2007) Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell*, **28**, 677–691.

102. Zeng,L., Zhang,Q., Li,S., Plotnikov,A.N., Walsh,M.J. and Zhou,M. (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*, **466**, 258–262.

103. Futterer,a, Raya,A., Llorente,M., Izpisu'a-Belmonte,J.C., de la Pompa,J.L., Klatt,P. and Mart'ınez-A,C. (2012) Ablation of Dido3 compromises lineage

commitment of stem cells in vitro and during early embryonic development. *Cell Death Differ.*, **19**, 132–143.

104. Gelato,K.A., Tauber,M., Ong,M.S., Winter,S., Hiragami-Hamada,K., Sindlinger,J., Lemak,A., Bultsma,Y., Houliston,S., Schwarzer,D. *et al.* (2014) Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. *Mol. Cell*, **54**, 905–919.

105. Su,X., Zhu,G., Ding,X., Lee,S.Y., Dou,Y., Zhu,B., Wu,W. and Li,H. (2014) Molecular basis underlying histone H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. *Genes Dev.*, **28**, 622–636.

106. Hiragami-Hamada,K., Soeroes,S., Nikolov,M., Wilkins,B., Kreuz,S., Chen,C., De La Rosa-Vel'azquez,I.A., Zenn,H.M., Kost,N., Pohl,W. *et al.* (2016) Dynamic and flexible H3K9me3 bridging via HP1_ dimerization establishes a plastic state of condensed chromatin. *Nat. Commun.*, **7**, 11310.

107. Norwood,L.E.,Moss,T.J., Margaryan,N. V., Cook,S.L., Wright,L., Seftor,E.A., Hendrix,M.J.C., Kirschmann,D.A. and Wallrath,L.L. (2006) A requirement for dimerization of HP1Hs in suppression of breast cancer invasion. *J. Biol. Chem.*, **281**, 18668–18676.

108. Trembecka-Lucas,D.O., Trembecka-Lucas,D.O., Szczurek,A.T., Szczurek,A.T., Dobrucki,J.W. and Dobrucki,J.W. (2013) Dynamics of the HP1_-PCNA-containing complexes in DNA replication and repair. *Nucleus*, **4**, 74–82.

109. Jacobson,R.H., Ladurner,A.G., King,D.S. and Tjian,R. (2000) Structure and function of a human TAF II 250 double bromodomain module. *Science*, **288**, 1422–1425.

110. McDaniel,S.L., Fligor,J.E., Ruan,C., Cui,H., Bridgers,J.B., DiFiore,J.V., Guo,A.H., Li,B. and Strahl,B.D. (2016) Combinatorial histone readout by the dual plant homeodomain (PHD) fingers of Rco1 mediates Rpd3S chromatin recruitment and the maintenance of transcriptional fidelity. *J. Biol. Chem.*, **291**, 14796–14802.

111. Huh,J., Wu,J., Lee,C., Yun,M., Gilada,D., Brautigam,C.A. and Li,B. (2012) Multivalent di-nucleosome recognition enables the Rpd3S histone deacetylase complex to tolerate decreased H3K36 methylation levels. *EMBO J.*, **31**, 3564–3574.

112. Lee,C.-H., Wu,J. and Li,B. (2013) Chromatin remodelers fine-tune H3K36me-directed deacetylation of neighbor nucleosomes by Rpd3S. *Mol. Cell*, **52**, 255–263.

113. Tabet,S., Douglas,S.F., Daze,K.D., Garnett,G.A.E., Allen,K.J.H., Abrioux,E.M.M., Quon,T.T.H.,Wulff,J.E. and Hof,F. (2013) Synthetic trimethyllysine receptors that bind histone 3, trimethyllysine 27 (H3K27me3) and disrupt its interaction with the epigenetic reader protein CBX7. *Bioorg. Med. Chem.*, **21**, 7004–7010.

114. Simhadri,C., Daze,K.D., Douglas,S.F., Quon,T.T.H., Dev,A., Gignac,M.C., Peng,F., Heller,M., Boulanger,M.J., Wulff,J.E. *et al.* (2014) Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J. Med. Chem.*, **57**, 2874–2883.

115. Stuckey,J.I., Dickson,B.M., Cheng,N., Liu,Y., Norris,J.L., Cholensky,S.H., Tempel,W., Qin,S., Huber,K.G., Sagum,C. *et al.* (2016) A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nat. Chem. Biol.*, **12**, 180–187.

116. Ren,C., Smith,S.G., Yap,K., Li,S., Li,J., Mezei,M., Rodriguez,Y., Vincek,A., Aguilo,F., Walsh,M.J. *et al.* (2016) Structure-guided discovery of selective antagonists for the chromodomain of polycomb repressive protein CBX7. *ACS Med. Chem. Lett.*, **7**, 601–605.

117. Milosevich,N., Gignac,M.C., McFarlane,J., Simhadri,C., Horvath,S., Daze,K.D., Croft,C.S., Dheri,A., Quon,T.T.H., Douglas,S.F. *et al.* (2016) Selective inhibition of CBX6: a methyllysine reader protein in the polycomb family. *ACS Med. Chem. Lett.*, **7**, 139–144.

118. Whitehead,L., Dobler,M.R., Radetich,B., Zhu,Y., Atadja,P.W., Claiborne,T., Grob,J.E., McRiner,A., Pancost,M.R., Patnaik,A. *et al.* (2011) Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. *Bioorg. Med. Chem.*, **19**, 4626–4634.

119. Rahman,S., Sowa,M.E., Ottinger,M., Smith,J.A., Shi,Y., Harper,J.W. and Howley,P.M. (2011) The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol. Cell. Biol.*, **31**, 2641–2652.

120. Filippakopoulos,P., Qi,J., Picaud,S., Shen,Y., Smith,W.B., Fedorov,O., Morse,E.M., Keates,T., Hickman,T.T., Felletar,I. *et al.* (2010) Selective inhibition of BET bromodomains. *Nature*, **468**, 1067–1073.

121. Nicodeme,E., Jeffrey,K.L., Schaefer,U., Beinke,S., Dewell,S., Chung,C., Chandwani,R., Marazzi,I., Wilson,P., Coste,H. *et al.* (2010) Suppression of inflammation by a synthetic histone mimic. *Nature*, **468**, 1119–1123.

122. James,L.I., Barsyte-Lovejoy,D., Zhong,N., Krichevsky,L., Korboukh,V.K., Herold,J.M., MacNevin,C.J., Norris,J.L., Sagum,C.A., Tempel,W. *et al.* (2013) Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat. Chem. Biol.*, **9**, 184–191.

123. Verma,S.K., Tian,X., LaFrance,L. V., Duquenne,C., Suarez,D.P., Newlander,K.A., Romeril,S.P., Burgess,J.L., Grant,S.W., Brackley,J.A. *et al.* (2012) Identification of potent, selective, cell-active inhibitors of the histone lysine methyltransferase EZH2. *ACS Med. Chem. Lett.*, **3**, 1091–1096.

124. Konze,K.D., Ma,A., Li,F., Barsyte-Lovejoy,D., Parton,T., MacNevin,C.J., Liu,F., Gao,C., Huang,X.-P., Kuznetsova,E. *et al.* (2013) An orally bioavailable chemical probe of the lysine methyltransferases EZH2 and EZH1. *ACS Chem. Biol.*, **8**, 1324–1334.

125. Horton,J.R., Liu,X., Gale,M., Wu,L., Shanks,J.R., Zhang,X., Webber,P.J., Bell,J.S.K., Kales,S.C., Mott,B.T. *et al.* (2016) Structural basis for KDM5A histone lysine demethylase inhibition by diverse compounds. *Cell Chem. Biol.*, **23**, 769–781.

126. Xu,W.S., Parmigiani,R.B. and Marks,P.A. (2007) Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene*, **26**, 5541–5552.

127. Kim,H. and Bae,S. (2011) Histone deacetylase inhibitors: molecular mechanisms of action and clinical trials as anti-cancer drugs. *Am. J. Transl. Res.*, **3**, 166–179.

128. 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.

129. Falkenberg,K.J. and Johnstone,R.W. (2014) Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat. Rev. Drug Discov.*, **13**, 673–691.

130. Oike,T., Ogiwara,H., Amornwichet,N., Nakano,T. and Kohno,T. (2014) Chromatin-regulating proteins as targets for cancer therapy. *J. Radiat. Res.*, **55**, 613–628.

131. West,A.C. and Johnstone,R.W. (2014) New and emerging HDAC inhibitors for c ancer treatment. *J Clin Invest.*, **124**, 30–39.

132. Delmore,J.E., Issa,G.C., Lemieux,M.E., Rahl,P.B., Shi,J., Jacobs,H.M., Kastritis,E., Gilpatrick,T., Paranal,R.M., Qi,J. *et al.* (2011) BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*, **146**, 904–917.

133. Mertz,J.A., Conery,A.R., Bryant,B.M., Sandy,P., Balasubramanian,S.,Mele,D.A., Bergeron,L. and Sims,R.J. (2011) Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16669–16674.

134. Cheung,K., Lu,G., Sharma,R., Vincek,A., Zhang,R., Plotnikov,A.N., Zhang,F., Zhang,Q., Ju,Y., Hu,Y. *et al.* (2017) BET N-terminal bromodomain inhibition selectively blocks Th17 cell differentiation and ameliorates colitis in mice. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 2952–2957.

135. Kassis,J.A. and Brown,J.L. (2013) Polycomb group response elements in Drosophila and vertebrates. In: *Advances in Genetics*. Elsevier Inc., Vol. **81**, pp. 83–118.

136. Cavalli,G. and Paro,R. (1998) The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell*, **93**, 505–518.

137. Cavalli,G. and Paro,R. (1999) Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science*, **286**, 955–958.

138. Dejardin,J. and Cavalli,G. (2004) Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *EMBO J.*, **23**, 857–868.

139. Muller,J. and Kassis,J.A. (2006) Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.*, **16**, 476–484.

140. Bauer,M., Trupke,J. and Ringrose,L. (2016) The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*, **125**, 471–496.

141. Woo,C.J., Kharchenko,P. V, Daheron,L., Park,P.J. and Kingston,R.E. (2010) A region of the human HOXD cluster that confers Polycomb-group responsiveness. *Cell*, **140**, 99–110.

142. Woo,C.J., Kharchenko,P. V, Daheron,L., Park,P.J. and Kingston,R.E. (2013) Variable requirements for DNA-binding proteins at Polycomb-dependent repressive regions in human HOX clusters. *Mol. Cell. Biol.*, **33**, 3274–3285.

143. Cuddapah,S., Roh,T.-Y., Cui,K., Jose,C.C., Fuller,M.T., Zhao,K. and Chen,X. (2012) A novel human Polycomb binding site acts as a functional Polycomb response element in Drosophila. *PLoS One*, **7**, e36365.

144. Sing,A., Pannell,D., Karaiskakis,A., Sturgeon,K., Djabali,M., Ellis,J., Lipshitz,H.D. and Cordes,S.P. (2009) A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell*, **138**, 885–897.

145. Vasanthi,D., Nagabhushan,A.,Matharu,N.K. and Mishra,R.K. (2013) A functionally conserved Polycomb response element from mouse HoxD complex responds to heterochromatin factors. *Sci. Rep.*, **3**, 3011.

146. Ringrose,L., Rehmsmeier,M., Dura,J. and Paro,R. (2003) Genome-wide prediction of polycomb/trithorax response elements in Drosophila melanogaster. *Dev. Cell*, **5**, 759–771.

147. Hauenschild,A., Ringrose,L., Altmutter,C., Paro,R. and Rehmsmeier,M. (2008) Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. *PLoS Biol.*, **6**, 2130–2143. 148. Gilman,J. and Love,J. (2016) Synthetic promoter design for new microbial chassis. *Biochem. Soc. Trans.*, **44**, 731–737.

149. Qin,J.Y., Zhang,L., Clift,K.L., Hulur,I., Xiang,A.P., Ren,B.-Z. and Lahn,B.T. (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One*, **5**, e10611.

150. Keung,A.J., Bashor,C.J., Kiriakov,S., Collins,J.J. and Khalil,A.S. (2014) Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell*, **158**, 110–120.

151. Engreitz,J.M., Ollikainen,N. and Guttman,M. (2016) Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.*, **17**, 756–770.

152. Somarowthu,S., Legiewicz,M., Chill on,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell*, **58**, 353–361.

153. Tsai,M.-C., Manor,O., Wan,Y., Mosammaparast,N., Wang,J.K., Lan,F., Shi,Y., Segal,E. and Chang,H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689–693.

154. Wu,L., Murat,P., Matak-Vinkovic,D., Murrell,A. and Balasubramanian,S. (2013) Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. *Biochemistry*, **52**, 9519–9527.

155. Portoso,M., Ragazzini,R., Brencic, Z., Moiani,A., Michaud,A., Vassilev,I., Wassef,M., Servant,N., Sargueil,B. and Margueron,R. (2017) PRC2 is dispensable for HOTAIR -mediated transcriptional repression. *EMBO J.*, **36**, 981–994.

156. Hacisuleyman,E., Shukla,C.J., Weiner,C.L. and Rinn,J.L. (2016) Function and evolution of local repeats in the Firre locus. *Nat. Commun.*, **7**, 11021.

157. Hasegawa,Y., Brockdorff,N., Kawano,S., Tsutui,K., Tsutui,K. and Nakagawa,S. (2010) The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell*, **19**, 469–476.

158. Hendrickson,D.G., Kelley,D.R., Tenen,D., Bernstein,B. and Rinn,J.L. (2016) Widespread RNA binding by chromatin-associated proteins. *Genome Biol.*, **17**, 28.

159. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.

160. Wang,X., Wei,X., Pang,Q. and Yi,F. (2012) Histone deacetylases and their inhibitors: molecular mechanisms and therapeutic implications in diabetes mellitus. *Acta Pharm. Sin. B*, **2**, 387–395.

161. Haumaitre,C., Lenoir,O. and Scharfmann,R. (2008) Histone deacetylase inhibitors modify pancreatic cell fate determination and amplify endocrine progenitors. *Mol. Cell. Biol.*, **28**, 6373–6383.

162. Bramswig,N.C., Everett,L.J., Schug,J., Dorrell,C., Liu,C., Luo,Y., Streeter,P.R., Naji,A., Grompe,M. and Kaestner,K.H. (2013) Epigenomic plasticity enables human pancreatic _ to _ cell reprogramming. *J. Clin. Invest.*, **123**, 1275–1284.

163. Khan,S. and Jena,G.B. (2014) Protective role of sodium butyrate, a HDAC inhibitor on beta-cell proliferation, function and glucose homeostasis through modulation of p38/ERK MAPK and apoptotic pathways: Study in juvenile diabetic rat. *Chem. Biol. Interact.*, **213**, 1–12.

164. Khan,S. and Jena,G. (2014) Sodium butyrate, a HDAC inhibitor ameliorates eNOS, iNOS and TGF-_1-induced fibrogenesis, apoptosis and DNA damage in the kidney of juvenile diabetic rats. *Food Chem. Toxicol.*, **73**, 127–139.

165. Berman,H.M.,Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H. and Shindyalov,I.N. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

166. Andrews,F.H., Strahl,B.D. and Kutateladze,T.G. (2016) Insights into newly discovered marks and readers of epigenetic information. *Nat. Chem. Biol.*, **12**, 662–668.

167. Wang,G.G., Song,J., Wang,Z., Dormann,H.L., Casadio,F., Li,H., Luo,J., Patel,D.J. and Allis,C.D. (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature*, **459**, 847–851.

168. Houbaviy,H.B., Usheva,A., Shenk,T. and Burley,S.K. (1996) Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 13577–13582.

169. PDB ID: 4XBF, Luka,Z., Loukachevitch,L.V., Martin,W.J., Wagner,C. and Reiter,N.J. (2016) Structure of LSD1:CoREST in complex with ssRNA.

CHAPTER 2

DESIGN, CONSTRUCTION, AND VALIDATION OF HISTONE-BINDING EFFECTORS *IN VITRO* AND IN CELLS

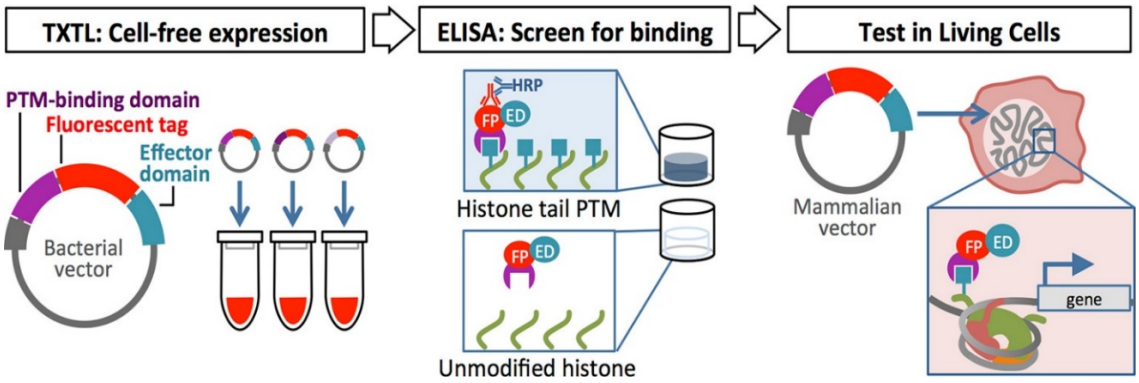


Figure 2.0 Abstract Figure.

Introduction

Chromatin engineering is a burgeoning field with applications in disease treatment (1), tissue engineering (2), and development of cell lines for industrial production of biologicals (3). Recently, proteins that bind specific histone posttranslational modifications (PTMs) have been used as probes for histone PTM states in cells and to target gene regulators to sites based on histone PTM status (4, 5). Protein motifs that interact with specific acetylated and methylated residues on histone H3 have been used to design fusion fluorescent proteins to label genomic regions that contain high levels of H3K14ac (6), H3K9me3 (7), and H3K4me3 and H3K27me3 (8). The wealth of basic research that has supported the design and application of recombinant, chromatin-based proteins is discussed in detail in reviews from our group (9) and others (10).

Here, we describe a workflow to measure binding of recombinant histone-binding proteins in vitro and to validate their activity in live cells using techniques that do not

require highly specialized equipment or high-yield protein preparations. Current approaches to validate the function of histone PTM binding domains (HBDs) include fluorescence polarization (FP) (11) isothermal calorimetry (ITC) (12), surface plasmon resonance (SPR) (13), and peptide arrays (14). The primary limitation of these approaches is the need for high yields of purified protein, which limits the number and variety of proteins that can be assayed efficiently. Recently, we demonstrated the rapid prototyping of novel multivalent histone-binding transcriptional activators built from a human orthologue called CBX8 (14). The results showed that small volumes (<12 μL) of cell-free transcription−translation lysates (TXTL) can be used to produce sufficient RFP-tagged histone binding fusion proteins to generate significant binding signal over background in an enzyme-linked immunosorbent assay (ELISA). Here as a proof of concept, we screened a small library of additional TXTL-expressed variants for H3K27me3 avidity. Open reading frames that encoded histone-binding candidates were then cloned into a mammalian expression vector and tested for gene-regulating activity in HEK293 cells. The cells used here include an ectopic heterochromatic site where the target histone PTM is located near the promoter of a reporter transgene (luciferase). Our report includes guidance for identifying other useful target PTMs and genomic loci.

Materials and Methods

Constructs for Histone-Binding Protein Fusions. BioBrick assembly (15) was used to build open reading frames that encoded full-length fusion proteins in the general purpose vector V0120. Full-length, annotated sequences are available at the Haynes lab Benchling Web site (https://benchling.com/ hayneslab/f_/rmSYkAAU-synthetic-chromatin-actuators-2-0/).

DNA Constructs for Cell-Free Transcription−Translation (TXTL) Expression. Constructs were assembled via directional cloning of PCR-amplified and digested fusion coding regions (from vector V0120) into pET28 as previously described (14) with the following modifications: EcoRI and HindIII restriction sites were used instead of BamHI and XhoI sites. All histone PTM-binding domains (HBDs) were separated by a [GGGGS] x4 linker. Human (Pcα) = CBX8 (amino acids 1−62). Fish (Pcβ) = CBX2 or pc1 (amino acids 2−63). Fly (Pcγ) = Pc (amino acids 16−77) (5). Plasmids were cloned in compatible strains: pET28 fusion-expressing constructs in *Escherichia coli* BL21 and σ70-T7 RNA polymerase plasmid in *E. coli* KL740.

Enzyme-Linked Immunosorbent Assays (ELISAs). Capture antibodies included mouse anti-6-histidine (Abcam 18184) and chicken anti-mCherry (Novus biologicals NBP2- 25158). Biotinylated peptides were H3K27me3 (21−44) (Anaspec AS-64367-025) and H3 (21−44) (Anaspec AS- 64440-025). Primary antibodies were chicken anti-mCherry (Novus biologicals NBP2-25158) and mouse anti-6-histidine (Abcam 18184). Secondary antibodies were rabbit anti-chicken HRP (Gentel, 0.5 mg/mL, RCYHRP) and goat anti-mouse HRP (KPL 074-1806). Data shown in this report were generated using a Synergy H1 Hybrid Reader with the following key settings: standard 96-well plate option; detection, absorbance; read type, end point; read speed, normal. We have obtained very similar A450 values [absorbance units (a.u.)] using two different plate reader models (Synergy H1 Hybrid Reader and PerkinElmer EnVision 2104 Multilabel Reader). For each Pc fusion, the A450 value for each H3K27me3 well (four replicates) was divided by the mean A450 value of four H3 (unmodified histone) control wells. “Histone peptide capture: relative binding” is the mean of the four normalized values.

Nonspecific binding was assessed by using a fusion protein that contained no HBD (PcΔ(14)). We used two methods to determine histone peptide capture as percent fusion protein input, where either the mean six-His capture A450 value or the TXTL RFP end point signal (533−610, a.u.) was used to represent input.

DNA Constructs for Mammalian Cell Expression. Constructs were assembled via nondirectional cloning of fusion-coding regions (from V0120) into vector MV10 as previously described (14). Colony PCR was used to screen for forward inserts: 1× GoTaq green master mix (Promega M7122), 0.4 μM forward primer 5′-caccatcgtggaacagtacg, and 0.4 μM reverse primer 5′-gcaactagaaggcacagtcg in a final volume of 25 μL. PCR was performed as follows: 95 °C for 2 min; 25 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min; 72 °C for 5 min; and held at 4 °C. Candidate clones were grown in 3 mL of liquid LB and 100 μg/mL ampicillin. Restriction digestion followed by electrophoresis (1% agarose in TAE) was used to analyze purified plasmid DNA (Sigma PLN350-1KT): 500−1000 ng of plasmid DNA, 1 μL each of FastDigest enzymes XbaI and PstI, and 1× FastDigest buffer (Thermo Fisher FD0684, FD0614, and B72).

Cell Culture and Transfection. HEK293 Gal4-EED/luc cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% tetracycline-free fetal bovine serum and 1% penicillin and streptomycin (pen/strep) at 37 °C in a humidified CO2 incubator. Silencing of the reporter gene (Tk-luciferase) was induced by supplementing the medium with 1 μg/mL dox for 96 h. For wash-out of doxycycline (to allow depletion of Gal4-EED), growth medium was removed and replaced with dox-minus medium. Prior to transfection, doxtreated cells were plated in 12-well culture dishes at 40% confluency (∼1.0×10^5 cells/well) in 2 mL of pen/strep-free growth

medium. Transient transfections were performed by adding 300 μL of DNA/Lipofectamine complexes to each well: 1 μg of plasmid DNA or ddH_2O for mock transfections (10 μL), 5 μL of Lipofectamine LTX (Invitrogen 15338100), or 285 μL of OptiMEM (Gibco 31985062). Plates were spun at 100 g for 5 min to increase the transfection efficiency and then incubated at 37 °C in a humidified CO2 incubator.

RFP, Hoechst, and Luciferase Plate Reader Assay Data Analysis. For each channel, luciferase (chemiluminescence), RFP, and Hoechst 33342, the average signal from the 1× PBS wells was subtracted from the value for every sample well. Next, the average background-subtracted luciferase value for untransfected (UT) cells was subtracted from each experimental luciferase value. UT-subtracted luciferase values were scaled by multiplying each value by [(mean UT Hoechst)/(sample Hoechst)] and then by [(mean Pcαα RFP)/(sample RFP)].

Statistical Analyses. Reported *p* values were determined by a Student's paired t test, with a two-tailed distribution.

Overview of Histone PTM-Binding and Gene Regulation Assays

Binding of TXTL-expressed, fluorescently tagged fusion proteins with target histone peptides is assessed by a user-friendly ELISA. Previous work has shown that bacterially expressed, eukaryotic histone PTM-binding domains (HBDs) have specific avidity for their ligands (reviewed in ref 9). Therefore, TXTL in bacterial lysates is suitable for producing ligand-binding HBDs and allows researchers to circumvent cell transformation or transfection. Ligands (modified histone tail peptides conjugated with biotin) are immobilized in neutravidin-coated microwells, and TXTLexpressed fusions are incubated in the wells to allow binding. Next, candidate regulators are transferred into

mammalian expression vectors to determine gene regulation activity in live cells, where the fusion proteins are expected to interact with whole nucleosomes. Excellent work reported by others has utilized single nucleosomes and nucleosome arrays as templates to study histone PTM-binding proteins (16−19). Because our workflow aims to identify proteins that interact with histone PTMs and activate the transcription of target genes, we use chromatin in cellulo rather than reconstituted nucleosomes or nucleosome arrays. However, we encourage researchers to perform experiments using whole nucleosomes to, for instance, determine the impact of PTM spacing over the nucleosome on the kinetics of binding.

Design and Construction of Histone PTM-Binding Fusion-Expressing Plasmids

In this report, we focus on PTMs within the unfolded N-terminal tail domain of histone H3. These PTMs represent some of the most well- characterized modifications in regard to the proteins that recognize these marks and their impact on gene regulation (9) Our group is interested in designing and optimizing histone binding transcriptional activators that recognize elevated levels of H3K27me3 in cancer cell epigenomes and induce activation of repressed anticancer loci. The procedure described herein uses a library of fusion proteins built for this purpose as an example (Figure 2.1A). The fusions contain conserved, H3K27me3-binding polycomb chromodomains (PCD) from human (CBX8), fish (Danio rerio CBX2, pc1), and/or fly (Drosophila melanogaster CBX2, pc1) orthologues. These PCD orthologues contain 62 amino acids, including three conserved residues that make up an aromatic pocket that binds H3K27me3 (20-21). The 59 remaining residues show a lower level of conservation; reports suggest that some of these residues may contribute to stabilizing the interaction of PCD with the histone tail (20-21).

In our previous work, we directly compared the human, fly, and fish PCD orthologues by testing each as a histone-binding module in a fusion activator protein (5). The fusion protein that contained a single human CBX8 PCD activated target genes more effectively than the fish or fly homologues. Unknown differences in protein processing, such as folding and degradation, prevented a full understanding of the differences in fusion protein activity. In our current work, we investigated the three PCD orthologues to determine H3K27me3 binding under controlled conditions in vitro and to compare the results with gene regulation in live cells.

Generally, we recommend keeping the N- or C-terminal position of the histone PTM-binding domain (HBD) in the fusion protein consistent with its native context (e.g., the H3K27me3-binding PCD is an N-terminal motif) (Figure 2.1A). In our hands, a linker peptide is apparently unnecessary at the HBD−fluorophore junction in fusion proteins that show histone-binding activity. However, we do recommend the inclusion of linker peptides between tandem HBDs. Recent work has shown that combinatorial PTM-binding motifs increase the overall avidity of the fusion for its target (14, 22-23), or allow for specific co-recognition of two distinct histone PTMs within a single nucleosome (8). Pairs of H3 as well as H4 nucleosome tails are expected to be oriented in cis so that they protrude away from the nucleosome in the same general direction (see Protein Data Bank entry 1AOI (24)). These histone tails are flexible enough to come into the proximity of each other. Therefore, a flexible (GS repeat) or rigid (α helix) peptide of 20 amino acids is a good starting point for HBD− HBD linker design. A three-dimensional (3D) structure based model of this format can be found in our previous work (14). It is critical to include in the fusion protein a fluorophore with efficient folding properties, such as the

monomeric Cherry (mCherry) we have used in our work. The tag enables protein detection after expression in TXTL or in cultured mammalian cells. The transcription-regulating domain (activator or repressor) should be included at the in vitro testing stage so that any effects of whole protein folding on binding with the target ligand can be observed. A standard directional cloning approach (e.g., BioBrick assembly (15)) or scarless assembly (e.g., Golden Gate (25)) is suitable for construction of the fusion protein-expressing plasmid DNA (Figure 2.1B). Two plasmids, one encoding a fusion protein and one expressing T7 RNA polymerase under the control of a constitutive σ70 promoter, are added to the TXTL reagent for the expression reaction.

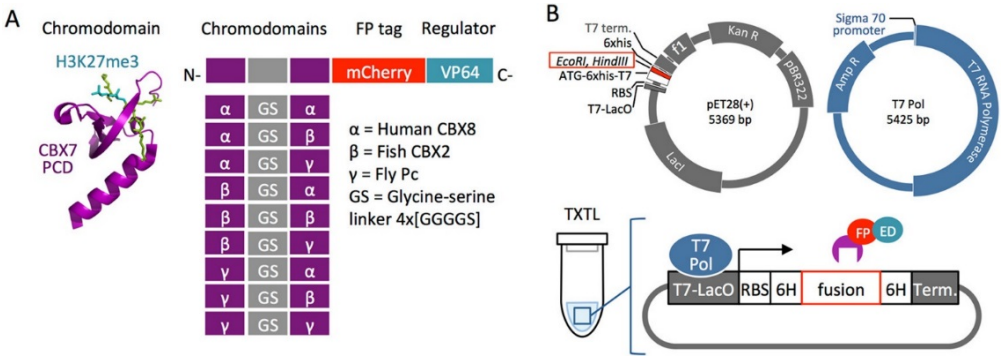


Figure 2.1 Constructs, plasmids, and a schematic for TXTL expression. (A) We constructed a library of nine fusion histone-binding protein open reading frames (ORFs) with tandem N-terminal HBDs separated by a 20-amino acid flexible linker (4x[GGGGS]). The 3D structure for CBX7 bound to H3K27me3 (Protein Data Bank entry 4X3K) is shown to represent CBX8/H3K27me3. (B) Plasmid maps, drawn to scale, used for expression in TXTL. Fusion protein ORFs were cloned into the *Eco*RI, *Hin*dIII site of pET28. A T7 RNA polymerase-expressing plasmid was included in the TXTL reactions to supply T7 polymerase, which is not present in the lysate mix. Legend: 6H, six-histidine; Term., transcription terminator sequence.

Real Time Detection of Cell-Free Transcription-Translation (TXTL)

TXTL is a flexible system that enables expression of recombinant, ready-to-use proteins without the need for cell lysis and purification. In reaction mixtures containing ∼100 ng of fusion-encoding plasmid, an increase in the intensity of the RFP signal began at just >2 h and continued to increase linearly up to 19 h (Figure 2.2). Our previous work with similar fusion proteins, including the Pcαα variant used here, showed a sigmoidal trend for RFP signal over time in TXTL reactions and began to plateau at 12−16 h.14 The slower, varying linear increases we observed in the current experiments could be due to incomplete translation, degradation, or misfolding of the fusion proteins. A Western blot with an antibody against mCherry (Figure 2.S1) showed that the full-length proteins were the predominant product in a representative set of TXTL reactions. Another possible cause of variability is inefficient production of fusion proteins due to batch-to-batch differences in components of the TXTL reagent (e.g., ribosomes, ATP, nucleases, proteases, etc.) or variable plasmid quality (i.e., supercoiled or nicked DNA). In the Western blot, anti-mCherry signal intensities generally agreed with end point RFP signal values. We conclude that RFP signal variation is likely due to differences in production.

The general procedure used for TXTL is as follows. (1) Set up a TXTL reaction by combining 9 μL of Sigma 70 Master Mix (myTXTL, Arbor Biosciences), 1 μL of 5 nM P70a-T7 RNA pol vector (Arbor Biosciences), and 2 μL of 50 ng/μL T7 promoter-driven template vector (pET28) in a total volume of 12 μL in one well of a white, 96-well plate (Roche 04729692001). (2) Seal the plate with a clear film (Roche 04729757001) and run a Roche 480 LightCycler II program (or machine of your choice) as follows: 29

°C for 10 min and 30 °C for 1 s, scan at 533−610 nm twice, and repeat 96 times (total of 16 h).

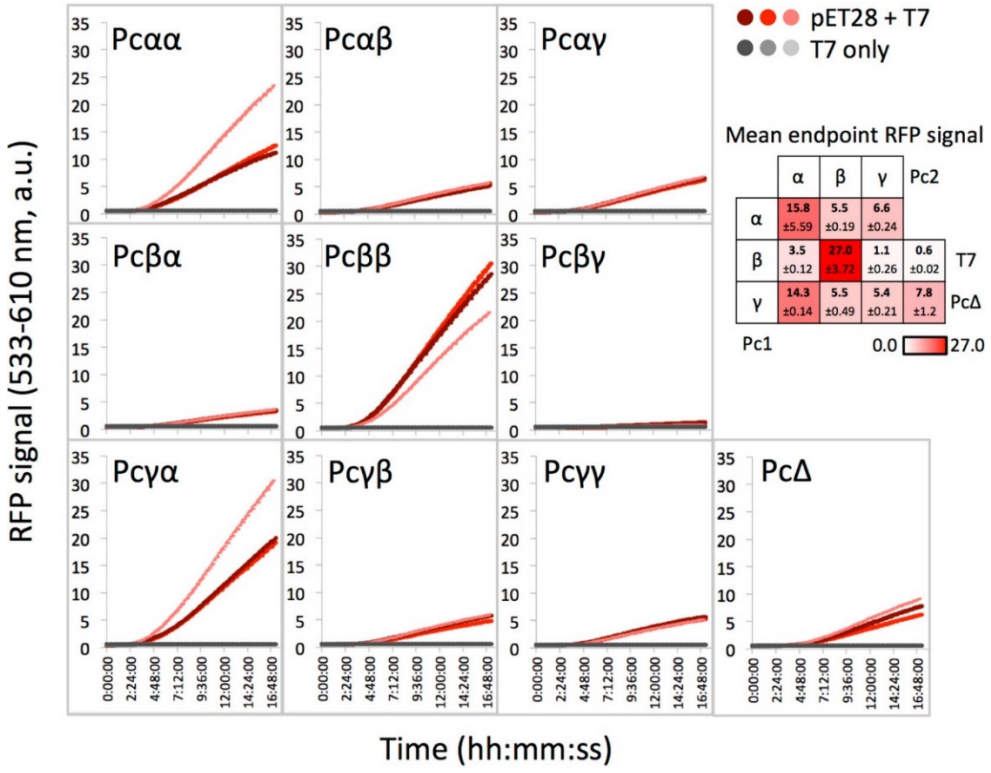


Figure 2.2 Real time detection of RFP-tagged fusion protein expression in TXTL. Graphs show the RFP signal captured every 10 min over ~16 h at a temperature of 29 °C, detected by a Roche Light Cycler 480 real time thermal cycler. The negative control (T7) contains the polymerase vector without the fusion-expressing pET28 vector. The grid shows the mean end point RFP signal values (bold) and standard deviations.

ELISA Capture for Quantification of Protein Generated by TXTL

Prior to histone PTM-binding experiments, total input (TXTL-generated fusion protein) should be determined with a reliable assay. In this report, we consider two parameters to represent input: end point fluorescent signal from the TXTL reaction (Figure 2.2) and signal values from control ELISA wells where antibodies are used to

capture the TXTL products. The TXTL RFP signal value is readily available and does not require additional experiments. However, this value may not correspond precisely with total full-length, histone-binding protein if incomplete folding of mCherry has occurred. Control ELISA wells using high-affinity antibodies against the fusion proteins circumvent the problem of misleading RFP signal values.

To determine the best epitopes for capture and detection, we compared different capture antibodies (Figure 2.3A): one against the fluorophore within the fusion (mCherry RFP) and the other against the six-histidine motif that appears at both the Nand C-termini of the fusion proteins. The ELISA HRP signals generated from the six-histidine capture were stronger than those from the anti-RFP capture. This is perhaps due to the higher ratio of six-histidine epitopes: two per fusion protein molecule, compared to the single mCherry tag (Figure 2.1B). The six-histidine capture data and the end point RFP signal values from real time detection of RFP in TXTL reactions show a modest correlation [R^2 = 0.61 with the Pcγα outlier removed (Figure 2.S2B)]. Therefore, the RFP signal may be used to roughly approximate the amount of fusion protein. However, epitope capture (e.g., via anti-six-histidine) should be used as a control to more accurately determine fusion protein input.

The general procedure for the ELISA control experiment is as follows. On the day the TXTL reaction is being set up and run (1 day before the ELISA experiment), prepare ELISA plates. (1) Add 50 µL of 2 µg/mL capture antibody in 200 mM carbonate buffer (pH 9.6) to each well. (2) Seal well with sealing foil (Cryostuff FS100) and incubate at 4 °C overnight. All subsequent incubations take place at 800 rpm at room temperature unless otherwise noted. (3) On the day of the ELISA, remove foil, shake out the coating

solution, and tap the inverted plate on paper towels to remove excess liquid. (4) Rinse the plate with 200 µL of 0.2% PBST (10 mM Na_2HPO4, 2 mM KH_2PO4, 2.7 mM KCl, 137 mM NaCl, and 0.2% Tween 20) three times for 3 min each. (5) Block the plate with 200µL of 5% BSA (Cell Signaling Technologies 9998S) in 0.2% PBST for 30 min. (6) Rinse the plate with 200 µL of 0.2% PBST three times for 3 min each. (7) Dilute the TXTL reaction mixture (~11 µL after evaporation) into 400 µL of 5% skim milk in 0.2% PBST. (8) Add 50 µL of the TXTL/skim milk solution to each well and incubate for 1 h. (9) Rinse the plate with 200 µL of 5% skim milk in 0.2% PBST three times for 3 min each. (10) Add 100 µL of the 1:3000 primary antibody diluted in 5% skim milk in 0.2% PBST to each well and incubate for 1 h. (11) Rinse the plate with 200 µL of 5% skim milk in 0.2% PBST three times for 3 min each. (12) Add 100 µL of the 1:3000 secondary, HRP-conjugated antibody diluted in 5% skim milk in 0.2% PBST to each well and incubate for 30 min. (13) Wash the plates five times with 200 µL of 0.2% PBST for 3 min each. (14) Remove buffer and add 100 µL of the one-step Ultra TMB-ELISA solution (Thermo 34029) to each well, protect from light, and incubate for 15 min. (15) Stop the reaction with 100 µL of 2.0 M sulfuric acid and incubate for 2 min. Measure A450 using a plate reader of choice (see details in Materials and Methods). See Figure 2.3A for an illustration of immunocapture and detection.

Initial First-Pass Screening for Histone PTM Binding

An enzyme-linked immunosorbent assay (ELISA) is used to assess histone PTM binding by measuring the amount of fusion protein that remains bound to an immobilized target ligand (histone peptide) after washing away unbound and nonspecifically bound proteins.

The procedure for measuring histone peptide capture includes the steps from "ELISA Capture for Quantification of Protein Generated by TXTL" with the following modifications. (1) Coat a 96-well plate (Greiner Bio-one 655101) with 20 ng/ μL neutravidin (Invitrogen 31000) in PBS [10 mM Na_2HPO_4, 2 mM KH_2PO_4, 2.7 mM KCl, and 137 mM NaCl (pH 8.0)] and add 50 μL to each well. (2) Follow steps 2−6 and then dilute biotinylated peptides to 1 μM in 0.2% PBST. Add 50 μL of the appropriate peptide solution to each well and incubate the plate at 800 rpm for 1 h. Following the peptide incubation, rinse the plate with 200 μL of 0.2% PBST three times for 3 min each. Block the plate with 200 μL of 5% skim milk (Apex 20- 241) in 0.2% PBST for 30 min. After this step, continue the protocol starting at step 7. See Figure 2.3B for an illustration of capture via immobilized histone peptides and detection.

We used the ELISA-based histone peptide capture procedure to generate data for the library of fusion proteins described in Figure 2.1. Normalized HRP signal values from the H3K27me3 capture showed that tandem human orthologues of the PCD motif (variant Pcαα) conferred the strongest binding (Figure 2.3B). ELISA trials using additional replicate TXTL products agreed with this general observation (Figure 2.S3). For the other variants, we observed relative binding values that were significantly higher ($p < 0.029$) than PcΔ, but at no more than half the level of Pcαα. Overall, these data suggest that Pcαα has the strongest preference for H3K27me3 versus unmodified H3 and that this activity is reduced when the second PCD is changed to the fly for fish orthologue. These results also provide the first in vitro evidence to suggest that low level of relative binding of the fly and fish HDBs with H3K27me3 may contribute to the weaker activity of the fusion gene regulators in human cells, as previously reported (5).

Next, we investigated histone peptide-binding activity by determining the proportion of input protein that remained bound in the H3K27me3 and H3 histone peptide capture wells (Figure 2.3C). When six-histidine capture was used to represent input, the values for all variants with the fish (β) or fly (γ) orthologue in position 1 were at least half of or slightly higher than Pcαα (Figure 2.3C, top chart). These variants also showed H3 peptide capture values significantly higher than Pcαα ($p < 0.017$), suggesting higher levels of nonspecific binding. When the data were calculated using the TXTL end point RFP signal to represent input, relatively high or low RFP values changed the relationships between values; e.g., the values for Pcβγ and Pcγα become higher and lower, respectively, than Pcαα (Figure 2.3C, bottom chart). Discrepancies between the six-histidine epitope capture values and RFP signal lead to different interpretations of the data and underscore the importance of using a reliable internal control to determine input protein levels in the ELISA-based experiments.

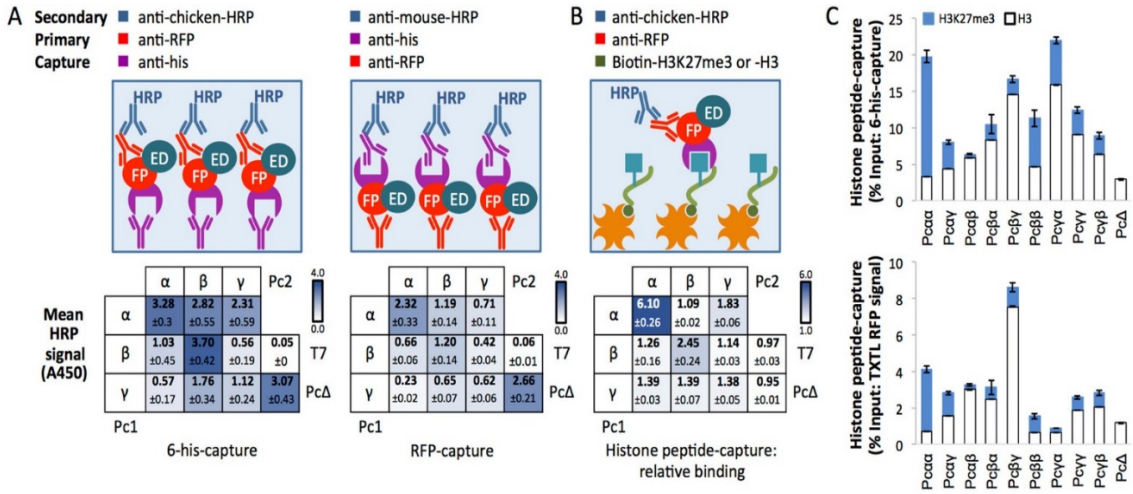


Figure 2.3 Relative protein levels and H3K27me3 binding of TXTL-expressed fusions determined by an ELISA. (A) To determine relative protein levels, three replicate TXTL reaction mixtures per fusion protein (from Figure 2.2) were combined and 1.4 µL of the

total TXTL reaction mixture was brought to 50 µL with 5% skim milk in 0.2% PBST and applied to each well containing immobilized anti-six-histidine (anti-his) or anti-mCherry (anti-RFP). Primary and secondary antibodies used in the detection step are shown in the table above each graphic. The grid shows mean values (bold) and standard deviations (±). (B) Interaction of each Pc fusion with the target ligand was assessed by histone peptide capture using H3K27me3 (green/blue) or unmodified H3K27 (negative control, not shown) anchored to immobilized neutravidin (orange). "Histone peptide capture: relative binding" is the mean A450 value from four replicate H3K27me3 wells, each normalized to the mean A450 for H3K27. (C) The top chart shows relative H3K27me3 capture scaled to six-His capture (A, bottom left). The bottom chart shows relative H3K27me3 capture scaled to mean RFP (from Figure 2.2). Legend: error bars, standard deviation; HRP, horseradish peroxidase; FP, fluorescent protein; ED, effector domain for gene regulation.

Determination of Target Specificity and Apparent K_d

After first-pass screening for fusions that preferentially bind the histone PTM of interest, ELISA and microspot arrays can be used to determine apparent dissociation constants (K_d^{app}) (14) as well as target histone PTM selectivity. Both histone peptide capture techniques detect the amount of protein that remains bound after washing away nonspecifically bound proteins. Therefore, the calculated K_d^{app} will be biased toward the off kinetics. Previously, we used *E. coli*-expressed, purified fusion proteins and an HRP ELISA to assess binding as a function of histone PTM concentration in mixtures of immobilized ontarget PTMs (H3K27me3) with unmodified peptides (H3). We used direct

visualization of mCherry (RFP) on microspot arrays to assess binding as a function of fusion protein concentrations.

TXTL potentially produces a suitable amount of protein for quantitative analysis. For our previously reported HRP ELISA binding curves (14) 100 nM (50 μL) E. coli-expressed, purified Pcαα showed a ratio of ~5:1 for signal (binding to 100% H3K27me3) to background (100% H3K27), as well as significant signal for 20−90% H3K27me3. In the current study, we observed that 1.4 μL of TXTL-expressed Pcαα produced a HRP ELISA signal:background ratio of ~6:1 (Figure 2.3). Therefore, we would expect to detect a range of binding values (above background) if TXTL-produced Pcαα protein were exposed to an H3K27me3 dilution series (as previously described (14)). This should hold true for other Pc fusion variants with similar HRP ELISA values.

An important feature of histone PTM-binding domains (HBDs) is their ability to discriminate between distinct histonetags within a cellular chromatin landscape that is decorated with a wide variety of modifications. Therefore, it is important to perform histone peptide capture using a variety of potential "off-target" PTMs, e.g., H3K27me3 versus K4me3, K9me3, and K36me3, or different degrees of methylation (me1−me3). In our previous work, ELISA experiments showed that Pcα fusions had no interaction with K4me3 or K9me3 that was significantly stronger than that of unmodified H3 (14). Microspot binding assays have a substantial advantage over an ELISA in that the former supports miniaturization, which allows a variety of histone PTMs at several concentrations to be tested in a single application of soluble protein. One challenge for this approach is that it does not allow signal amplification (i.e., via HRP) prior to detection and thus requires a sufficient amount of protein to visualize linked tags (e.g.,

mCherry used here) or immunostaining of the histone-peptide-bound fusion proteins. In our previous work, purified Pcαα (14) produced a detectable signal (~2000 fluorescence units) on H3K27me3 microspots (1 mm each) with no background signal from unmodified H3 peptides when 0.5 mL of 100 nM fusion protein was applied over an area of 187 cm^2. Theoretically, if 1.4 μL of TXTL is equivalent to 50 μL of 100 nM purified protein (as suggested by the HRP ELISAs that used purified proteins), we could expect to require at least ~35 μL of TXTL to produce a detectable signal on a microspot array. Because the reaction volume used here is 11 μL, a few reactions may need to be combined for a sufficient yield. We encourage researchers to review the work of Tekel etal (14) and Filippakopoulos et al (26) for examples of glass slide- and cellulose membrane-based histone peptide capture experiments, respectively.

Construction of Plasmids for Mammalian Cell Expression

We have constructed a vector called MV10 (14) for the overexpression of fusion proteins in frame with a nuclear localization sequence and six-histidine tag. A single XbaI cloning site is included to accept BioBrick RFC23 standard inserts flanked by XbaI and SpeI overhangs (Figure 2.4). Nondirectional cloning theoretically yields equal proportions of forward (desired) and reverse insertions. We recommend the following colony PCR protocol to quickly screen for forward insertions. (1) Assemble constructs via nondirectional cloning of fusion-coding regions into vector MV10 using restriction digests and ligations as previously described (14). (2) Prepare colony PCR mixes as follows: 1× GoTaq green master mix (Promega M7122), 0.4 μM forward primer 5′-caccatcgtggaacagtacg, and 0.4 μM reverse primer 5′-gcaactagaaggcacagtcg in a final volume of 25 μL in each 0.5 mL tube. (3) Pick colonies grown from ligation-transformed

E. coli with a sterile, disposable micropipette tip, streak onto a small area (1 cm^2) on a gridded and labeled LB agar, 100 µg/mL ampicillin plate, swirl the same tip in each PCR mix, and discard the tip. Incubate the streak plate overnight at 37 °C. (4) Perform PCR as follows: 95 °C for 2 min; 25 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min; 72 °C for 5 min; and held at 4 °C. (5) Analyze 20−50% of each reaction mixture via gel electrophoresis. (6) Inoculate 5 mL of liquid LB, 100 µg/mL ampicillin cultures with candidate clones from the streak plate; grow for 18 h at 37 °C with shaking, and miniprep the plasmid DNA (Sigma PLN350-1KT). (7) Perform restriction digests as follows: 500−1000 ng of plasmid DNA, 1 µL of each of FastDigest enzymes XbaI and PstI, and 1× FastDigest buffer (Thermo Fisher FD0684, FD0614, and B72). Analyze the products using standard agarose gel electrophoresis with TAE buffer.

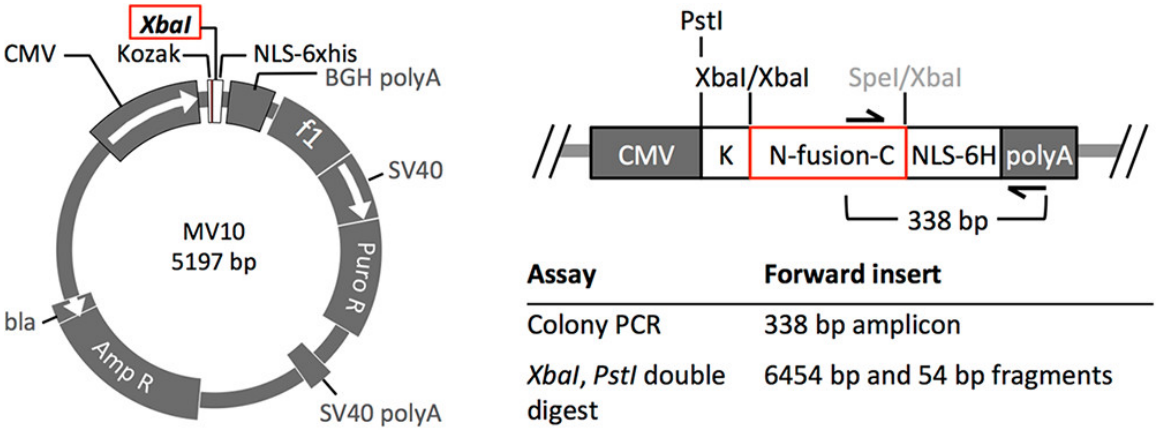


Figure 2.4 Vector for mammalian cell expression and scheme for verifying constructs. The MV10 vector is designed to accept *Xba*I–*Spe*I donor fragments at a *Xba*I site so that the open reading frame (red) is cloned in frame with the translation start site within the yeast-derived Kozak (K) sequence and the downstream nuclear localization sequence, six-histidine tag, and stop codon. The downstream *Spe*I/*Xba*I junction is a mixed site (5′-actaga) that cannot be recut. Concurrence of both results shown for the validation assays

is unique for a single forward insert for any construct that includes the C-terminal mCherry-VP64 sequence used here.

Determination of Transcriptional Regulation in Mammalian Cells: Microwell Assay for Luciferase

Standard transfection techniques (polyplex or electroporation) can be used to deliver fusion-expressing plasmids into cells. In our hands, transiently transfected cells (nonhomogeneous) for which the transfection efficiency is ≥50% (determined by fluorescence microscopy and flow cytometry) is sufficient for quick validation of large sets of plasmids. For more rigorous quantitative analyses of histone PTM-binding candidate fusion transgenic line in Limitations.

We have used an H3K27me3-enriched locus as a target to test the function of PCD fusion regulators (Figure 2.5A). In this system, accumulation of H3K27me3 near the promoter of a luciferase target gene in HEK293 cells is controlled by adding doxycycline to the growth medium (as described in Cell Culture and Transfection). Other mammalian ectopic chromatin systems have been developed for H3K27me3 (27), H3K9me3 (27, 28), and acetylation of H3 and H4 (29). These systems provide a powerful platform for comparing the activity of libraries of histone PTM-binding regulators but represent a very limited set of PTMs; we discuss endogenous loci as an alternative option in the next section.

Regulation of a reporter gene at ectopic chromatin is performed as follows. (1) Transfect cells and grow for 3 days to allow fusion protein expression and regulation of the target. (2) Collect semiadherent cells with 1× PBS washes, pellet (200 g, room temperature, 5 min), and resuspended in 1.5 mL of 1× PBS. (3) Transfer cells to Black

Costar Clear Bottom 96-Well Plates (Corning 3631). To test each sample in triplicate, load three wells with a 200 µL cell suspension for RFP detection, three wells with a 100 µL cell suspension and 100 µL of 2× Hoechst 33342 stain (2 µg/mL, Invitrogen H3570), and three wells with 100 µL of cells and 100 µL of prepared Luciferase Assay Buffer with D-luciferin substrate (Biotium 30085). Add 200 µL of 1× PBS to one or more columns of wells to determine background noise. Allow the plate to incubate at room temperature protected from light for 10 min, allowing cells to settle for optimal fluorescence reads. (4) Scan the plate in a Biotek Synergy H1 microplate reader with the following parameters: read, fluorescence at 580−610 nm (RFP), gain set to 100; read fluorescence, 360−460 nm (Hoechst 33342), gain set to 75; orbital shake, 5 min; read chemiluminescence (luciferase activity). (5) Determine fluorescence signal detection limits of the plate reader with a set of standard samples as described in Figure 2.S4. (6) Analyze the data from each channel separately. Subtract the average background (1× PBS wells) from each value. See Figure 2.5B for an illustration of a microwell plate configuration.

Results from the in vitro histone peptide capture assays suggest that Pcα confers H3K27me3-selective binding to HDB fusion proteins (Figure 2.3B). The other orthologues might enhance general, nonspecific interaction with unmodified histone H3, as indicated by relatively high input-normalized values (Figure 2.3C). To investigate the significance of the observed histone peptide-binding activities for gene regulation activity, we tested five fusion protein variants in the mammalian cell assay: two that showed high and moderate levels of relative binding with H3K27me3 and low levels of H3 binding (Pcαα and Pcββ), one that showed moderate levels of relative binding and

high levels of H3 binding (Pcγα), other variants that showed low levels of relative binding and contained at least one human PCD orthologue (Pcαβ, Pcαγ, or Pcβα), and the truncated PcΔ protein (negative control). Cells expressing Pcαα showed the highest levels of luciferase reactivation (luciferase enzyme activity), and the only mean value that was higher (p = 0.058) than PcΔ after scaling by the RFP signal (Figure 2.5C). This result suggests that a high relative binding level of a fusion protein with H3K27me3 from the ELISA-based assay may predict effective transcriptional activation at a silenced chromatin site in cells. Overall, our results demonstrate that an ELISA using TXTL-expressed proteins is an effective method for identifying histone PTM-binding regulators that can activate target genes in live cells.

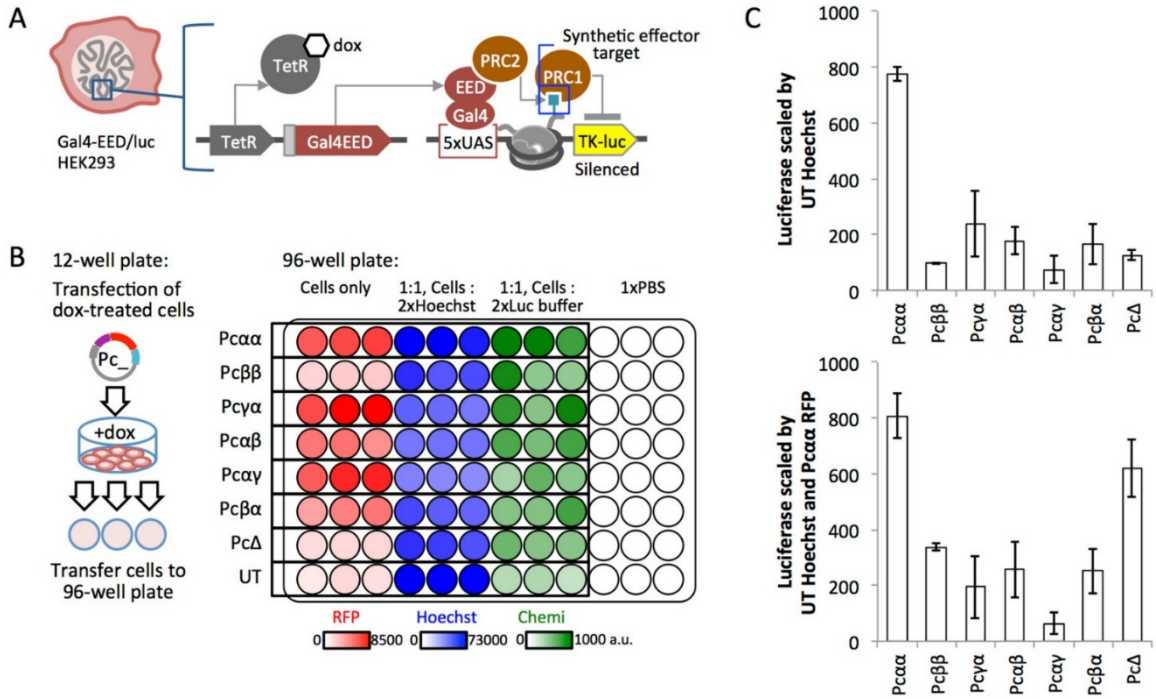


Figure 2.5 Microwell plate reader assay for assessing fusion protein expression and regulation of a target reporter gene. (A) Gal4-EED/luc cells were treated with dox to induce ectopic recruitment of polycomb repressive complex 1 (PRC1) and accumulation of H3K27me3 and PRC2 at a *luciferase* reporter. (B) After *luciferase* silencing, cells

were transfected with each fusion-expressing plasmid, harvested, and aliquoted into a 96-well plate. In this procedure, RFP is used to assess fusion protein expression, the signal from the Hoechst 33342 DNA stain is a proxy for cell loading, and luciferase activity indicates gene expression induced by each Pc activator fusion. (C) Bar charts show mean background-subtracted values from triplicate wells, normalized and scaled as described in Materials and Methods. Legend: error bars, standard deviation; UT, untransfected cells.

Determination of Gene Regulation in Mammalian Cells: RT-qPCR Assay for Endogenous Genes

If the Gal4-EED/luc system or other similar reporters are not available or relevant for one's own project, we recommend an expression assay based on an endogenous target locus (or several loci) with a well-characterized epigenetic state. These genes can be identified as those reported to show changes in expression when histone-modifying enzymes are disrupted by inhibitors or RNAi-mediated knockdown. For instance, our group identified a set of 14 genes for which polycomb-mediated silencing was supported by genetic and pharmacologic disruption studies in human cancers and stem cells (4). Epigenetic regulation can be confirmed by investigating genomic chromatin feature maps. Table 2.1 presents publicly available data from whole-genome chromatin profiling experiments (chromatin immunoprecipitation followed by deep sequencing, ChIP-seq) to assist in the identification of target loci in widely used model cell lines.

Reverse transcription followed by quantitative PCR (RTqPCR) (30) is a standard, flexible method for determining changes in gene transcription. For each experiment, a minimum of three transcripts must be measured: a housekeeping gene (e.g., GAPDH), the fusion regulator, and the regulated target gene. We have successfully used RT-qPCR to

measure changes in luciferase expression in the Gal4-EED/luc reporter cell line to validate luciferase assay results (14) and to measure the expression of several endogenous genes in the presence of an H3K27me3- binding fusion activator called PcTF (4-5). In our hands, it is not necessary to sort for fusion protein-positive cells prior to RTqPCR because the gene expression values can be scaled by RFP signal (flow cytometry) or fusion mRNA levels as needed. (1) Transfect ∼1.0×10^5 cells per well as described in Materials and Methods. Include a mock-transfection control: empty vector or water instead of fusion protein-expressing DNA. Grow for 2−3 days in an appropriate cell culture medium. (2) Prepare total RNA from ∼1.0×10^6 cells using a method similar to that previously described by our group (2-3). We recommend direct lysis of pelleted cells with 500 μL of TRIzol (Thermo Fisher 15596026) followed by spin column purification of RNA from the aqueous phase of chloroformextracted samples (Qiagen RNeasy Mini kit 74104). (3) Generate cDNA libraries by using 2 μg of total RNA and the SuperScript III first-strand synthesis system (Invitrogen 18080051). (4) Perform qPCR (15 μL each) using 2 μL of a 1:10 cDNA dilution for the endogenous gene(s) of interest or a 1:1000 dilution for the highly expressed genes (i.e., reference gene GAPDH, ACTB, etc., or the overexpressed fusion transgene). Follow the manufacturer's recommendations for other reagents (primers, probes, polymerase, dNTPs, and buffer). (5) Calculate the mean quantification cycle (C_q, also described as C_t and C_p) for three replicate wells per unique reaction. Calculate the expression level with the equation $\Delta Cq = {}^{2\text{mean Cq reference}-\text{mean Cq target}}$. Calculate the log 2-fold change in gene expression as $\log 2(\Delta C_{q\ \text{transfected cells}}/\Delta C_{q\ \text{mock}})$.

Table 2.1 Recommended Human Cell Lines For Identifying Endogenous Target Sites

Histone H3 PTM	Corresponding HBD	Cells lines and ChIP-seq data sources			
		HEK293 (kidney)	MCF7 (breast)	K562 (blood)	H1-hESC (stem)
H3K4me3	PHD[8]	ENCSR000DTU	ENCSR000DWJ	ENCSR668LDD	ENCSR000AMG
H3K9me3	CD[7]	ENCSR000FCJ	ENCSR000EWQ	ENCSR000APE	ENCSR000APZ
H3K14ac	BRD[6]	N/A	N/A	N/A	ENCSR057BTG
H3K27me3	PCD[5,8,14]	N/A	ENCSR000EWP	ENCSR000EWB	ENCSR000ALU
H3K27ac*	YEATS[33]	ENCSR000FCH	ENCSR000EWR	ENCSR000AKP	ENCSR880SUY

[a]Detailed information for HBDs tested in vitro and in cells is available in the cited references and in our recent review (9). Data ID numbers from the ENCODE project online database at https://www.encodeproject.org/experiments are listed. These can be used to access and download BAM files. To view ChIP enrichments, we recommend loading the BAM file into the Integrated Genome Viewer from the Broad Institute (free download, http://software.broadinstitute.org/software/igv/). [b]Abbreviations: PHD, plant homeodomain; PCD, polycomb chromodomain; CD, chromodomain; BRD, bromodomain; YEATS, "Yaf9, ENL, AF9, Taf14, Sas5" domain. [c]Not available in ENCODE. [d]Also expected to interact with H3K9ac and H3K18ac.

Chromatin Immunoprecipitation to Determine Fusion Protein Binding at Target Loci

A critical step in validating the behavior of fusion proteins that interact with histone PTMs is to detect the interaction of the fusion protein with target loci in cells. For fusion transcriptional regulators, it is important to identify instances in which the histone PTM binding fusion interacts with chromatin sites but does not influence gene

expression. Furthermore, several HDB fusion proteins have been used to label histone PTM-enriched regions and are not intended to artificially regulate genes (6, 8, 31). Crosslinked chromatin immunoprecipitation (X-ChIP), first reported in 1988 (33) has become a gold standard method for measuring the accumulation of proteins at specific genomic loci. In brief, cells are treated with formaldehyde to cross-link chromatin proteins with target DNA, the chromatin is sheared (sonicated or enzymatically digested) into soluble particles, particles containing the protein of interest (e.g., the fusion protein) are immunoprecipitated, and the co-immunoprecipitated DNA is purified and analyzed by quantitative PCR of specific sites (ChIP−qPCR) or by next-generation sequencing and computational alignment of all DNA fragments in the sample (ChIPseq). Fusion proteins expressed from transgenes are highly amenable to ChIP because they can be designed to include a reliable epitope tag, and uniform expression across all cells can be achieved in a stable transgenic cell line. In our previous work, we used ChIP-seq to profile the genome wide distribution of a Pcα fusion (called PcTF) and its target histone PTM H3K27me3 in a transgenic U2OS cell line (4). As an alternative to the resource-intensive ChIP-seq assay, we have used ChIP− qPCR to measure changes in Pc fusion and histone PTM levels over time near a specific PcTF-responsive gene (4). In cases in which the HBD fusion shows strong, specific localization within nuclei, immunofluorescence cytology with antibodies against specific histone PTMs can be used to assess the co-localization of HBD fusion proteins and their targets (34).

Limitations

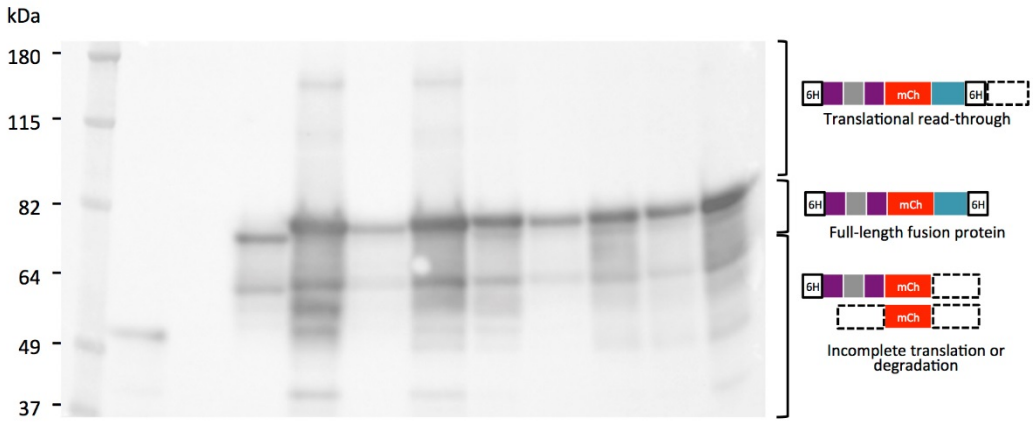
TXTL May Not Be Sufficient for Single-HBD Fusion Proteins. In previous trials, we observed very little HRP signal over background for a TXTL-expressed Pc fusion

called PcTF, which carried a single copy of Pcα (unpublished). However, using overexpressed and purified protein from transformed *E. coli*, we demonstrated specific binding of PcTF to H3K27me3 (vs H3Kme3 and H3K9me3) in an ELISA. We also observed H3K27me3 binding of PcTF at levels higher than that of the negative control (mCherry-VP64) using a histone peptide microspot array. These assays showed a weaker K_d^{app} value (~0.5-fold) for PcTF than for the double- HBD variant Pc_2TF [identical to Pcαα (Figure 2.1)]. Thus, the lower-affinity fusion PcTF functions in vitro, but the TXTL reaction as described here may not produce a sufficiently high yield for a detectable HRP signal after binding and washing during the ELISA.

Fusion Protein Dose and Cell Heterogeneity Are Difficult To Measure and Control in Transiently Transfected Cells. The transiently transfected HEK293 cells that we generated for this study produced RFP signal levels that varied between expression plasmids and were near the lower end of the detection limits for the microwell plate reader that we used (Figure 2.S4). Therefore, we recommend transient transfections only for preliminary assessment of expression levels and fusion regulator activity in a model system. For follow-up studies that require precise measurements of gene expression in response to the dose (expression level) of the fusion regulator or require a homogeneous population of transgenic cells for optimal results (e.g., ChIP-seq), we recommend generating stable cell lines. Inducible transgene expression systems prevent premature expression of the fusion, which may broadly regulate genes in the host cells prior to treatment and analysis. There are many available options that do not require viral transduction. For instance, we have used the popular T-REx system (Invitrogen) to express PcTF in U-2 OS cells (4-5). We have also successfully used an all-in-one

doxycycline-inducible expression cassette based on the sleeping beauty transposon-based system pSBtet- GP (35) to express PcTF in MCF7 cells.

Associated Content



Variant	PcΔ	T7	Pcαα	Pcγα	Pcβα	Pcαγ	Pcγγ	Pcβγ	Pcαβ	Pcγβ	Pcββ
Expected size (kDa)	37.1	n/a	53.8	55.8	55.7	55.8	56.0	55.9	55.7	55.9	55.7
Endpoint RFP signal (533-610, a.u.)	31.8	0.7	37.9	49.3	9.4	41.9	11.7	3.5	14.1	18.8	33.9

Figure 2.S1 Western blot analysis of TXTL-expressed fusion proteins. TXTL reactions containing each fusion construct ("Variant") plus T7 RNA polymerase vector, or T7 RNA polymerase alone (negative control) were run as described in Materials and Methods. Samples were subjected to PAGE, blotted onto nitrocellulose, and immunostained with anti-mCherry and an HRP-conjugated secondary antibody. The procedure is described in detail in Supporting Methods. Predicted molecular weights (expected size, kDa) were calculated using the protein report tool within the CLC Main Workbench program (Qiagen). The positions of the TXTL-expressed fusion protein bands relative to the protein standard (lane 1) and to each other are consistent with our previously published Western blots for mammalian cell-expressed proteins (1). The protein standard shown here (Thermo Fisher #10748010) may not be an accurate indicator of protein size.

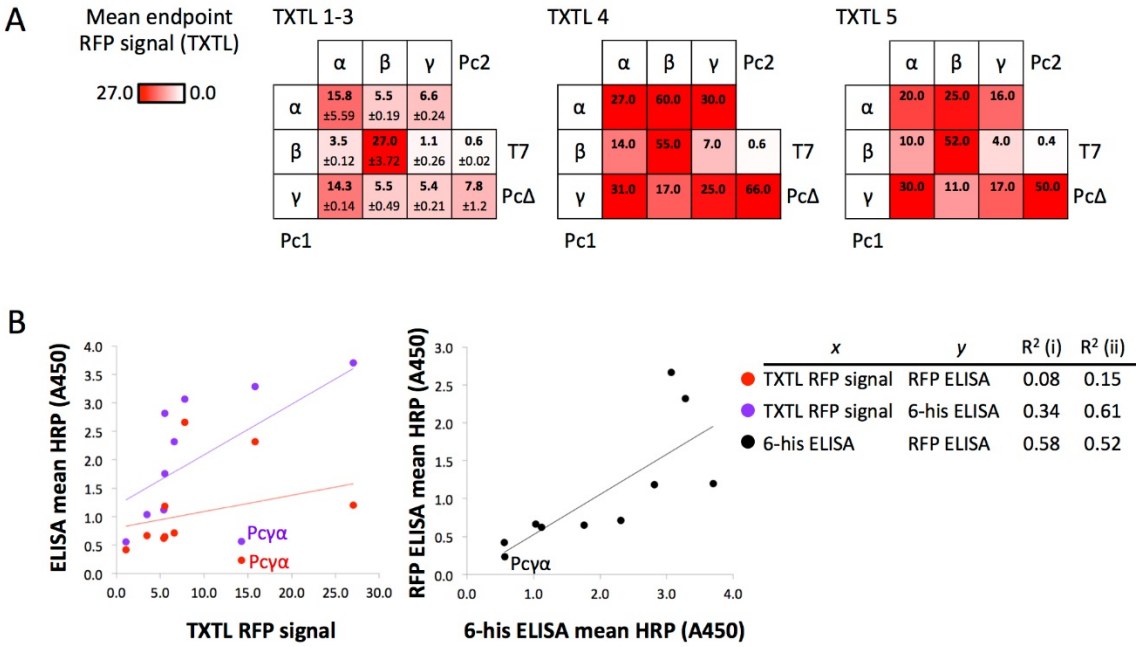


Figure 2.S2 Comparison of TXTL RFP signal with total protein determined by RFP- or 6-his capture ELISA. (A) To determine reproducibility of expression via TXTL, we ran two additional TXTL reactions (TXTL 4 and 5) per fusion construct using the same plasmid DNA preparations are TXTL 1-3. "TXTL 1-3" represents the samples and values from Figure 2.2. Each value shown for TXTL 4 and 5 is from a single reaction. (B) Samples TXTL 1-3 were combined and used for ELISA experiments with different capture antibodies, anti-RFP or anti-6-histidine. Left dot plot: comparison of the endpoint TXTL RFP signal values (TXTL 1-3 in panel A) with the mean HRP signal values from the ELISA data (Fig. 3A). Right dot plot: comparison of the mean HRP signal values for 6 his versus RFP-capture. R squared values are shown for comparisons of all data (i) or with the Pcγα outlier excluded (ii).

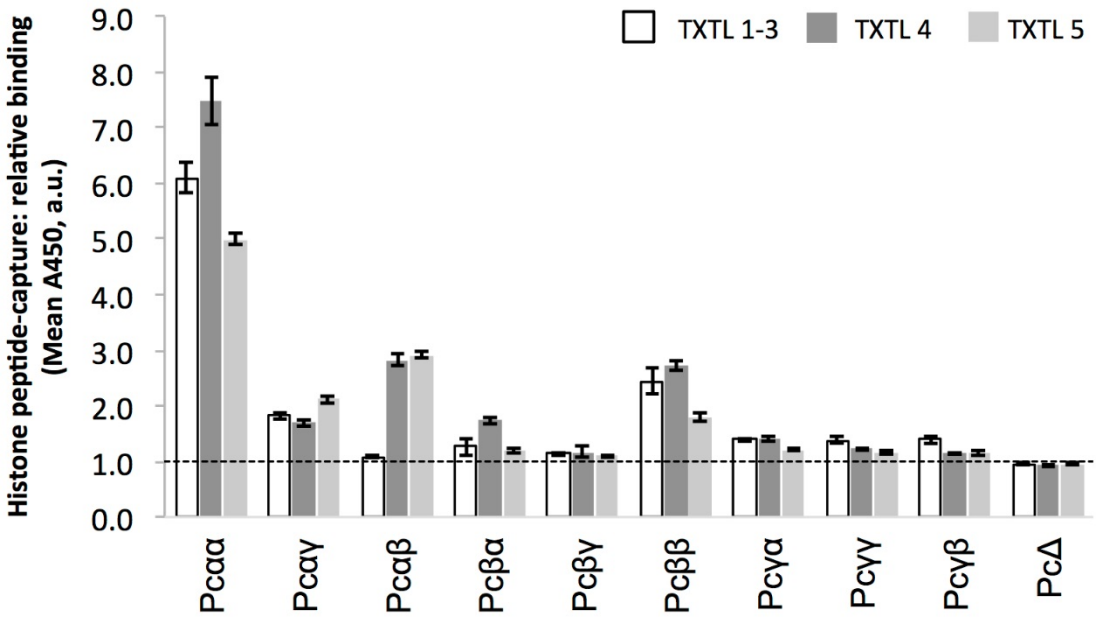


Figure 2.S3 ELISA assays for additional replicate TXTL samples. To determine reproducibility of histone capture for TXTL-expressed fusion proteins we carried out ELISA for two additional TXTL reactions (4 and 5). Endpoint RFP signal values for these samples are shown in Fig. S2A. Error bars, standard deviation.

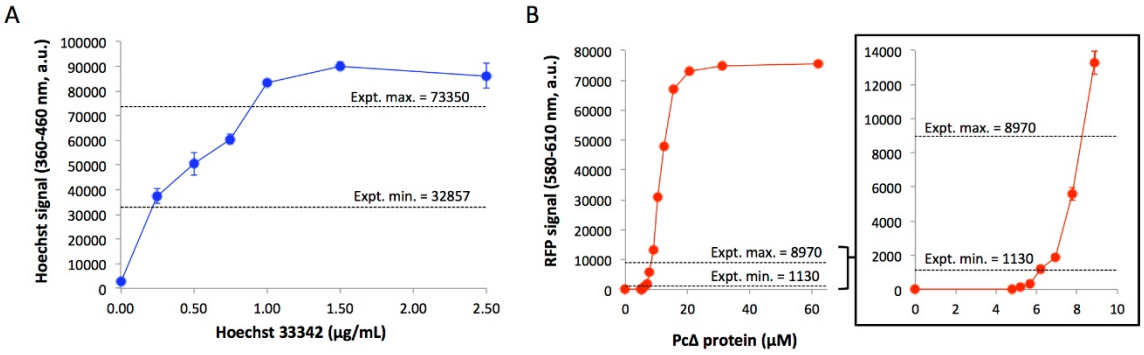


Figure 2.S4 Determination of plate reader detection limits for Hoechst and RFP (mCherry). Samples of varying signal intensity were used to determine the linear range of signal detection for the Biotek Synergy H1 plate reader using the settings described in our report. (A) To prepare samples for the Hoechst channel (360-460 nm), HEK293 cells were grown to 90% confluency in a 12-well plate, collected, stained with varying

concentrations of dye. Dots show mean values from triplicate cell samples (bars, standard deviation). The maximum and minimum Hoechst signal values from the plate reader data from Figure 2.5 are marked with dotted lines. (B) Samples for the RFP channel contained 5.0 to 62.0 μM 6-histidine-tagged recombinant PcΔ protein diluted in 1X PBS (200 μl per well). PcΔ was over-expressed in *E. coli* cultures and purified on a nickel-column as described in detail in Tekel et al. Dots show mean values from triplicate protein samples (error bars, standard deviation).

Western blot and immunostaining of TXTL-expressed fusion proteins. Samples for denaturing polyacrylamide gel electrophoresis (PAGE) were prepared by heating 1 μL of TXTL reaction plus 4 μl NuPAGE LDS Sample Buffer 4x (Thermo Fisher #NP0007) and 1 μl 1 M Dithiothreitol (DTT, Millipore Sigma #D0632-1G) at 95°C for 5 min in a total volume of 20 μL. Samples (cooled to room temperature) and a pre-stained protein standard (10 μl, Thermo Fisher #10748010) were electrophoresed at 200 V for 45 minutes in a 4-12% Bis-Tris gel (Thermo Fisher #NP0322BOX) with MOPS-SDS buffer [(50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7) diluted 1:20] in an XCell SureLock vertical chamber (Invitrogen #EI0001). 0.5 mL antioxidant solution (Invitrogen #NP0005) was added to the inner (cathode) chamber prior to electrophoresis. PAGEseparated proteins were transferred onto a nitrocellulose membrane (Bio-Rad #1704158) via semidry transfer in a Trans-Blot Turbo system (Bio Rad #1704150). Complete transfer was verified by S6 staining the membrane with 1x Ponceau-S (G-Biosciences #786-576). Immunostaining was carried out with the following: blocking buffer, 5% nonfat dry milk in 1x PBST (1x PBS, 0.2% Tween-20); primary, chicken polyclonal anti-mCherry 1:1000 (Novus Biologicals #NBP2-25158); secondary, HRP-

conjugated rabbit anti-chicken 1:5000 (Millipore Sigma #AP162P). Immunostaining was performed at 4°C overnight (primary) or at room temperature for 1 hour (secondary) with rotation in a Parafilm pouch. Immunostained blots were washed 4x, 10 min each in 1x PBST, with orbital shaking at room temperature. HRP signal was detected using the SuperSignal West Femto substrate kit (Thermo Fisher #34095) and a PXi4 imager (Syngene) with GeneSys software.

Abbreviations

PTM, post-translational modification; HBD, histone PTM binding domain; PCD, polycomb domain; PHD, plant homeodomain; HRP, horseradish peroxidase; EED, embryonic ectoderm development; VP64, viral protein 64; BSA, bovine serum albumin; RFP, red fluorescent protein; ELISA, enzyme linked immunosorbent assay; Pc, polycomb; TXTL, E. coli based in vitro transcription and translation system; mCh, mCherry; BRD, bromodomain; CD, chromodomain; PCR, polymerase chain reaction.

References

1. Jones, P. A., Issa, J.-P. J., and Baylin, S. (2016) Targeting the Cancer Epigenome for Therapy. Nat. Rev. Genet. 17 (10), 630−641.

2. Thakore, P. I., Black, J. B., Hilton, I. B., and Gersbach, C. A. (2016) Editing the Epigenome: Technologies for Programmable Transcription and Epigenetic Modulation. Nat. Methods 13 (2), 127−137.

3. Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempl, P. M., Schmidl, C., et al. (2016) Comprehensive Genome and Epigenome Characterization of CHO Cells in Response to Evolutionary Pressures and over Time. Biotechnol. Bioeng. 113 (10), 2241−2253.

4. Nyer, D. B., Daer, R. M., Vargas, D., Hom, C., and Haynes, K. A. (2017) Regulation of Cancer Epigenomes with a Histone-Binding Synthetic Transcription Factor. npj Genomic Med. 2, 1.

5. Haynes, K. A., and Silver, P. A. (2011) Synthetic Reversal of Epigenetic Silencing. J. Biol. Chem. 286 (31), 27176−27182.

6. Sanchez, O. F., Mendonca, A., Carneiro, A. D., and Yuan, C. (2017) Engineering Recombinant Protein Sensors for Quantifying Histone Acetylation. ACS Sens 2 (3), 426−435.

7. Lungu, C., Pinter, S., Broche, J., Rathert, P., and Jeltsch, A. (2017) Modular Fluorescence Complementation Sensors for Live Cell Detection of Epigenetic Signals at Endogenous Genomic Sites. Nat. Commun. 8 (1), 649.

8. Delachat, A. M.-F., Guidotti, N., Bachmann, A. L., Meireles-Filho, A. C. A., Pick, H., Lechner, C. C., Deluz, C., Deplancke, B., Suter, D. M., and Fierz, B. (2018) Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. Cell Chem. Biol. 25 (1), 51−56.e6.

9. Tekel, S. J., and Haynes, K. A. (2017) Molecular Structures Guide the Engineering of Chromatin. Nucleic Acids Res. 45 (13), 7555−7570.

10. Park, M., Keung, A. J., and Khalil, A. S. (2016) The Epigenome: The next Substrate for Engineering. Genome Biol. 17 (1), 183.

11. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., et al. (2011) Recognition and Specificity Determinants of the Human Cbx Chromodomains. J. Biol. Chem. 286 (1), 521−529.

12. Flynn, E. M., Huang, O. W., Poy, F., Oppikofer, M., Bellon, S. F., Tang, Y., and Cochran, A. G. (2015) A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications. Structure 23 (10), 1801−1814.

13. Jung, M., Philpott, M., Müller, S., Schulze, J., Badock, V., Eberspächer, U., Moosmayer, D., Bader, B., Schmees, N., Fernández- Montalván, A., et al. (2014) Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the Histone H4 Tail and the Small Molecule Inhibitor JQ1. J. Biol. Chem. 289 (13), 9304−9319.

14. Tekel, S. J., Vargas, D. A., Song, L., LaBaer, J., Caplan, M. R., and Haynes, K. A. (2018) Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. ACS Synth. Biol. 7 (3), 842−852.

15. Phillips, I., and Silver, P. (2006) A New Biobrick Assembly Strategy Designed for Facile Protein Engineering.

16. 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., et al. (2011) Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. Cell 145 (5), 692−706.

17. Huh, J.-W., Wu, J., Lee, C.-H., Yun, M., Gilada, D., Brautigam, C. A., and Li, B. (2012) Multivalent Di-Nucleosome Recognition Enables the Rpd3S Histone Deacetylase Complex to Tolerate Decreased H3K36 Methylation Levels. EMBO J. 31 (17), 3564−3574.

18. Lee, C.-H., Wu, J., and Li, B. (2013) Chromatin Remodelers Fine-Tune H3K36me-Directed Deacetylation of Neighbor Nucleosomes by Rpd3S. Mol. Cell 52 (2), 255−263.

19. Khuong, M. T., Fei, J., Cruz-Becerra, G., and Kadonaga, J. T. (2017) A Simple and Versatile System for the ATP-Dependent Assembly of Chromatin. J. Biol. Chem. 292 (47), 19478−19490.

20. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003) Molecular Basis for the Discrimination of Repressive Methyl-Lysine Marks in Histone H3 by Polycomb and HP1 Chromodomains. Genes Dev. 17 (15), 1870−1881.

21. Min, J. R., Zhang, Y., and Xu, R.-M. (2003) Structural Basis for Specific Binding of Polycomb Chromodomain to Histone H3Methylated at K27. Genes Dev. 17, 1823−1828.

22. Su, X., Zhu, G., Ding, X., Lee, S. Y., Dou, Y., Zhu, B., Wu, W., and Li, H. (2014) Molecular Basis Underlying Histone H3 Lysine- Arginine Methylation Pattern Readout by Spin/Ssty Repeats of Spindlin1. Genes Dev. 28 (6), 622−636.

23. Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I. A., Zenn, H. M., Kost, N., Pohl, W., et al. (2016) Dynamic and Flexible H3K9me3 Bridging via HP1β Dimerization Establishes a Plastic State of Condensed Chromatin. Nat. Commun. 7, 11310.

24. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal Structure of the Nucleosome Core Particle at 2.8 A Resolution. Nature 389 (6648), 251−260.

25. Engler, C., and Marillonnet, S. (2011) Generation of Families of Construct Variants Using Golden Gate Shuffling. Methods Mol. Biol. 729, 167−181.

26. Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., et al. (2012) Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. Cell 149 (1), 214−231.

27. Bintu, L., Yong, J., Antebi, Y. E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., and Elowitz, M. B. (2016) Dynamics of Epigenetic Regulation at the Single-Cell Level. Science 351 (6274), 720−724.

28. Snowden, A. W., Gregory, P. D., Case, C. C., and Pabo, C. O. (2002) Gene-Specific Targeting of H3K9Methylation Is Sufficient for Initiating Repression in Vivo. Curr. Biol. 12 (24), 2159−2166.

29. Kwaks, T. H. J., Sewalt, R. G. A. B., van Blokland, R., Siersma, T. J., Kasiem, M., Kelder, A., and Otte, A. P. (2005) Targeting of a Histone Acetyltransferase Domain to a Promoter Enhances Protein Expression Levels in Mammalian Cells. J. Biotechnol. 115 (1), 35−46.

30. Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., et al. (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin. Chem. 55 (4), 611−622.

31. Lungu, C., Pinter, S., Broche, J., Rathert, P., and Jeltsch, A. (2017) Modular Fluorescence Complementation Sensors for Live Cell Detection of Epigenetic Signals at Endogenous Genomic Sites. Nat. Commun. 8 (1), 649.

32. Wan, L., Wen, H., Li, Y., Lyu, J., Xi, Y., Hoshii, T., Joseph, J. K., Wang, X., Loh, Y.-H. E., Erb, M. A., et al. (2017) ENL Links Histone Acetylation to Oncogenic Gene Expression in Acute Myeloid Leukaemia. Nature 543 (7644), 265−269.

33. Solomon, M. J., Larsen, P. L., and Varshavsky, A. (1988) Mapping proteinDNA Interactions in Vivo with Formaldehyde: Evidence That Histone H4 Is Retained on a Highly Transcribed Gene. Cell 53 (6), 937−947.

34. Hattori, N., Niwa, T., Kimura, K., Helin, K., and Ushijima, T. (2013) Visualization of Multivalent Histone Modification in a Single Cell Reveals Highly Concerted Epigenetic Changes on Differentiation of Embryonic Stem Cells. Nucleic Acids Res. 41 (15), 7231−7239.

35. Kowarz, E., Löscher, D., and Marschalek, R. (2015) Optimized Sleeping Beauty Transposons Rapidly Generate Stable Transgenic Cell Lines. Biotechnol. J. 10 (4), 647−653.

CHAPTER 3

TANDEM HISTONE-BINDING DOMAINS ENHANCE THE ACTIVITY OF A SYNTHETIC CHROMATIN EFFECTOR

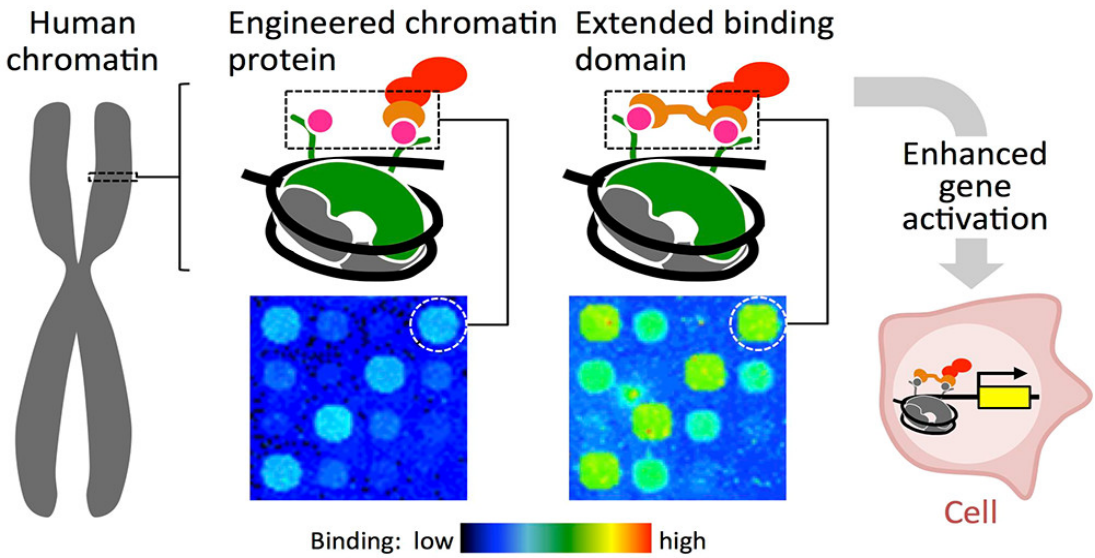


Figure 3.0 Abstract Figure.

Introduction

The discovery of histone post-translational modifications (PTMs) and the peptides that specifically interact with these marks has enabled scientists and cell engineers to manipulate chromatin, the DNA−protein structure that regulates gene expression states in eukaryotic cells. Structure-based models have informed targeted knockdown of chromatin subunits and the rational design of low molecular weight inhibitor compounds (reviewed in ref 1). DNA-binding domains fused with structural chromatin proteins and histone-modifying enzymes have been used to generate ectopic chromatin conformations at specific loci (2-3). Until recently, scientists had not yet leveraged PTM-binding peptides from natural effector proteins to "read" the rich biological information encoded

in histone marks in living cells. Peptides that recognize specific histone PTM signals are essential for synthetic systems that integrate epigenetic regulatory signals. In order to use PTM binding peptides in synthetic fusion proteins, the peptides must be portable, that is, maintain their intrinsic function within a new protein sequence. Early studies established important foundational knowledge by demonstrating that the interaction of the chromodomain motif (CD) with trimethylated histone H3 lysine 27 (H3K27me3) is an intrinsic activity that is maintained by the CD in the context of recombinant, fusion proteins (4-5). Other protein folds including the bromodomain (BRD) and plant homeodomain finger (PHD) function as isolated peptides (6−8) and within fusion proteins (7, 9) to specifically interact with acetylated histone lysines (BRD) and H3K4me3 (PHD).

We constructed the polycomb-based transcription factor (PcTF) using a histone PTM-binding motif from the natural protein CBX8 (10). The CBX8 effector protein binds to histone H3 trimethylated at lysine 27 (H3K27me3) through its Nterminal polycomb chromodomain (PCD) and establishes a silenced transcriptional state. Expression of PcTF, an artificial transcriptional activator with an N-terminal PCD, mCherry tag, and C-terminal VP64 activation domain, led to increased expression of H3K27me3-enriched genes in three different cancer-derived cell lines (11) These results show promise for designing transcription factors that can read chromatin marks to rewire aberrant epigenetic programming. However, binding affinities observed in vitro for isolated PCDs is poor, reported as 5 to >500 μM (12-13) compared to DNA-binding domains with target affinities in the pico- to nanomolar range such as transcription activator-like effectors (TALEs, ∼3−220 nM) (14), zinc fingers (∼0.01−16 nM) (15, 16),

and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated domains (Cas) (~0.5 nM) (17). In other work, we observed stronger gene upregulation when mCherry-VP64 was targeted to a promoter via a Gal4 DNA binding domain compared to the PCD histone-binding domain (10) PcTF mediated gene activation is dose dependent (11) and high PcTF expression levels are required for optimal activity. This limits the usefulness of PcTF for therapeutic applications where barriers to delivery severely limit the number of proteins that ultimately reach the nuclei of target cells. Although pharmacokinetic barriers to DNA and protein delivery in vivo are not trivial, increasing the effective dose of PcTF could significantly advance this technology toward clinical use.

The appearance of tandem histone binding domains within natural proteins suggests that the performance of histonebinding regulators can be customized and tuned through multivalency, defined as contact with more than one histone PTM via multiple domains (reviewed in refs 3, 18, and 19). Multivalent chromatin proteins can engage adjacent PTMs within a single histone tail, such as K4me3 and R8me2 on histone H3 bound by Spindlin1 (20) or K5ac and K12ac on histone H4 bound by TAFII250 (21). PTM targets can also reside on two distinct histone tails, such as H4K16ac and H3K4me3 bound by BPTF (7). Dual recognition of histone PTMs is accomplished by tandem protein motifs within the histonebinding protein. Comparisons of natural mono- and divalent proteins (22) as well as histone peptide on- and off-target binding studies (20, 23) have produced compelling evidence that tandem motifs contribute to avidity and specificity. The idea that combinatorial avidity allows proteins to read a “histone code” has been the topic of some controversy. Until recently, multivalency had not been

demonstrated using a rationally designed composition of binding domains. Tandem histone binding domains have been used to design protein probes to fluorescently label regions that are enriched for specific histone modifications (24-25) To date, multivalency has not been used to design a transcriptional regulator and tandem PCDs have not been reported. In order to compensate for the modest affinity of the CBX8 PCD (12) for its target, we added a second copy of H3K27me3-binding PCD to the N-terminus of PcTF to build Pc_2TF. Here, we demonstrate that Pc_2TF shows stronger avidity for H3K27me3 in vitro. This activity corresponds with enhanced activation of an H3K27me3-repressed gene in cultured cells. Our results have important implications for building and tuning fusion proteins that target sites of polycomb-mediated silencing, which plays a central role in cancer and stem cell plasticity.

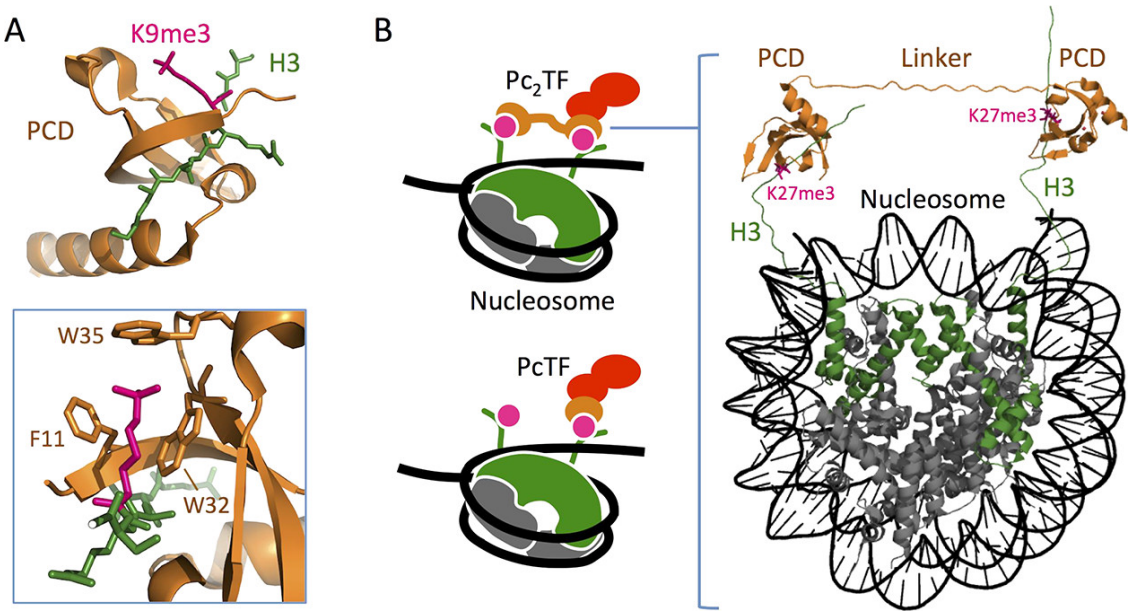


Figure 3.1 Three-dimensional model layout to show the plausibility of Pc2TF binding to adjacent H3K27me3 marks. (A) PCD (CBX8) in complex with trimethyl lysine (PDB 3I91) (37). Three residues form a hydrophobic cage and surround the Kme3 moiety (inset). (B) H3K27me3 recognition by synthetic fusion proteins that carry a single or

tandem PCD domains (PcTF and Pc_2TF, respectively). The 3D rendering was composed in the PyMOL Molecular Graphics System, version 1.3, Schrödinger, LLC (https://www.pymol.org), using data for CBX8/H3K9me3 (PDB 4X3K) (51) and a whole nucleosome assembly (PDB 5AV8) (51-52) from the Protein Data Bank.

Design of a Bivalent Synthetic Chromatin-Based Transcriptional Regulator

We designed the Pc_2TF protein to simultaneously recognize two copies of the histone posttranslational modification H3K27me3. The polycomb chromodomain motif (PCD) consists of three β strands packed against a C-terminal α helix and a hydrophobic pocket formed by three aromatic residues that interact with a methyl-lysine side chain (13, 26) (Figure 3.1A). The arrangement of histones within the nucleosome octamer suggests that Pc_2TF might bind adjacent trimethylated H3K27 residues. A single nucleosome includes eight individual histone proteins. The central tetramer contains two copies of histone H3 and H4. The H3 proteins are oriented in cis so that the unfolded N-terminal tails protrude away from the nucleosome in the same direction (27) (Figure 3.1B). One or both H3 tails (28) can become trimethylated at lysine 27 by the enzyme enhancer of zeste (EZH) (29). Therefore, tandem PCDs in the multivalent protein Pc_2TF might interact with two histone post translational modifications (PTMs) in a single nucleosome (Figure 3.1B) or single PTMs on adjacent nucleosomes.

To quickly and efficiently identify a linker that would allow contact of each PCD with an H3K27me3 ligand, we used an in vitro expression and ELISA procedure to test four Pc_2TF variants. Different lengths and physical characteristics were explored by using flexible glycine−serine linkers (30) and rigid α- helical (31-32) linkers. Glycine and serine, amino acids with small side chains, have been used in a wide range protein

engineering applications to build linker peptides that have minimal interference with the function of tethered proteins. However, as was demonstrated by mutagenesis of a rigid linker in the bivalent protein BPTF, added flexibility can destabilize protein−histone interactions (7). Rigid linkers might perform better by stabilizing the distance between PCDs to support interactions with neighboring K27me3 moieties (33-34). The Pc_2TF constructs included two tandem copies of the PCD separated by one of four linkers: flexible $(GGGGS)_4$, long flexible $(GGGGS)_{16}$, rigid $(EAAAR)_4$, and long rigid $(EAAAR)_{16}$. Based on a simplified layout of the interacting components (PCDs and a nucleosome carrying two H3K27me3 modifications; Figure 3.1B), we predicted that 20 amino acids would provide sufficient length for adjacent PCDs to bind simultaneously. The 80-amino-acid linkers were used to determine the impact of increased spacing between PCDs.

To expedite the prototyping stage, we used a cell-free expression system (35). Pc_2TF variants and a control protein with no binding domain ($Pc_{\Delta}TF$) were expressed from a pET28 vector (Figure 3.2A) in TXTL solution. Real-time detection of mCherry fluorescence in a Roche thermal cycler confirmed expression of recombinant proteins. For ELISAs, biotinylated histone peptides were immobilized on a neutravidin-coated 96 well plate. HRP-conjugated anti-mCherry was used for immunodetection of bound fusion proteins. Significantly higher HRP signal was detected compared to background (unmodified K27 and K27ac) for variants that contained the flexible $(GGGGS)_4$, long flexible $(GGGGS)_{16}$, and long rigid $(EAAAR)_{16}$ linkers (Figure 3.2C). The implications of these results are discussed in depth in the Conclusions. Assuming that HRP signal is proportional to Pc-fusion molecules bound, the flexible $(GGGGS)_4$ linker conferred the

strongest avidity in this assay. Therefore, we used this variant in subsequent experiments to determine the impact of bivalency on the activity of synthetic, histone-binding effectors.

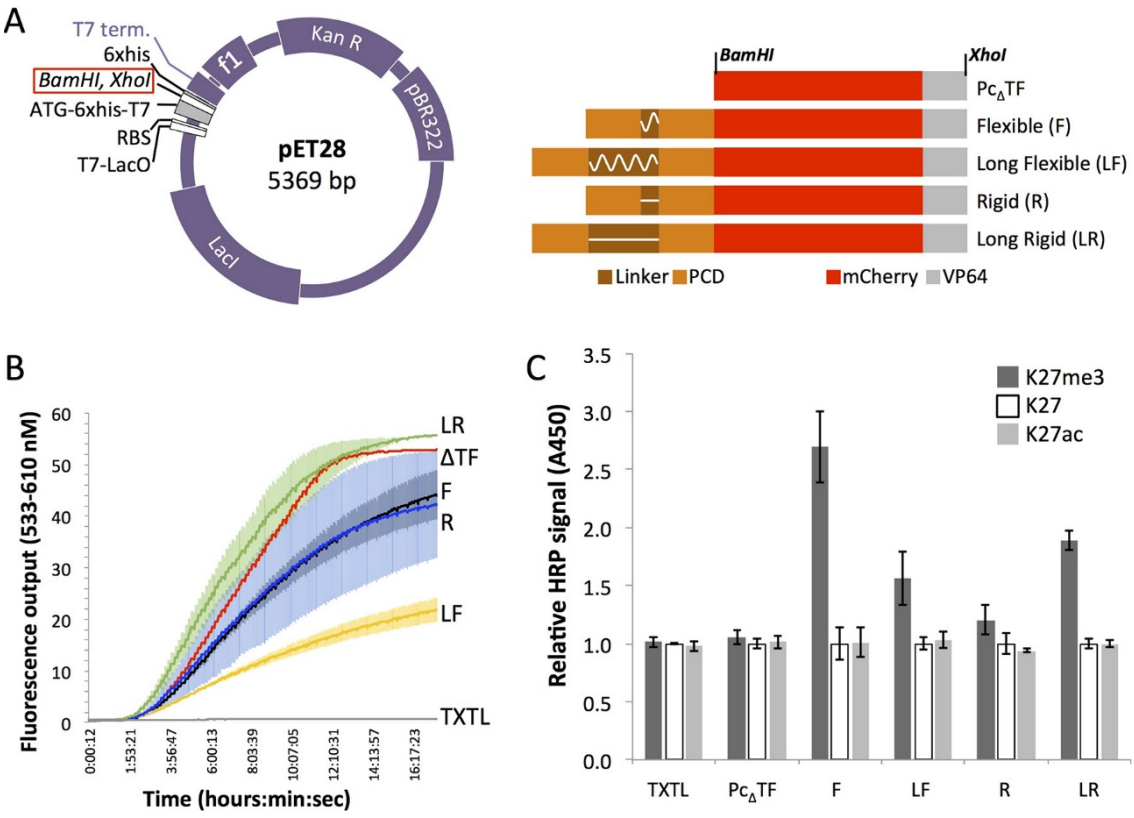


Figure 3.2 Comparison of Pc_2TF variants that were expressed in a bacterial cell-free expression system. (A) Map of the expression vector and open reading frames (ORFs). Fusion-encoding ORFs were cloned in the pET28 vector at *Bam*HI and *Xho*I. (B) Real-time detection of mCherry fluorescence was used to determine expression of recombinant protein in TXTL in a 96-well PCR plate in a Roche thermal cycler. Each replicate is an independent TXTL reaction in a single well (1 replicate for $Pc_\Delta TF$ and TXTL without DNA, 3 replicates for others). Replicates were pooled for ELISAs in panel C. Solid line = mean, shaded regions = SDM. (C) The bar chart shows mean signal from anti-mCherry-HRP signal at an absorbance of 450 nm (3 ELISA wells) from TXTL-expressed fusion

proteins or plasmid-free "blank" TXTL captured by tethered trimethyl-K27 (K27me3), unmodified (K27), or modified nontarget (K27ac) histone H3 peptides. For each TXTL product, individual values are normalized to the unmodified H3 mean value within the set (error bars = SDM).

Bivalency Strengthens the Avidity of the Pc-Fusion for H3K27me3

To compare known concentrations of Pc fusion proteins in subsequent experiments, we overexpressed and purified recombinant proteins from *E. coli*. Denaturing polyacrylamide electrophoresis (PAGE) of lysates from isopropyl β-D-1-thiogalactopyranoside (IPTG)-treated and untreated *E. coli* confirmed inducible production of the proteins at roughly the expected sizes: 37, 44, and 52 kDa for $Pc_{\Delta}TF$, PcTF, and Pc_2TF, respectively (Figure 3.3A). Nickel-NTA column-purified proteins were soluble in 1× phosphate buffered saline (PBS). The visible red hue under white light, which is typical of the mCherry protein (36), indicated proper protein folding (Figure 3.3A).

To determine impact of the additional PCD domain on PcTF avidity, we exposed tethered histone peptides to varying concentrations of soluble PcTF and Pc_2TF. Liquid phase ligand (H3K27me3 peptide) binding assays (fluorescence polarization, FP) reported by other groups have determined affinities of N-terminal PCD motifs from the Drosophila Pc protein (residues 1−90, $K_d = 5.0 \pm 1$ μM (13)) and mammalian CBX8 protein (mouse residues 1−62, $K_d = 165 \pm 20$ μM (12), human residues 8−61, $K_d > 500$ μM (37)). The amino acid sequence of the PCD in our fusion proteins (human CBX8 residues 1−62) is identical to the mouse ortholog. To acquire data that is relevant to the full-length fusion proteins (295 to 445 residues) that we had previously tested as gene

regulators in cancer cell lines (11), we used a histone peptide microspot array. We used a mathematical model to predict relative RFP signal levels after Pc-fusion binding and subsequent washing of the microarray. PcTF has a higher mCherry/PCD ratio than Pc_2TF (1:1 versus 1:2). Therefore, PcTF should show higher relative RFP signal when all targets (H3K27me3 peptides) within a microspot are saturated by PCD binding (Figure 3.S1). Assuming that bivalency supports an additive increase in avidity, a higher fraction of Pc_2TF molecules should remain bound at the microspot during washing, resulting in higher total RFP signal. Results from a test array were consistent with this prediction (Figure 3.3B). We tested concentrations of the recombinant proteins over 2 orders of magnitude (0.2−20 μM) to determine the apparent dissociation constant (K_d^{app}) of each protein for 10, 20, and 50 μM tethered H3K27me3 ligand (Figure 3.S2). We detected no interaction with unmodified histone H3 peptides and very little signal above background for the $Pc_{\Delta}TF$ negative control. The K_d app of monovalent PcTF was 5.14−8.95 μM for four independent trials (Figure 3.3C and 3.S2). The micromolar K_d app values are comparable to K_d values from the aforementioned FP experiments, although the wash steps in microspot assay may bias K_d app toward the off kinetics of the binding process. We conclude that PCD retains its intrinsic affinity for H3K27me3 as an N-terminal motif within a fusion protein.

Overall, analysis of the microspot array data suggest that at 10 and 20 μM of H3K27me3 the K_d^{app} of Pc_2TF is roughly 2- fold smaller than PcTF (Figure 3.S2). Assuming that the second PCD fold (PCD2) maintains its intrinsic affinity, PCD2 should approximately double the overall association rate for Pc_2TF since there is twice the chance of a PCD−H3K27me3 collision. Avidity is related to the inverse of the

equilibrium constant, and the equilibrium constant is proportional to the ratio of association rate over dissociation rate. Thus, the effect we observed is most likely due to increasing the association rate or decreasing the dissociation rate, which would decrease the K_d^{app} value (compared to PcTF) roughly 2-fold.

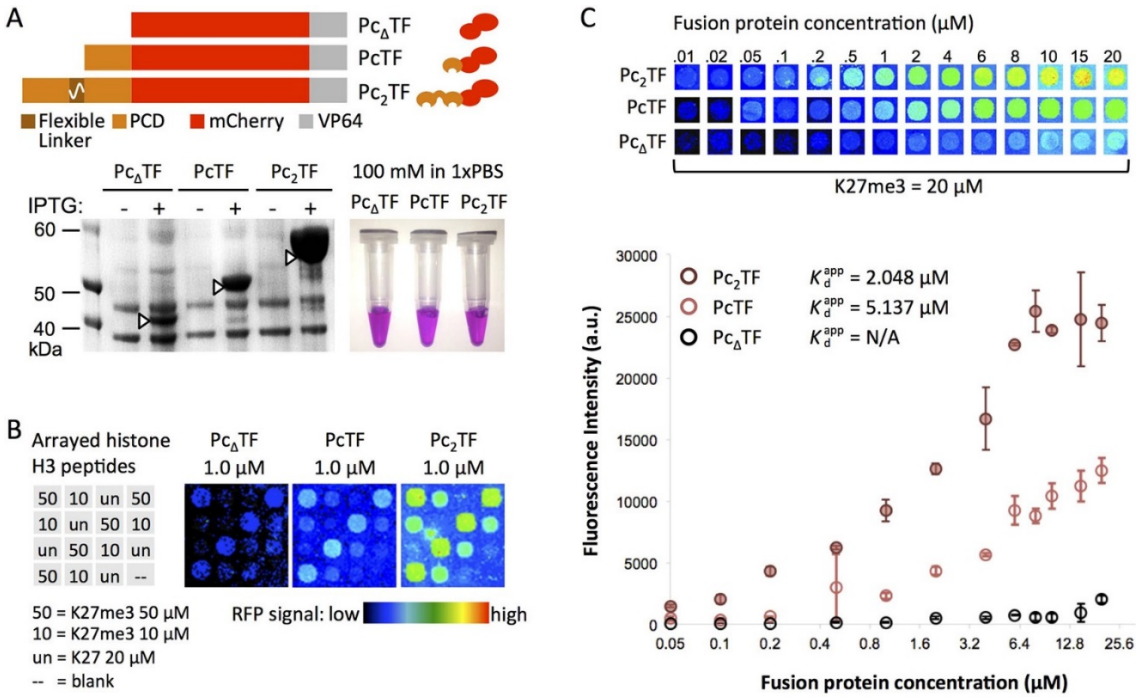


Figure 3.3 A bivalent PCD fusion peptide shows enhanced avidity for H3K27me3 in microspot array experiments. (A) For high-yield expression, *E. coli* was transformed with pET28 plasmids encoding Pc$_\Delta$TF (negative control), PcTF (single PCD), and the Pc$_2$TF containing the flexible linker $(GGGGS)_4$. Native polyacrylamide gel electrophoresis (PAGE) of overexpressed proteins purified from *E. coli*. (B) Test slides were spotted with histone H3 peptides (K27me3 or unmodified K27) as indicated in the grid for qualitative analysis. Pseudocolored images show mCherry signal after an application of 1.0 µM fusion protein to individual arrays. (C) New arrays were spotted with 10, 20, and 50 µM H3K27me3 for quantitative analysis. Fluorescence signal versus the concentration of fusion protein applied to the array was used to calculate the apparent dissociation

constant (K_d^{app}, not applicable for $Pc_{\Delta}TF$). Each point in the graph is the mean signal from four spots in one application (error bars = SDM). The data displayed in the graph are from representative applications (out of four total) for 20 μM immobilized H3K27me3.

Bivalent Pc_2TF Shows Cooperative, On-Target Binding with Solid Phase Target Ligands

Next, we investigated the binding properties of the mono- and bivalent PCD proteins over a range of target ligand densities to approximate dynamic distributions of H3K27me3 that may occur in chromatin. We assumed that random distribution within each mixture would decrease the spacing between H3K27me3 targets as their concentration was increased. We applied dilutions of the target ligand (0−100% H3K27me3 mixed with unmodified H3K27, 1000 nM final concentration) to ELISA wells and exposed the immobilized ligands to the highest concentration of fusion proteins that produced minimal background signal in preliminary ELISA trials (0.1 μM). At 0−15% H3K27me3, HRP signal for PcTF or Pc_2TF was not significantly greater than the negative control fusion protein ($Pc_{\Delta}TF$). In this range, the number of fixed H3K27me3 ligands may not have captured enough fusion proteins to yield detectable signal after washing. At 20−30% and higher, the HRP signal from the PcTF and Pc_2TF wells increased with H3K27me3 concentration.

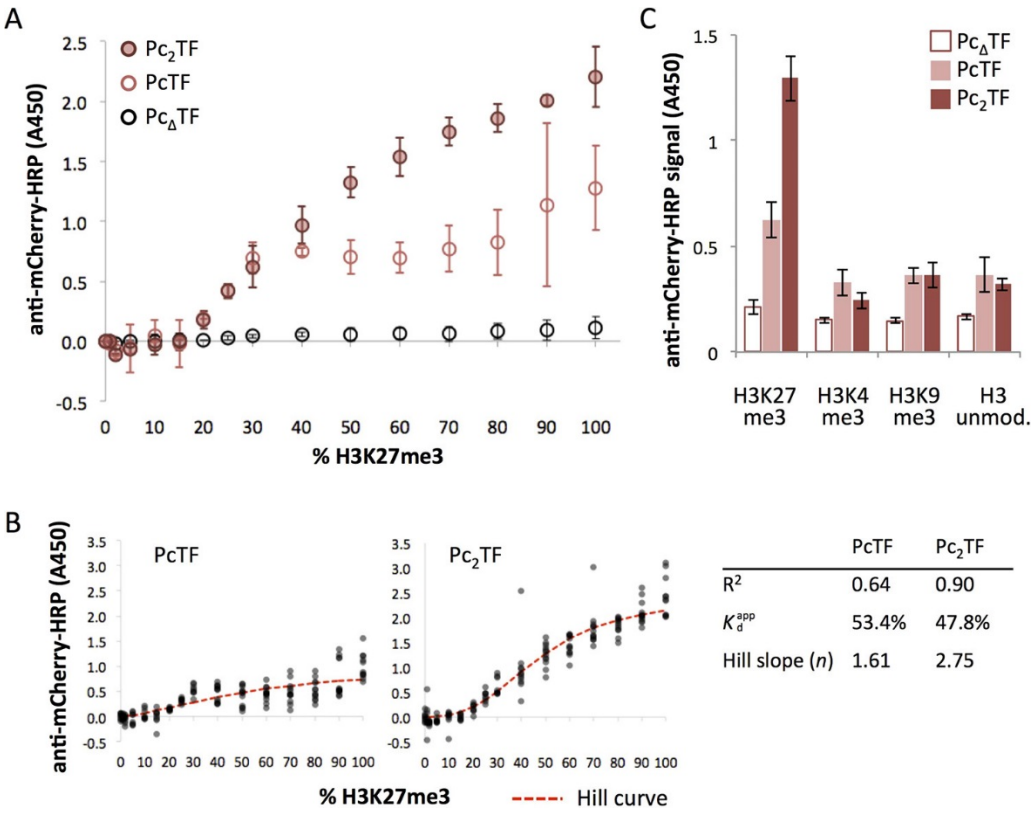
A Hill slope of 2.75 from the nonlinear regression ($R^2 = 0.90$) for Pc_2TF (Figure 3.4B) indicates that binding scales nonlinearly with the concentration of its ligand. It is difficult to fit a Hill curve to the data for PcTF ($R^2 = 0.64$) because the increase in HRP signal is interrupted by a plateau at 30−80% H3K27me3. The cause of the plateau is

unclear; however it is possible that the increase observed above 80% is due to the binding avidity between PcTF and H3K27me3 being exceeded at these concentrations. Overall, we can conclude from the ELISA data that Pc_2TF binding is cooperative.

To investigate ligand selectivity, 50 nM purified PcTF, Pc_2TF, or $Pc_\Delta TF$ was tested for interaction with histone peptides that were trimethylated at different lysine residues. PcTF and Pc_2TF showed significant binding with H3K27me3 peptides compared to the control protein $Pc_\Delta TF$ (Figure 3.4C). HRP signal from the H3K27me3 wells was significantly higher than what appears to be nonspecific binding with unmodified H3. This was not the case for off-target ligands H3K4me3 and H3K9me3, suggesting that the Pc-fusions can discriminate between the different methyl marks. No significant increase in HRP signal in the off-target wells was observed for Pc_2TF, suggesting that target preference was not lost as avidity was enhanced. One might expect cross-reactivity with H3K9me3 since this PTM appears within a similar motif (ARKS) as H3K27me3 (13, 37). Others have reported that in vitro, chromodomain peptides from different orthologues (CBX1−8) have varying preferences for the two histone modifications (12, 37-38) CBX8, the PCD used for PcTF in our work, has shown weak affinity for H3K27me3 and none for H3K9me3 (12, 37), which is consistent with our results.

The results from the assays with purified proteins led us to ask; what is the biological consequence of increased binding in living cells where the physical distribution of H3K27me3 is much different? In the cellular chromatin environment, H3K27me3 can occur in cis on the radial surface of a single nucleosome (Figure 3.1B), in trans where DNA bending brings the H3 tails of neighboring nucleosomes close together,

or sparsely distributed across many nucleosomes. Furthermore, H3K27me3 marks in living cells are dynamic. The enzyme EZH1/2 adds methyl groups to H3K27, and the enzymes KDM6A (UTX) and KDM6B (JMJD3) removes these marks (39). We set out to compare Pc_2TF to PcTF in a cellular milieu.



	PcTF	Pc_2TF
R^2	0.64	0.90
K_d^{app}	53.4%	47.8%
Hill slope (*n*)	1.61	2.75

Figure 3.4 Bivalent Pc_2TF shows cooperative and on-target binding with H3K27me3 ligands. (A) The scatter plot shows mean HRP signal at an absorbance of 450 nm (one ELISA trial, means of four technical replicate wells, error bars = SDM) from wells in which 0.1 μM purified protein (Pc_2TF, PcTF, or $Pc_\Delta TF$) was allowed to bind with different proportions of H3K27me3 biotinylated peptides (0–100%) mixed with unmodified H3 and tethered to neutravidin-coated surfaces. (B) Hill curves were fit to data for three ELISA trials (dots = technical replicate wells from all trials). (C) ELISA was used to detect interaction of 0.05 μM purified proteins with immobilized histone H3

peptides that were trimethylated at lysine 27, 4, or 9 or unmodified. The bar chart shows mean signal values from anti-mCherry-HRP at an absorbance of 450 nm (4 technical replicates, bars = SDM).

Bivalent Pc_2TF Activates a Target Gene in a Partially Silenced State

Previously, we demonstrated that PcTF activated a reporter gene near ectopic H3K27me3 in HEK293 cells (10). Here, we determined the biological significance of PCD bivalency by comparing the gene-regulation activities of Pc_2TF and PcTF at the same reporter. Doxycycline (dox)-mediated induction of Gal4-EED in HEK293 Gal4-EED/ luc cells leads to accumulation of H3K27me3 at and silencing of a chromosomally integrated Tk-luciferase transgene (Figure 3.5A) (40-41) Tk-luciferase repression reaches steady state at 96 h (40), and repression is maintained by epigenetic inheritance after loss of Gal4-EED (41).

We transfected dox-treated cells (96 h) with the $Pc_{\Delta}TF$ negative control, PcTF, or Pc_2TF cloned into a mammalian expression vector (Figure 3.5B). Fluorescence microscopy and Western blots confirmed nuclear localization and expression of full length proteins. Using reverse transcription followed by quantitative PCR (RT-qPCR), we detected higher luciferase transcript levels in Pc_2TF and PcTF-expressing cells compared to $Pc_{\Delta}TF$ (Figure 3.5C). These results indicate that bivalent Pc_2TF is a stronger activator than PcTF. Luciferase (luc) activity levels detected by an enzymatic assay (Figure 3.5D) and normalized to RFP signal-to-noise ratios from flow cytometry corroborated the RT-qPCR results; Pc_2TF stimulated greater luc expression than PcTF. We did not detect significantly higher luc activity for PcTF-expressing cells versus the

negative control in this trial (Figure 3.5D), but did so in additional experiments (Figure 3.5E).

Expression of fusion regulators in cells that were treated with dox for 0, 48, and 96 h showed that Pc_2TF had roughly twice the activity as PcTF (96 h) and that Pc_2TF activated Tk-luciferase without prior dox-induced silencing (Figure 3.5E). The latter result can be explained by an intermediate, partially silenced level of Tk-luciferase expression (40) compared to fully active Tk-luciferase in a "Luc14" parental cell line that lacks the Gal4-EED gene, as previously observed (40). H3K27me3 was detected via chromatin immunoprecipitation (ChIP) coupled with qPCR (ChIP-qPCR) near the luciferase promoter (Tk) in uninduced Gal4-EED/luc cells at significantly higher levels than in Luc14 cells. Dox treatment resulted in a further decrease in Tk-luciferase expression and a significant increase in H3K27me3 accumulation. In the experiments reported here, basal Tk-luciferase expression (Figure 3.5A) agrees with independent experiments from our prior study (0.02−0.07 luciferase activity per cell, au) (40). The uninduced state may have low levels of H3K27me3 at nucleosomes near the reporter gene in all cells or high levels of H3K27me3 at the reporter gene in a small proportion of cells in the population. In contrast to Pc_2TF, monovalent PcTF only activated Tk-luciferase after silenced chromatin had been induced for 96 h. These results suggest that Pc_2TF is more tolerant of low levels of H3K27me3 in cellular chromatin.

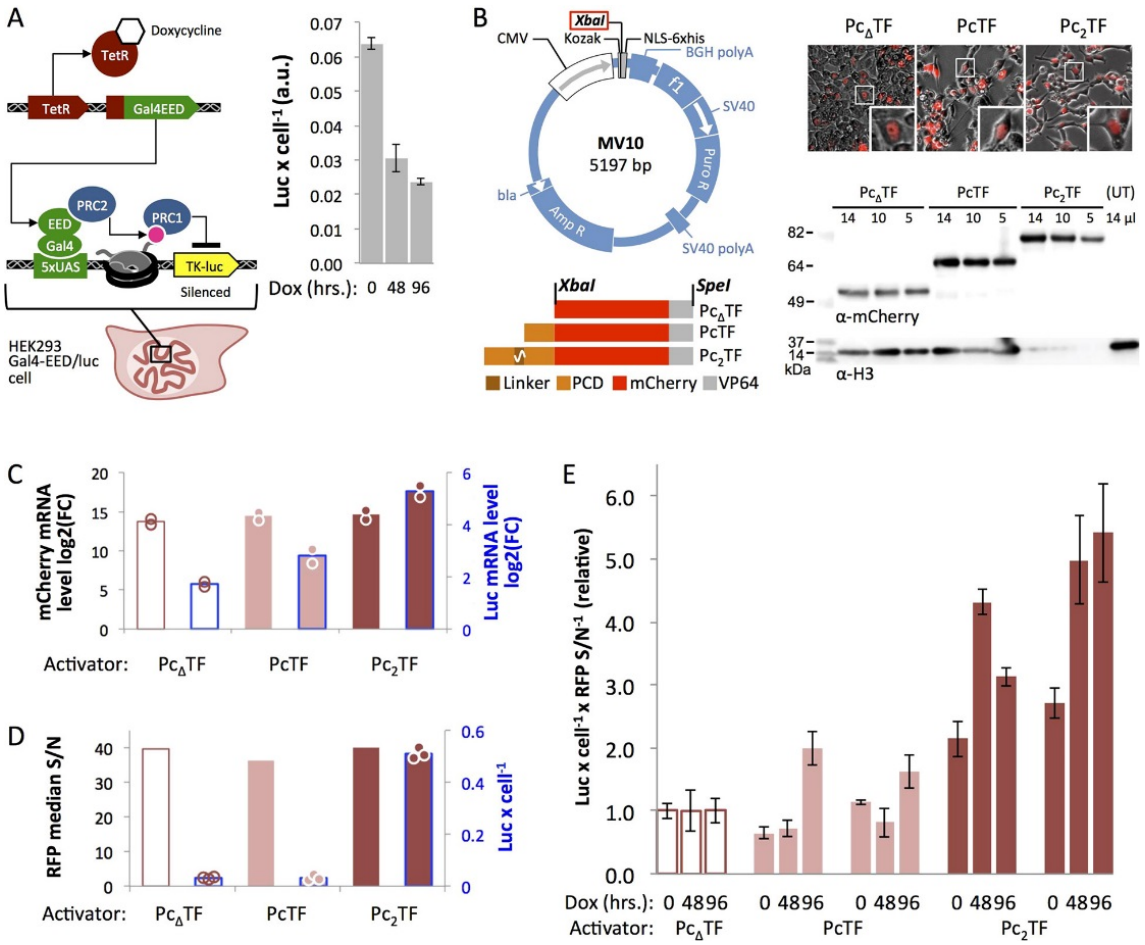


Figure 3.5 Pc_2TF stimulates expression at a polycomb-silenced reporter gene. (A) An engineered HEK293 cell line, Gal4-EED/luc, was used for doxycycline-mediated control of H3K27me3 and PRC-mediated silencing at a *Tk-luciferase* reporter. Expression is partially silenced prior to dox treatment, as demonstrated previously (40) and becomes fully repressed at 96 h. (B) Fusion constructs were cloned into the MV10 vector at *Xba*I. Fluorescence microscopy confirmed nuclear localization of the fusion proteins in transfected cells. The same samples were used for Western blots to confirm cellular expression of full length proteins, RT-qPCR to measure mRNA levels (C), and flow cytometry to measure RFP signal (D). Changes in the expression states of *Tk-luciferase* in 96-h dox-treated cells were determined by RT-qPCR (C) and luc activity assays (D)

(circles = replicates, described in Methods). (E) Fusion proteins were expressed in cells treated with dox for 0, 48, or 96 h to determine the activity of the fusion activators at intermediate repressed states (bars = mean values for 3 luciferase assays from one transfected sample, each scaled to mean $Pc_{\Delta}TF$ luc/cell; error bars = SDM).

Conclusions

This report presents the first demonstration of modular, synthetic multivalency of a chromatin-derived histone-binding protein with gene-regulating activity. In our previous work, we have demonstrated the use of a monovalent synthetic effector to activate chromatin-silenced genes in live cells. Natural bivalent chromatin proteins that recognize two histone posttranslational modifications at once suggest a broader design space for synthetic chromatin effectors. Our application of bivalency to design a synthetic fusion protein produced two important advances for engineering synthetic chromatin effectors. First, we determined that synthetic linkers allow tethered histone PTM-binding peptides to function within the context of a fusion protein in vitro and in live cells. Second, we have established that doubling the valency with tandem PCDs strengthens avidity and increases gene regulation activity by at least 2-fold.

Here, we demonstrated that different synthetic linkers allow tethered histone PTM-binding peptides to bind in vitro to varying degrees. We observed weaker binding for the longer flexible linker (80 amino acids) compared to the shorter linker in our ELISA experiment. This result is likely due to lower production of the long flexible linker variant in TXTL. Given that both variants showed binding above background, GGGGS repeat number may not significantly affect bivalent PCD engagement with H3K27me3 in vitro. For the rigid linker tethered PCDs, only the longer length (80 amino

acids) appeared to support binding. Assuming that this variant protein was properly folded, lack of binding over background for the shorter EAAAR-repeat variant could be caused by suboptimal rotation, that is, in trans instead of in cis, of the second PCD away from the 2-D binding surface in the ELISA well. This mechanism was demonstrated with mutated α-helical linkers in bivalent BPTF7 and with tandem zinc finger DNA binding domains (32). In the context of cellular chromatin where looping and folding occurs, H3K27me3 would not necessarily be constrained to one face of the Pc_2TF protein. Valuable insights and perhaps greater Pc_2TF performance might be acquired by exploring additional linker variants in cells as well as *in vitro*. Such work is beyond the scope of the studies reported here, which accomplished a major step by identifying a functional bivalent synthetic effector protein that specifically interacts with its target H3K27me3.

We have established that tandem PCDs strengthen avidity for H3K27me3 in a cooperative manner in vitro and increase gene regulation activity in live cells by at least 2-fold compared to a monovalent PCD. The wide distribution of multivalency within bromodomain family (6) and other effector proteins (7) suggests that multivalent engagement has an important, evolutionarily conserved biological role. Multivalency appears to largely be represented by cell-cycle and gene-activating effectors. Relatively few multivalent proteins that recognize silencing marks have been studied in biophysical detail. Examples include the chromodomains of the Arabidopsis protein CMT342 and the mammalian protein HP1β (43) as bivalent homodimers, these proteins show enhanced interaction with their respective ligands H3K9meK27me and H3K9me3. Pc_2TF is novel in its composition of histone-binding motifs: adjacent, identical polycomb

chromodomains within a single peptide. Therefore, its activity in vitro and in cells provides new insights into the recognition of histone marks by effector proteins.

In the context of cellular chromatin, Pc_2TF appears to be active at the target gene prior to full repression (Figure 3.5E, 0 h dox), whereas detectable activity of monovalent PcTF required a prolonged period of induced repression at the target (Figure 3.5E, 96 h dox). Our previous ChIP mapping data (40) confirm that compared to the fully active and fully silenced states, intermediate levels of H3K27me3 appear at Tk-luciferase (on average) without the addition of doxycycline. It is likely that in the pretreated state, leaky Gal4-EED expression causes a few cells in the population have one or two H3K27me3 marks at a nucleosome near the Tk promoter. Stronger avidity, supported by the additional PCD module, may increase the likelihood of an activation event at the target in this small population of cells. This idea is consistent with the behavior of synthetic zinc finger-based DNA-binding regulators, where stronger affinity of the regulator for its DNA target is associated with stronger gene activation (44). Similar behavior can also be observed for multivalent receptor-binding peptides, which bind with high avidity and specificity to a small number of receptor-positive cells (45).

Further engineering efforts to achieve greater nonlinear enhancement of PcTF/Pc_2TF may require changes within the PCD binding motif. The hydrophobic interaction between the methylammonium cage and the methyl-lysine moiety (Figure 3.1A) depends upon proper positioning of PCD residues that appear discontinuously in the primary sequence; this positioning requires specific intramolecular contacts of peptide residues within the PCD fold. Reverse engineering and de novo design of a new binding pocket through randomization of sequences would likely yield many nonfunctional

proteins. K27-adjacent interactions that contribute to interactions with the histone tail (13) could be leveraged to enhance affinity. However, increasing the stability by introducing additional hydrogen bonding could overwhelm the hydrophobic, K27me3-specific interaction and allow PCD to recognize unmodified tails or off target modifications. Trade-offs between affinity and specificity pose formidable challenges to enhancing PCD affinity. Therefore, the most practical strategy for identifying alternative PCDs is to leverage H3K27me3-specific orthologues and paralogues from various species (46). It will be important to determine cross-reactivity with different histone modifications since certain CBX PCD peptides have been shown to bind H3K9me3 (12).

Multivalent engagement of combinatorial histone marks has recently become a key line of evidence to support the controversial histone code hypothesis. Rationally designed synthetic multivalency will advance this important area of research by exploring functions beyond the limits of preexisting natural multivalent proteins. Furthermore, engineered chromatin effectors provide a practical tool to support artificial regulation of gene expression states through direct engagement with highly conserved components of chromatin, that is, histone tails and their modifications. Therapeutic synthetic gene regulators that leverage this mechanism could help circumvent the shortcomings of epigenetic inhibitors, which target chromatin enzymes that can gain drug-resistant mutations (47-48). In conclusion, our findings demonstrate that synthetic biology is a powerful tool for fundamental investigations of chromatin biology and epigenetic engineering.

Methods

Plasmid Constructs for TXTL and Bacterial Expression. Constructs (Figure 3.2A, Figure 3.3A) were assembled as BioBrick compatible fragments in vector V0120 (49). Fragments were PCR amplified with Phusion polymerase using primers 1−6 (Table 3.S1) and a protocol adapted from New England Biolabs Phusion High Fidelity DNA Polymerase (98 °C 0:45, [98 °C 0:10, 67 °C 0:20, 72 °C 0:45] × 25, final extension of 5:00), column purified (Qiagen PCR Cleanup Kit), and double-digested with BamHI and XhoI (New England BioLabs). BamHI/XhoIdigested inserts and 50−75 ng of BamHI/XhoI linearized pET28(+) vector were ligated at a 3:1 molar ratio in a 20 μL reaction as described in the New England BioLabs (NEB) protocol for T4 ligase (M0202). Five microliters of each ligation was incubated with 50 μL of Turbo competent DH5- alpha E. coli (NEB) on ice for 5 min, transferred to 42 °C for 45 s, then to ice for 5 min, and allowed to recover in 350 μL of SOC medium at 37 °C with shaking for 30 min. Pelleted cells were resuspended in 50 μL of SOC, plated on LB agar (50 ug/ mL kanamycin), and grown at 37 °C overnight. Colony PCR was performed to identify positive ligation results using primers 6 and 7 (Table 3.S1) and the GoTaq Promega protocol. Plasmids were cloned, extracted (Sigma GeneElute Plasmid Miniprep Kit), and Sanger sequenced for verification prior to protein expression in cell-free TXTL or in *E. coli*. Annotated sequences for all pET28 constructs are available online at Benchling-Haynes lab: Synthetic Chromatin Actuators 2.0 (https:// benchling.com/hayneslab/f_/rmSYkAAU-synthetic-chromatinactuators- 2-0).

TXTL: Cell-free Expression. TXTL reactions were set up with the following conditions as previously described (35). 9 μL of lysate, 10 nM final template vector, 0.5

nM σ70-T7 RNA pol vector to a total of 12 μL. A Roche Lightcycler 480 was used to detect mCherry fluorescence with the following protocol: 29 °C for 10 min, bring to 30 °C for 1 s, scan 533−610 nm, repeat 96 times (total 16 h).

E. coli Expression and Purification of Proteins. All selection media contained 50 μg/mL kanamycin. $Pc_{\Delta}TF$, PcTF, and Pc_2TF in pET28 were transformed into Rosetta 2pLys DE3 cells and plated on LB agar and grown at 37 °C overnight. The next day, a single colony from each was used to inoculate 50 mL of LB and grown overnight at 37 °C at 300 rpm. The next day, 1 L of LB in a baffled Erlenmeyer flask was inoculated to an OD_{600} of 0.1. The cultures were grown to an $OD_{600} = 0.6$, induced with IPTG (1 mM final concentration), and allowed to express $Pc_{\Delta}TF$ and PcTF at 37 °C for 5 h with shaking (220 rpm). Pc_2TF expression was carried out overnight at room temperature with shaking (220 rpm) to aid solubility of the protein. Cell disruption and protein purification are described in detail in Supporting Information and Methods. Purification of recombinant protein from *E. coli* is described in Supporting Information (Methods).

Enzyme-Linked Immunosorbent Assays (ELISAs). All steps were carried out at room temperature except specifically noted, and all incubations and washes were agitated at 800 rpm on an Eppendorf Thermomixer R. Clear bottom plates (Greiner bio-one #655101) were coated in 50 μL of 20 ng/ μL neutravidin in PBS pH 8.0 overnight at 4 °C. The plates were washed the next day 3× with 200 μL of 0.2% PBS-Tween (PBST) with 5 min of shaking at 800 rpm between washes. The plate was blocked for 30 min at 800 rpm at room temperature with 200 μL of 5% BSA in 0.2% PBST followed by 3× washes of 200 μL of 0.2% PBST for 5 min each at 800 rpm. Fifty microliters of 1 μM biotinylated peptides (Anaspec, H3 (21− 44), H3K4me3 (1−21), H3K9me3 (1−21),

H3K27me3 (21− 44), or H3K27Ac (21−43)) in 0.2% PBST) were incubated at room temperature for 1 h at 800 rpm, followed by 3× washes of 200 μL of 0.2% PBST for 5 min at 800 rpm. The plate was blocked with 200 μL of 5% skim milk in 0.2% PBST (room temperature, 800 rpm, 30 min). TxTL, 1.5 μL (Figure 2), or 50 μL of 0.1 μM (Figure 3.3A) or 0.05 nM (Figure 3.3C) purified proteins in 50 μL of 5% skim milk in PBST were incubated in each well for 1 h (room temperature, 800 rpm). The wells were washed 3× with 200 μL of 5% skim milk in PBST with 5 min of 800 rpm shaking. After adding 100 μL of 1:3000 chicken polyclonal anti-mCherry (Novus Biologicals #NBP2−25158) in 5% nonfat milk in 0.2% PBST, wells were incubated for 1 h, followed by 3× of 200 μL of 5% skim milk in 0.2% PBST for 5 min each (room temperature, 800 rpm). After addition of 100 μL of 1:3000 rabbit anti-chicken−HRP (RCYHRP Genetel 0.5 mg/mL) in 5% skim milk in 0.2% PBST wells were incubated for 30 min (room temperature, 800 rpm). The plate was washed 5× with 200 μL of 0.2% PBST for 3 min each at 800 rpm. The plate was incubated with 100 μL of 1-step Ultra TMB-ELISA (Thermo-Fisher #34029) for 15 min while protected from light. Reactions were stopped with 100 μL of 2.0 M sulfuric acid, incubated for 2 min, and read at 450 nm. Each plate contained four technical replicates per H3K27me3 concentration per fusion protein. Two ELISA plates (trials) were run for each of two purified protein samples per construct. One trial failed to show significant signal over background (for all recombinant proteins) and was omitted from the final analysis. In Figure 3.4 “anti-mCherry-HRP (A450)” = HRP signal − mean HRP signal for 0% H3K27me3. The Microsoft Excel Solver tool was used to fit the Hill equation, $1 / ((K_d^{app} / [L])^n + 1)$ to the data by minimizing the sum of the squared errors between the equation and data (varying K_d^{app} and n). R^2 was calculated

as $1 - (SS_{reg}/SS_{tot})$, where total sum of squares $SS_{tot} = \Sigma_i(y_i - \bar{y})^2$ and regression sum of squares $SS_{reg} = \Sigma_i(f_i - \bar{y})^2$.

Peptide Microspot Arrays. APTES functionalized glass slides were coated with 200 μL of 1:1 (v/v) 40 mg/mL BS3 cross-linking solution and 1 mg/mL neutravidin with a cover slide (Thermo Scientific, #651-2-5251) and incubated overnight at 4 °C. The next day, the cover slide was removed, and the slide was rinsed 3× with 0.2% PBST for 5 min each. Slides were deactivated by incubation with $Na_2CO_3/NaHCO_3$ buffer, pH 9.4, for 30 min. The slides were quickly rinsed with ddH_2O and centrifuged to dry at 1200 rpm for 2 min. Slides were printed with biotinylated peptides (Anaspec) at concentrations of 10, 20, or 50 μM in 20% glycerol and PBS with a pin-printer (spot to spot distance = 600 μm) and incubated at room temperature for 1 h. The slide was rinsed with ddH_2O as described above and blocked with superblock for 1 h at room temperature. Proteins were diluted in superblock and incubated on the slide for 1 h at room temperature. The slides were rinsed with 0.2% PBST for 3 min each followed by quick rinsing with ddH_2O 3× and centrifuged dry (as described). Red fluorescent protein (mCherry) signal was detected at 50% gain and 50% intensity on a PowerScanner at 635 and 535 nm, 10 μm resolution. Slides were also scanned at 75%−75% and 100%− 100% to obtain a suitable signal-to-noise ratio. Arraypro software was used to quantify the median intensity values for each spot and background levels. Graphpad Prism software was used to fit the binding saturation nonlinear regression equation $y = (B_{max} * X) / (K_d^{app} + X)$ to the data, where B_{max} is the highest binding value, and X is the concentration of protein.

Plasmid Constructs for Mammalian Expression. MV10 was constructed from pcDNA3.1(+) (Invitrogen) with the following modifications. The CMV promoter was

removed via SpeI digestion and T4 ligase recircularization. A dsDNA fragment that encodes *Kozak* (ribosome binding site), *XbaI*, a nuclear localization sequence, 6x histidine, and a stop codon (5' cccgccgccaccatggagtctagacccaagaaaaagcgcaaggtacaccatcaccaccatcacgcgtaaagctgag) with *SpeI* overhangs at both ends (ctag/t) was inserted at *XbaI*. CMV (SpeI/XbaI fragment) was reintroduced upstream of Kozak at *SpeI*. Proper orientation of inserts was confirmed by Sanger sequencing. Constructs PcTF and Pc_2TF (Fig. 3.5B) were PCR-amplified (Phusion) with primers 9 and 10 (Table 3.S1), double-digested with XbaI and SpeI, and column-purified (Qiagen PCR Purification, 28104). Construct $Pc_{\Delta}TF$ (Fig. 3.5B) was double-digested with XbaI and SpeI (Thermo Fisher FastDigest) and isolated by electrophoresis and gel purification. XbaI/SpeI fragments and 25 ng XbaI-linearized, dephosphorylated MV10 vector were ligated at a 2:1 molar ratio in a 10 µL reaction as described in the Roche protocol for Rapid DNA Ligation (11635379001 Roche), using 1.0 µL NEB T4 ligase instead of the supplied enzyme. All 10 µL of each ligation was incubated with 50 µL Turbo competent DH5-alpha *E. coli* (New England Biolabs) on ice for 5 min, transferred to 45℃ for 45 seconds, then to ice for 5 min. Cells were plated directly on pre-warmed LB agar (100 ug/mL ampicillin) without recovery and grown at 37℃ overnight. Plasmid DNA was prepared (Sigma GeneElute Plasmid Miniprep Kit) from 5 mL cultures inoculated with single colonies. Forward orientation of the inserts was determined by XbaI and PstI double-digestion of prepped plasmids and Sanger sequencing. Annotated sequences for all MV10 constructs are available online at Benchling-Hayneslab: Synthetic Chromatin Actuators 2.0 (https://benchling.com/hayneslab/f_/rmSYkAAU-synthetic-chromatin-actuators-2-0).

Cell Culture and Transfection. HEK293 Gal4-EED/luc cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% tetracycline-free fetal bovine serum and 1% penicillin and streptomycin at 37°C in a humidified CO_2 incubator. Silencing of the reporter gene (*Tk-luciferase*) was induced by supplementing the media with 1 μg/mL of dox for 48 or 96 hours. For wash-out of doxycycline (to allow depletion of Gal4-EED), growth medium was removed and replaced with dox-minus medium supplemented with 0.5 μg/mL puromycin to select for the transgenic anti-Gal4-EED shRNA (41), and grown for 5 days. Prior to transfection, dox treated or untreated cells were plated in 12-well culture dishes at 40% confluency (~1.0E5 cells per well) in 2 mL pen/strep-free growth medium. Transient transfections were carried out by adding 100 μL of DNA/Lipofectamine complexes to each well: 1 μg pDNA or ddH_2O for mock transfections (10 μL), 3 μL Lipofectamine LTX (Invitrogen), 87 μL Opti-MEM. 48 hours after transfection, cellular mCherry (580/610 excitation/emission) was imaged in culture dishes on a Nikon Eclipse Ti wide field inverted fluorescent microscope (MEA53100) at 200x magnification (eyepiece = 10x; objective = CFI S Plan Fluor ELWD 20x, numerical aperture = 0.45), 25°C, without oil immersion, and with either phase contrast or an mCherry filter set (TE2000 cube, excitation FF01-562/40-25, emission FF01-641/75-25). Images from each channel were acquired with a digital monochrome camera (Coolsnap ES2 12 bit, 20 MHz) and overlaid using NIS-Elements software. For downstream assays (RT-qPCR, Western blots, and flow cytometry), the growth medium was removed, semi-adherent cells were gently collected with 1x PBS washes, pelleted (200 g, room temperature, 5 min) and resuspended in 1x PBS. Six replicate samples (wells) were pooled for assays in Figure 3.5B, C, and D.

RT-qPCR. Preparation of total RNA, cDNA synthesis, and qPCR were performed as previously described (11) using ~1.0E6 HEK293 Gal4-EED/luc cells that were pelleted (500 g, room temperature, 5 min) and lysed with 500 µL TRIzol (Thermo Fisher #15596026). DNA/LNA oligos for qPCR were: *mCherry* - forward 5'-cctgaagggcgagatcaag, reverse 5'-ttgacctcagcgtcgtagtg, LNA probe #41 (Millipore Sigma #04688007001); *luciferase* - forward 5'-caggtcttcccgacgatg, reverse 5'-gtctttccgtgctccaaaac, LNA probe #70 (Millipore Sigma #04688937001); *GAPDH* (reference) - Roche human G6PD assay (Millipore Sigma #5046246001). Mean Crossing point (Cp), the first peak of d^2y/dx^2 (fluorescence over cycle number), was calculated by the Roche LightCycler 480 software for three replicate wells per unique reaction. For each biological replicate (one transfection per fusion protein) two replicate cDNA synthesis reactions (from one RNA prep) were completed. Expression level was calculated as delta Cp = $2^{[Cpreference - Cp]}$. "mRNA level $\log_2$(FC)" = delta C_p transfected cells / delta C_p mock.

Western blots. Total protein was prepared from roughly 250,000 cells. Sample preparation, polyacrylamide gel electrophoresis (PAGE), and membrane blotting are described in detail in Supporting Information. Immunostaining was carried out with the following: blocking buffer - 5% nonfat dry milk in 1x PBST (1x PBS, 0.1% Tween-20); primary 1 - chicken polyclonal anti-mCherry, 1:2000 (Novus Biologicals #NBP2-25158); secondary 1 - HRP-conjugated rabbit anti-chicken, 1:2000 (Millipore Sigma #AP162P); primary 2 - anti-histone H3, 1:1000 (Abcam #ab1791); secondary 2 - HRP-conjugated goat anti-rabbit, 1:2000 (Cell Signalling Technology #7074). Immunostaining was performed at 4°C overnight (primary) or at room temperature for 1 hour (secondary) with

nutation in a Parafilm pouch (50). Immunostained blots were washed 4x 10 min in 1x PBST, with orbital shaking at room temperature. HRP signal was detected using the SuperSignal West Femto substrate kit (Thermo Fisher #34095) and a PXi4 imager (Syngene) with GeneSys software.

Imaging and flow cytometry. Cells were passed through a 35 μm nylon strainer (EMS #64750-25). Red fluorescent signal from mCherry was detected on a BD Accuri C6 flow cytometer (675 nm LP filter) using CFlow Plus software. Data were further analyzed using FlowJo 10.0. One run (~10,000 live cells, gated by forward and side scatter) was completed per sample. "RFP median signal/noise (S/N)" = median RFP signal from live RFP-positive cells / median RFP noise from live untransfected cells.

Luciferase Assays. Cell counts (per 100 μL) were determined by flow cytometry (BD Accuri C6). 100 μL of cells or 1x PBS (blank) were incubated with 100 μL of complete luciferase assay reagent as described in the protocol for the Biotium Firefly Luciferase Assay Kit (89138-960) and in previous work (40) in Corning and Costar 96-well Cell Culture Plates, opaque, white (Corning 3789A). Chemiluminescence was detected using a Synergy H1 Multi-Mode Reader (Biotek). Replicates included three samples (100 μL each) taken from a single population of transfected cells. "Luc x $cell^{-1}$(a.u.)" = [Sample Luciferase signal] − 1x PBS blank signal / [cell count × (100 μL/20 μL)]. For fusion protein-expressing cells, normalization was performed by dividing Luc x $cell^{-1}$ by the RFP median signal/noise value (from flow cytometry).

Associated Content

Table 3.S1 DNA templates and primers used in experiments

Template(s)	Primer Name	Primer Sequence (5'-...)
PcTF_V0120	1. PcTF.pET28.For.2	atgtcaGGATCCATGGAG CTTTCAGCGGTG
PcTF_V0120	2. PcTF.pET28.Rev	GCGCTTTTTCTTGGGCT CGAGCAACATGTCCAA GTCG
PcTF_pET28	3. mcVP64.For	aatgcctGGATCCATGGTG AGCAAGGGCGAGGA
PcTF_pET28	4. mcVP64.Rev	ATCTCAGTGGTGGTGG TGG
Pc2TF_V0120	5. DD.pET28.For	cataacaGGATCCGCGGCC GCATCTAGAATG
Pc2TF_V0120	6. DD.pET28.Rev	acttgggCTCGAGGCGGCC GCTACTAGT
pET28	7. T7.For	TAATACGACTCACTAT AGGGGAATTG
pET28	8. T7.Rev	GCTAGTTATTGCTCAG CGG
PcTF_V0120, Pc2TF_V0120	9. Biobrick For	TCACTGACTGACTGAC TGCGTCTCAA
PcTF_V0120, Pc2TF_V0120	10. Biobrick Rev	TTCCAGTCAGTCAGTC AGTCGTCTCTTG

Primers 1-6 were used for addition of restriction sites to pre-assembled constructs via Phusion PCR to build the bacterial/TXTL expression plasmids. Primers 7 and 8 were used for verification by Sanger sequencing. Primers 9 and 10 were used to amplify inserts for the MV10 vector. Non-binding overhangs are shown in lowercase. Restriction site nucleotides are underlined.

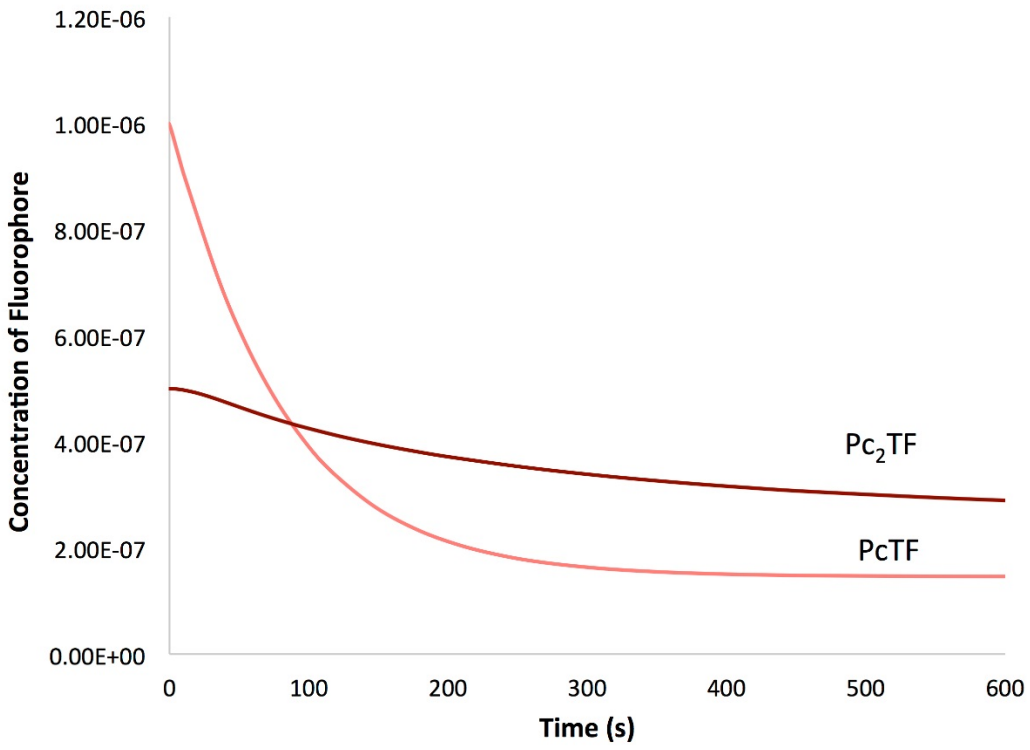
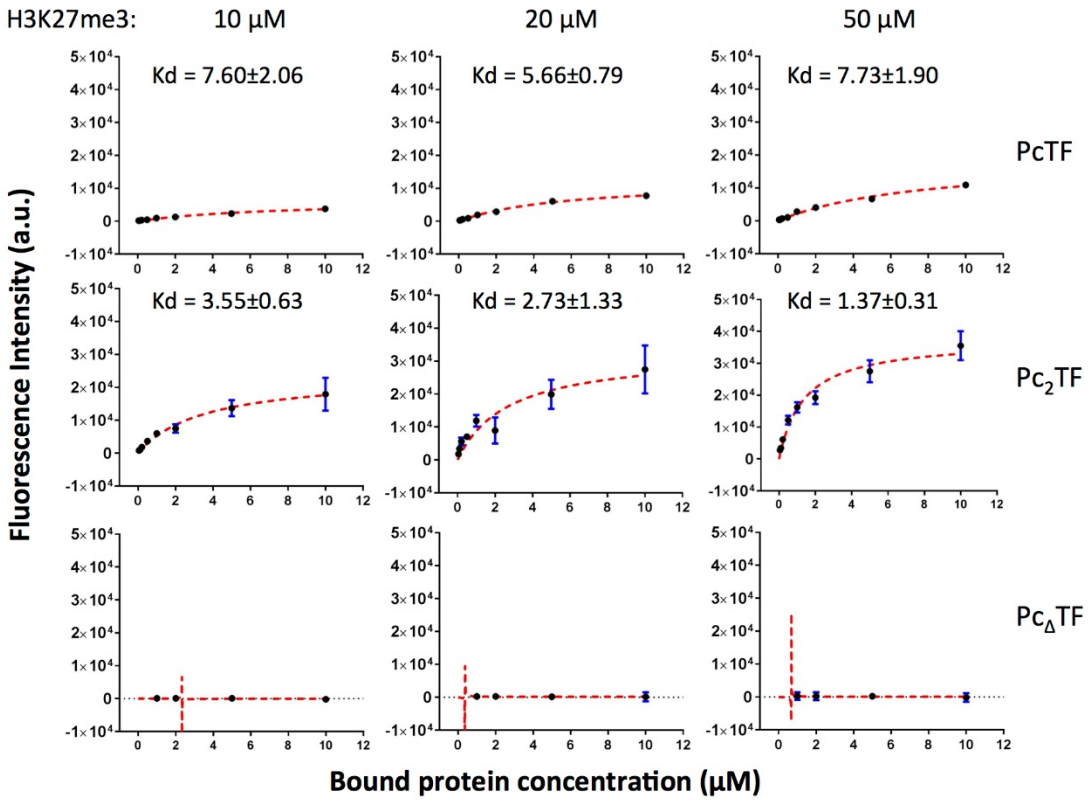


Figure 3.S1 Predicted fluorescence after Pc-fusion binding and subsequent washing. A mathematical model was used to predict mCherry (RFP) signal (fluorophore concentration) when all H3K27me3 targets were saturated by PCD binding (t = 0 sec) and over time during washing with buffer. A mechanistic model of washing of the spot assays was created using differential equations and solved using Matlab. Initial conditions were written to describe the initial amount of bound PcTF (1×10^{-6} M) and bound Pc_2TF (0.5×10^{-6} M) to simulate saturated ligands – each Pc_2TF is bound to two ligands, thus half as many Pc_2TF molecules would saturate the surface. In one simulation, for washing a spot array with PcTF, an equation was written to describe PcTF association and dissociation with a single ligand as: $dPb/dt = ka * Pu * Lu - kd * Pb$. ka is the association rate constant, kd is the dissociation rate constant, Pu is the concentration of unbound PcTF or Pc_2TF, Lu is the concentration of unbound ligand, and Pb is the

complex of PcTF bound to ligand. Algebraic equations also track unbound PcTF (Pu = Ptotal – Pb) and unbound ligand (Lu = Ltotal – Pb). In the other simulation, for washing a spot array with Pc_2TF, two equations were written. The first equation describes Pc_2TF association with a ligand to form a complex bound by one PCD to one ligand (Pb1) which can be lost either by dissociation or association of a second PCD or can be regained by one of the two PCDs of a Pb2 dissociating: dPb1/dt = ka * Pu * Lu – kd * Pb1 – ka * Vr * Pb1 * Lu + kd * Pb2. These parameters are the same as above but also include Vr which is the ratio of overall concentration to effective local concentration (see Shewmake et al. 2008) and Pb2 which is a Pc_2TF bound to two ligands via a PCD for each ligand. The second equation describes the second PCD of Pb1 binding to a ligand and dissociation of one PCD from a ligand: dPb2/dt = ka * Vr * Pb1 * Lu – kd * Pb2. Algebraic equations also track unbound Pc_2TF (Pu = Ptotal – (Pb1 + Pb2)) and unbound ligand (Lu = Ltotal – 1* Pb1 – 2* Pb2). Parameters were set to: ka = 2000 M-1 s-1, kd = 0.01 s-1, Ptotal = initial value of Pb (PcTF) or Pb2 (Pc_2TF), Ltotal = 1 x 10-6 M, Vr = 100. Simulations can be run using most ordinary differential equation (ODE) solvers, but we used ode15s with absolute tolerance of 1e-9 and relative tolerance of 1e-6. Pb vs. time is plotted for PcTF. (Pb1 + Pb2), Pb1, and Pb2 are plotted vs. time for Pc_2TF. Matlab codes and equations used to generate these simulations can be found online: https://github.com/khaynes5/PcTF_kinetics



H3K27me3	Bound Protein	Kd values ±error (R-squared value)			
10 µM	PcTF	1.54±n (0.86)	1.55±n (0.96)	**7.6±2.06 (0.98)**	n/s
	Pc_2TF	0.72±n (0.59)	**3.55±0.63 (0.99)**	3.83±n (0.93)	10.61±n (0.99)
20 µM	PcTF	5.14±n (0.98)	**5.66±0.79 (1.00)**	6.45±n (0.99)	8.95±n (0.92)
	Pc_2TF	0.93±n (0.82)	1.35±n (0.91)	2.05±n (0.98)	**2.73±1.33 (0.89)**
50 µM	PcTF	3.50±n (0.99)	**7.73±1.9 (0.99)**	---	---
	Pc_2TF	0.83±n (0.77)	**1.37±0.31 (0.97)**	---	---

Figure 3.S2 Additional peptide array binding data. Dot plots show representative results from microspot assay trials where varying concentrations of soluble fusion proteins were applied to 10, 20, or 20 µM of H3K27me3 (dots = means of 2 replicate spots, error bars = SDM). Apparent dissociation constant (K_d^{app}) values were determined by lines of best fit (dashed red line). The table shows calculated K_d^{app} values, error (SDM), and R^2 values for all experimental replicates (n/s = no signal, --- = no additional replicate). Bold text indicates K_d^{app} values from curves that are shown in the representative dot plots.

Purification of recombinant protein from *E. coli.* IPTG-induced cells were pelleted by centrifugation at 4,000 RCF for 10 min, resuspended in 30 mL of purification buffer (10% glycerol, 250 mM NaCl, 50 mM Na_2PO_4, pH 8.0), and frozen at -80°C overnight. Disruption of thawed cells (on ice) was performed by sonication with a QSONICA instrument (model Q500) at 50% power: (1 second on, 2 seconds off) x 6000 cycles. After addition of imidazole (10 mM final concentration), insoluble material was pelleted at 16,000 RPM for 30 min at 4°C in a Beckman Coulter Avante J-E Rotor JA-17. Purification columns were prepared by washing 2.5 mL of Ni-NTA Agarose (Qiagen #30210) with 10 mL of ddH_20 on a 50 mL polyprep column, then equilibrated with 15 mL purification buffer plus 10 mM imidazole. Soluble fractions of cell lysates were loaded onto a plugged column, vortexed briefly to homogenize resin and supernatant, and incubated with rotation at 4°C for 2 h. The cap and bottom plug were removed to empty the unbound fraction, and flow-through was applied back to the column once. Protein-bound resin was washed with 10 mL binding buffer plus 10 mM imidazole, followed by 5 mL binding buffer plus 20 mM imidazole and 5 mL binding buffer plus 50 mM imidazole. Resin was incubated with 1 mL of binding buffer plus 250 mM imidazole for 10 min in a plugged column before elution. Elution was repeated with binding buffer plus 500 mM, and then binding buffer plus 1.0 M imidazole. Proteins were concentrated and buffer-exchanged into PBS using a 15 mL 30,000 kDa centrifugal filter and repeated washes of PBS followed by centrifugation at 4000 RCF for 10 min, and stored at 4°C in a final volume of about 1 mL. Concentration of protein was determined using a denaturing polyacrylamide gel with bovine serum albumin (BSA) standards.

Western blot detailed protocol. Roughly 250,000 cells were pelleted at 300 g for 5 min, lysed by resuspension in 500 µL Mammalian Cell PE LB (G-Biosciences #786-180) plus 5 µL 100x ProteaseArrest (G-Biosciences #786-108), and vortexed for 2 min. Insoluble debris was pelleted at 16,000 xg, 4°C, 5 min and discarded. Samples for denaturing polyacrylamide gel electrophoresis (PAGE) were prepared by heating 15 µL lysate plus 4 µL NuPAGE LDS Sample Buffer 4x (Thermo Fisher #NP0007) and 1 µL 1 M Dithiothreitol (DTT, Millipore Sigma #D0632-1G) at 100°C for 5 min. Samples (cooled to room temperature) and a pre-stained protein standard (10 µL, Thermo Fisher #10748010) were electrophoresed at 120 V in a 4-12% Bis-Tris gel (Thermo Fisher #NP0322BOX) with MOPS-SDS buffer [50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7] diluted 1:20 in an XCell SureLock vertical chamber (Invitrogen #EI0001). Proteins were transferred onto a nitrocellulose membrane (Bio-Rad #1704158) via semi-dry transfer in a Transblot Turbo system (Bio-Rad #1704150). Complete transfer was verified by staining the membrane with 1x Ponceau-S (G-Biosciences #786-576).

Abbreviations

BPTF, bromodomain PHD finger transcription factor; ELISA, enzyme-linked immunosorbent assay; H3K27me3, histone H3 trimethylated at lysine 27; PCD, polycomb chromodomain motif; PCR, polymerase chain reaction; PcTF, monovalent polycomb-based transcription factor; Pc_2TF, bivalent polycomb- based transcription factor; PRC, polycomb repressive complex; PTM, post-translational modification; RFP, red fluorescent protein; RT-qPCR, reverse transcription followed by quantitative polymerase chain reaction; TXTL, *E. coli*-based cell-free transcription-translation system

References

1. Tekel, S., and Haynes, K. A. (2017) Molecular Structures Guide the Engineering of Chromatin. Nucleic Acids Res. 45, 7555−7570.

2. Cano-Rodriguez, D., and Rots, M. G. (2016) Epigenetic Editing: On the Verge of Reprogramming Gene Expression at Will. Curr. Genet. Med. Rep. 4, 170−179.

3. Park, M., Keung, A. J., and Khalil, A. S. (2016) The epigenome: the next substrate for engineering. Genome Biol. 17, 183.

4. Platero, J. S., Hartnett, T., and Eissenberg, J. C. (1995) Functional analysis of the chromo domain of HP1. EMBO J. 14, 3977−3986.

5. Messmer, S., Franke, A., and Paro, R. (1992) Analysis of the functional role of the Polycomb chromo domain in Drosophila melanogaster. Genes Dev. 6, 1241−1254.

6. 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., and Knapp, S. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149, 214−231.

7. 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.

8. Gatchalian, J., Fütterer, A., Rothbart, S. B., Tong, Q., Rincon- Arano, H., Sánchez de Diego, A., Groudine, M., Strahl, B. D., Martínez-A, C., van Wely, K. H. M., and Kutateladze, T. G. (2013) Dido3 PHD modulates cell differentiation and division. Cell Rep. 4, 148−158.

9. Wang, R., and You, J. (2015) Mechanistic analysis of the role of bromodomain-containing protein 4 (BRD4) in BRD4-NUT oncoprotein- induced transcriptional activation. J. Biol. Chem. 290, 2744−2758.

10. Haynes, K. A., and Silver, P. A. (2011) Synthetic reversal of epigenetic silencing. J. Biol. Chem. 286, 27176−27182.

11. Nyer, D. B., Daer, R. M., Vargas, D., Hom, C., and Haynes, K. A. (2017) Regulation of cancer epigenomes with a histone-binding synthetic transcription factor. npj Genomic Med. 2, 1.

12. Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., and Allis, C. D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. 26, 2560−2569.

13. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. 17, 1870−1881.

14. Meckler, J. F., Bhakta, M. S., Kim, M.-S., Ovadia, R., Habrian, C. H., Zykovich, A., Yu, A., Lockwood, S. H., Morbitzer, R., Elsäesser, J., Lahaye, T., Segal, D. J., and Baldwin, E. P. (2013) Quantitative analysis of TALE−DNA interactions suggests polarity effects. Nucleic Acids Res. 41, 4118−4128.

15. Jantz, D., and Berg, J. M. (2010) Probing the DNA-binding affinity and specificity of designed zinc finger proteins. Biophys. J. 98, 852−860.

16. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science 275, 657−661.

17. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62−67.

18. 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.

19. Su, Z., and Denu, J. M. (2016) Reading the Combinatorial Histone Language. ACS Chem. Biol. 11, 564−574.

20. Su, X., Zhu, G., Ding, X., Lee, S. Y., Dou, Y., Zhu, B., Wu, W., and Li, H. (2014) Molecular basis underlying histone H3 lysine arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. Genes Dev. 28, 622−636.

21. Jacobson, R. H. (2000) Structure and Function of a Human TAFII250 Double Bromodomain Module. Science 288, 1422−1425

22. Voigt, P., and Reinberg, D. (2011) Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. Chem- BioChem 12, 236−252.

23. Oliver, S. S., Musselman, C. A., Srinivasan, R., Svaren, J. P., Kutateladze, T. G., and Denu, J. M. (2012) Multivalent recognition of histone tails by the PHD fingers of CHD5. Biochemistry 51, 6534−6544.

24. Sanchez, O. F., Mendonca, A., Carneiro, A. D., and Yuan, C. (2017) Engineering Recombinant Protein Sensors for Quantifying Histone Acetylation. ACS Sens 2, 426−435.

25. Delachat, A. M.-F., Guidotti, N., Bachmann, A. L., Meireles- Filho, A. C. A., Pick, H., Lechner, C. C., Deluz, C., Deplancke, B., Suter, D. M., and Fierz, B. (2018) Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. Cell Chem. Biol. 25, 51.

26. Min, J. R., Zhang, Y., and Xu, R.-M. (2003) Structural Basis for specific binding of polycomb chromodomain to histone H3 methylated at K27. Genes Dev. 17, 1823.

27. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251−260.

28. Voigt, P., LeRoy, G., Drury, W. J., 3rd, Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A., and Reinberg, D. (2012) Asymmetrically modified nucleosomes. Cell 151, 181−193.

29. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 16, 2893−2905.

30. Wriggers, W., Chakravarty, S., and Jennings, P. A. (2005) Control of protein functional dynamics by peptide linkers. Biopolymers 80, 736−746.

31. Merutka, G., Shalongo, W., and Stellwagen, E. (1991) A model peptide with enhanced helicity. Biochemistry 30, 4245−4248.

32. Yan, W., Imanishi, M., Futaki, S., and Sugiura, Y. (2007) α-Helical Linker of an Artificial 6-Zinc Finger Peptide Contributes to Selective DNA Binding to a Discontinuous Recognition Sequence. Biochemistry 46, 8517−8524.

33. Shewmake, T. A., Solis, F. J., Gillies, R. J., and Caplan, M. R. (2008) Effects of Linker Length and Flexibility on Multivalent Targeting. Biomacromolecules 9, 3057−3064.

34. Chen, X., Zaro, J. L., and Shen, W.-C. (2013) Fusion protein linkers: property, design and functionality. Adv. Drug Delivery Rev. 65, 1357−1369.

35. Caschera, F., and Noireaux, V. (2014) Synthesis of 2.3 mg/mL of protein with an all Escherichia coli cell-free transcription−translation system. Biochimie 99, 162−168.

36. Shen, Y., Chen, Y., Wu, J., Shaner, N. C., and Campbell, R. E. (2017) Engineering of mCherry variants with long Stokes shift, redshifted fluorescence, and low cytotoxicity. PLoS One 12, e0171257.

37. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G. A., Li, Z., Vedadi, M., Schapira, M., Min, J., and Arrowsmith, C. H. (2011) Recognition and specificity determinants of the human cbx chromodomains. J. Biol. Chem. 286, 521−529.

38. Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967−980.

39. Swigut, T., and Wysocka, J. (2007) H3K27 demethylases, at long last. Cell 131, 29−32.

40. Daer, R. M., Cutts, J. P., Brafman, D. A., and Haynes, K. A. (2017) The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. ACS Synth. Biol. 6, 428−438.

41. Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008) A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. 10, 1291−1300.

42. Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S., and Jacobsen, S. E. (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J. 23, 4146−4155.

43. Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I. A., Zenn, H. M., Kost, N., Pohl, W., Chernev, A., Schwarzer, D., Jenuwein, T., Lorincz, M., Zimmermann, B., Walla, P. J., Neumann, H., Baubec, T., Urlaub, H., and Fischle, W. (2016) Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. Nat. Commun. 7, 11310.

44. Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N.C., Joung, J. K., and Collins, J. J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. Cell 150, 647−658.

45. Rosca, E. V., Gillies, R. J., and Caplan, M. R. (2009) Glioblastoma targeting via integrins is concentration dependent. Biotechnol. Bioeng. 104, 408−417.

46. Senthilkumar, R., and Mishra, R. K. (2009) Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. BMC Genomics 10, 549.

47. Gibaja, V., Shen, F., Harari, J., Korn, J., Ruddy, D., Saenz-Vash, V., Zhai, H., Rejtar, T., Paris, C. G., Yu, Z., Lira, M., King, D., Qi, W., Keen, N., Hassan, A. Q., and Chan, H. M. (2016) Development of secondary mutations in wild-type and mutant EZH2 alleles cooperates to confer resistance to EZH2 inhibitors. Oncogene 35, 558−566.

48. Baker, T., Nerle, S., Pritchard, J., Zhao, B., Rivera, V. M., Garner, A., and Gonzalvez, F. (2015) Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. Oncotarget 6, 32646−32655.

49. Phillips, I., and Silver, P. (2006) A New Biobrick Assembly Strategy Designed for Facile Protein Engineering. http://dspace.mit. edu/handle/1721.1/32535.

50. Quadri, S. M. S. (2015) Parafilm-M®, An Available Cost- Effective Alternative for Immuno-blot Pouches. Methods Mol. Biol. 1314, 313−323.

51. Ren, C., Morohashi, K., Plotnikov, A. N., Jakoncic, J., Smith, S. G., Li, J., Zeng, L., Rodriguez, Y., Stojanoff, V., Walsh, M., and Zhou, M.-M. (2015) Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. Chem. Biol. 22, 161−168.

52. Wakamori, M., Fujii, Y., Suka, N., Shirouzu, M., Sakamoto, K., Umehara, T., and Yokoyama, S. (2015) Intra- and inter-nucleosomal interactions of the histone H4 tail revealed with a human nucleosome core particle with genetically-incorporated H4 tetra-acetylation. Sci. Rep. 5, 17204.

CHAPTER 4

CELL-PENETRATING SYNTHETIC CHROMATIN PROTEINS REGULATE GENES IN HUMAN U-2 OS CELLS

Abstract

FDA-approved epigenetic drugs are low molecular weight compounds that indirectly correct gene expression in cancer cells by inhibiting enzymes that modify chromatin. The ease of screening libraries of compounds, high intrinsic membrane permeability, and low rates of degradation have made small molecule inhibitors effective in treating some epigenetic diseases. However, small molecule inhibitors often target cytosolic proteins, and do not directly target nucleosomes. Protein fusion-based epigenetic drugs could help mitigate this problem by allowing customizable treatments that target one specific histone mark with a tuned gene response, allowing for greater efficacy compared to small molecules. Unfortunately, transgene delivery can be generally cytotoxic and lead to DNA lesions with unpredictable effects. To address this issue, we fused the histone binding domain from CBX8, an mCherry fluorescent reporter, and the transcriptional activator VP64 (PcTF) to generate an epigenetic transcription factor capable of activating large sets genes independently of DNA sequence. Herein we report the delivery of recombinantly expressed and purified polycomb transcription factor regulator (PcTF) fused in frame with a nuclear localization signal (NLS) and a cell penetrating peptide tag (TAT). We optimized the delivery of purified, cell penetrating PcTF-NLS-TAT (CP-PcTF) in vitro with both 2D and 3D spheroids of U-2 OS bone cancer cells. We then performed RNAseq on cells treated with CP-PcTF to identify a panel of up and downregulated genes. In summary, we have demonstrated that an

epigenetic effector fused to a CPP can be delivered to U-2 OS cells and can influence gene expression through DNA independent interactions.

Introduction

Protein based therapeutics have become more popular since the introduction of recombinant insulin decades ago (1). However, a significant hurdle for precision, protein-based medicine lies in the efficient delivery of treatments (2). While many small molecule drugs are able to easily pass through a membrane, many protein therapeutics are unable to be efficiently delivered (3). Protein therapeutics offer several advantages over small molecule drugs, such as high specificity, potency, and lower likelihood to induce immune responses (1, 4). In addition, delivering recombinantly purified proteins offer additional benefits over nucleic acid based deliveries such as transfection and viral transduction. First, producing protein recombinantly enables precise control of protein purity, post translational modifications, and quantity of delivered protein (1). Furthermore, recombinant protein delivery avoids nucleic acid induced cytotoxicity and off target effects from viral or DNA particles. Third, nucleic acid based delivery can be cytotoxic and induce recombination with the endogenous genome. Recombinant protein delivery avoids these caveats and has been shown to enhance target binding specificity (5). Recently, researchers have used a variety of techniques to deliver recombinant Cas9 fusions (6-9). Successful delivery and gene editing in plants and animals has demonstrated the versatility and utility of recombinant protein delivery for gene therapy. Delivery of novel protein therapeutics could provide an alternative treatment for defective proteins over gene therapy (10). There is an expansive collection of techniques to deliver proteins to cells, ranging from codelivery with small molecules, peptides,

polymers, or lipids (reviewed in 1, 11). One technique uses cell penetrating peptides (CPPs) to induce endocytosis and deliver cargo through a co-incubation of cargo and CPP, or though covalent conjugation of cargo and CPP (12). Covalent fusions of CPPs and proteins have been successfully used to deliver enzymes to cells (13), expand stem cells (14), and inhibit cancer cell growth (15). For example, protein fusions with an HIV protein derived cell penetrating peptide TAT have been used to kill cancer cells and induce gene expression in embryonic stem cells (16-17). Understanding design rules to efficiently deliver protein-based drugs could lead to new insight in disease, and help lead to more precise treatments.

Designer epigenome effectors can be used to silence genes with high specificity (18). However, the majority of epigenetic based drugs are small molecule based and can end up in off target locations (19). With the emergence of computational design and new high throughput assays, therapeutic protein engineering is becoming more attainable (20). Engineering chromatin is a growing field (21), with recent work producing assays for determining *in vitro* function of synthetic histone binding domains (22). However, less work has been done to investigate the in cell function of these proteins in 2D and 3D cell models.

We have previously demonstrated the *in vitro* function of the recombinantly purified epigenetic effector PcTF, a fusion of an H3K27me3 binding domain CBX8, an mCherry domain fluorescent reporter, and a VP64 transcriptional activator (22). In addition, the function of a transiently transfected PcTF and a stably integrated PcTF has been characterized in 2D models with several cancer cell lines (23). These results provide great insight into the function of PcTF, but do not accurately recapitulate the impact of

PcTF on a 3D tumor environment or demonstrate function with a therapeutic protein delivery method. In order to elucidate the function of a recombinant PcTF delivered to a 2D and 3D environment, PcTF would have to be delivered without nucleic acid delivery.

We explored a variety of CPPs with varying properties to assess which peptide would deliver PcTF with minimal toxicity. First, we co-incubated a variety of cell penetrating peptides TAT (1), HA2 (24), E5 (25), and L17E (26) with purified PcTF in 2D U-2 OS cells. We detected mCherry signal inside of U-2 OS cells co-incubated with TAT and the L17E peptide after 24 hours (Supplemental Figure 4.S1). In order to maximize protein delivery and reduce possible cytotoxicity of exogenous peptides, we fused our cell penetrating peptides in-frame with PcTF. We used this PcTF-NLS-TAT fusion (CP-PcTF) to investigate the delivery and function of a recombinant epigenetic effector on 2D and 3D U-2 OS osteosarcoma cells.

Here, we demonstrate that CP-PcTF is delivered to cells in a time-dependent manner, and that CP-PcTF influence gene expression in 2D U-2 OS. In addition, we show that CP-PcTF slows growth in 3D spheroids of U-2 OS cells. Our results have important implications for delivering recombinant epigenetic effectors to cells in 2D and 3D models.

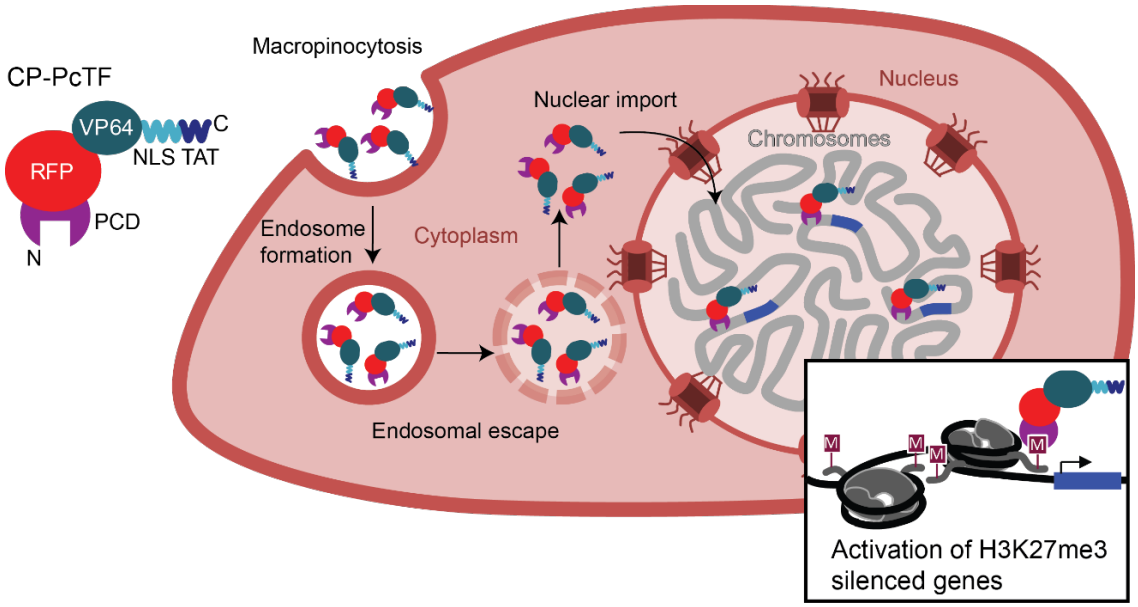


Figure 4.1. Delivery and cellular transport of the fusion chromatin protein PcTF. *E. Coli* expressed and purified CP-PcTF is incubated with U-2 OS cells. CP-PcTF enters the cell through macropinocytosis, followed by endosomal escape into the cytoplasm. CP-PcTF is then imported into the nucleus where it activates H3K27me3-silenced genes.

Design of a Cell Penetrating PcTF

In order to determine the optimal delivery vehicle for PcTF, we assayed the ability of several CPPs to deliver PcTF to U-2 OS cells (Supplemental Figure 4.S1). We synthesized four distinct peptides with varying properties to maximize our likelihood of finding a robust delivery partner. First, we synthesised a TAT peptide, a peptide fragment from the human immunodeficiency virus 1 (HIV-1) that has robust cell penetrating ability (1). However, cargo being delivered by pinocytosis has been known to become stuck in endosomes, leading to loss of function (27). Therefore, we synthesized TAT fusion peptides such as TAT-HA2, which consists of the N terminal 20 amino acids of the influenza virus hemagglutinin protein fused to a TAT peptide. HA2 is hypothesized to help with endosomal escape, enabling delivery of functional protein (24). E5-TAT, an

enhanced HA2 with increased membrane disruption ability fused with TAT, was also synthesized (25). Finally, we synthesized an engineered membrane-lytic peptide L17E, which was previously shown to enable robust cellular delivery of protein cargo (26).

To quickly and efficiently identify peptides that would allow PcTF delivery, we incubated each peptide with a high concentration of recombinant PcTF in 2D U-2 OS cells. We reasoned that the mCherry domain in PcTF would be a good fluorescent reporter for delivery. Each peptide was incubated with PcTF for 24 hours before washing and imaging with fluorescent microscopy. After 24 hours of incubation, the cells treated with PcTF and TAT or L17E peptides showed detectable mCherry signal inside the cells, while the TAT-HA2 and E5-TAT treated cells had no detectable signal (Supplemental Figure 4.S1). While the L17E cells showed more overall mCherry signal, we observed a localization of signal in the center of each cell with the TAT peptide (Supplemental Figure 4.S1). In an attempt to improve cellular delivery and signal for TAT and L17E, we constructed PcTF-CPP fusions for recombinant expression and purification. However, only the TAT fused PcTF was soluble in *E coli*. We call this PcTF-NLS-TAT fusion CP-PcTF (Figure 4.1). CP-PcTF is hypothesized to enter the cell through macropinocytosis, where the positively charged TAT residues interact with the negatively charged cell membrane, inducing cell entry (Figure 4.1). Once in the cell, CP-PcTF escapes the endosome and is imported into the nucleus where it binds to H3K27me3 and upregulates proximal genes (Figure 4.1).

The TAT peptide signal supports delivery of purified PcTF into U-2 OS cells in 2D and 3D cultures

To obtain purified, CP-PcTF in sufficient quantities for cell delivery, we cloned the open reading frame into an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible pET28 vector. We overexpressed CP-PcTF in *E coli*, with visible mCherry signal accumulating after three hours of induction. Cells were harvested, pelleted, and lysed before purifying CP-PcTF on a Nickel-NTA column. Purified proteins were concentrated into phosphate buffered saline (PBS) and quantified using BSA dilutions on a denaturing polyacrylamide gel to obtain delivery ready protein.

To determine the optimal dose and incubation time for delivery of functional CP-PcTF, we treated 2-D U-2 OS cells with soluble, purified CP-PcTF fusion proteins and observed accumulation of the RFP tag via microscopy and flow cytometry (Figure 4.2). Previous reports with TAT-protein fusions as a delivery vehicle used concentrations of protein ranging from 2 nM- 5 μM protein (14, 16-17, 28). We treated the cells with CP-PcTF ranging from 100 nM to 1 uM for 24 hours. We saw an increase of mCherry positive cells (RFP+) cells as we increased the concentration of CP-PcTF (Figure 4.2A). At 100 nM CP-PcTF, we detected RFP in 85.6% of the live cells, followed by 97.8% for 200 nM, 99.1% for 300 nM, 99.7% for 500 nM, and 100% for 1 μM.

To optimize the incubation time of CP-PcTF, we performed a time course by varying U-2 OS incubation times with 200 nM CP-PcTF (Figure 4.2B). We observed the beginning of cell penetration after an hour of incubation with RFP signal surrounding the cell post wash. After 24 hours of incubation, we observed punctate RFP signal appearing in the center of the cell. This signal was still detectable 48 hours after washing out unpenetrated protein, indicating that CP-PcTF is able to endogenous survive protease activity for at least 72 hours (Figure 4.2B). To ensure localization of RFP signal

correlated to our expected nuclear target of H3K27me3, we treated U-2 OS cells with CP-PcTF (Figure 4.2C). We treated U-2 OS cells with CP-PcTF and PcTF for 24 hours, followed by 48 hours of growth. We stained the cells for nuclear signal (Hoescht) and imaged the cells using fluorescent microscopy (Figure 4.2C). We observed colocalization of RFP signal and Hoescht (nuclear stain) signal, indicating that CP-PcTF is being imported into the nucleus.

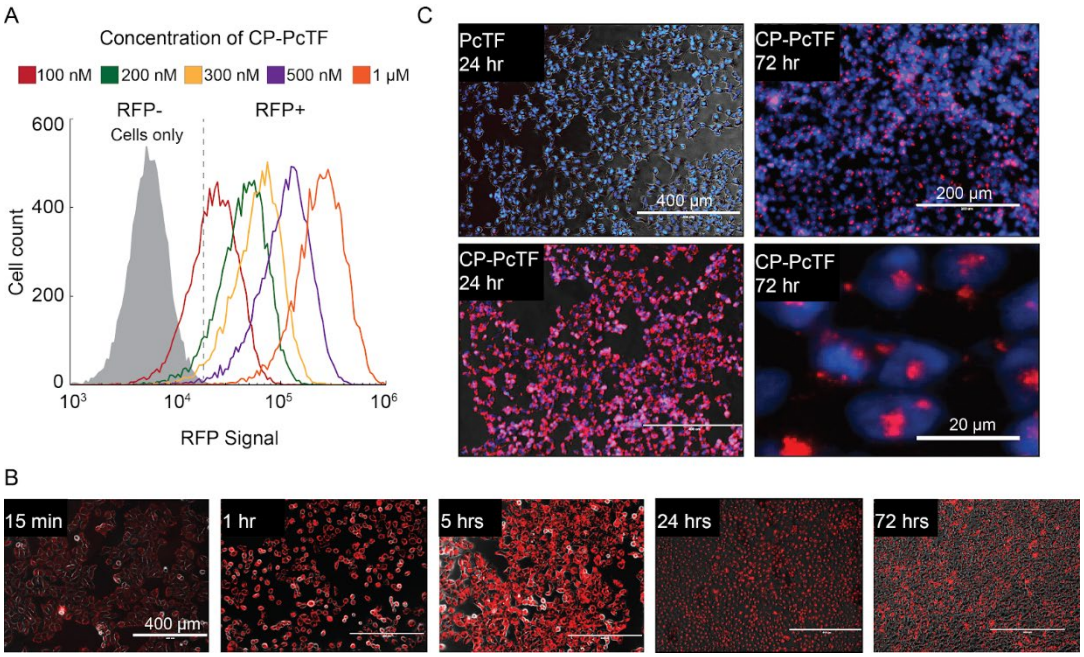


Figure 4.2. Uptake of soluble CP-PcTF by 2D U-2 OS cells (A) Flow cytometry of 2D U-2 OS cells treated with varying concentrations of CP-PcTF for 24 hours. (B) CP-PcTF shows time dependent delivery into U-2 OS cells. Cells were treated with 200 nM of CP-PcTF for 15 minutes, 1 hour, 5 hours, or 24 hours before washing 3 times with PBS and imaging with fluorescent microscopy. The 72-hour time point is from cells treated with CP-PcTF for 24 hours. (C) C-terminal TAT peptide supports dose-dependent accumulation of PcTF in 2-D cultured U-2 OS cells.

Genes become upregulated by CP-PcTF

To determine if CP-PcTF induced a transcriptional response upon delivery, we treated 2D U-2 OS cells with CP-PcTF for 24 hours and harvested the genomic RNA for next generation RNA sequencing (RNAseq). We also treated U-2 OS cells with CP-$Pc_{\Delta}TF$ as a control.

RNAseq revealed a set of genes significantly ($p<0.05$) upregulated compared to untreated cells. Compared to an untransfected control, we observed 103 and 82 significantly upregulated (2 fold) genes in the CP-PcTF and the CP-$Pc_{\Delta}TF$ respectively. Of these genes, 69 of them were unique to CP-PcTF, while 48 were unique to CP-$Pc_{\Delta}TF$. Many of these genes code for hypothetical proteins, are untranslated, or have unknown function. Regulation of these transcripts could be explained by H3K27me3-mediated silencing at these loci, but further investigation such as ChIP and qPCR is required to validate these targets.

We removed genes that code for uncharacterized predicted proteins, or have no known function, resulting in 58 annotated genes for CP-PcTF and 38 for CP-$Pc_{\Delta}TF$. We compared the filtered lists to each other, resulting in 16 shared genes and 42 uniquely upregulated genes for CP-PcTF (Figure 4.3A). In order to compare the upregulated genes to previous work with PcTF, we mapped the fold change from the significantly upregulated CP-PcTF genes to the fold change of the same genes from the previous transgenic PcTF (TG-PcTF) work (23) (Figure 4.3B). Interestingly, only 12 of the genes uniquely upregulated by CP-PcTF were also upregulated by TG-PcTF at the same time point. These genes mainly fall into the chemokine category of genes and suggest that PcTF might induce transcription of immune response genes in U-2 OS cells. The large

difference in significantly upregulated genes between CP-PcTF and TG-PcTF could be do to differing time scales for PcTF effectiveness, or different modes of action due to delivery. Of the 16 genes commonly upregulated by both CP-PcTF and CP-Pc_{Δ}TF, only two were shared with TG-PcTF, suggesting that these genes might be upregulated in response to protein delivery and not PcTF itself (Figure 4.3B).

In order to investigate common gene functions between the genes upregulated by CP-PcTF, we performed an overrepresentation test using Protein ANalysis THrough Evolutionary Relationships (PANTHER) (29). This analysis can help reveal evolutionarily related proteins and families based on omics data. The output of the analysis produces three classes: molecular function, biological process, and pathway. Our data revealed significant enrichment (p <0.05, FDR < 0.05) of gene families for each analysis. There were four clusters for biological processes, three for molecular functions, and one for cellular components (Figure 4.3C). The number of upregulated genes involved in the immune response suggests that these genes might be epigenetically silenced by H3K27me3.

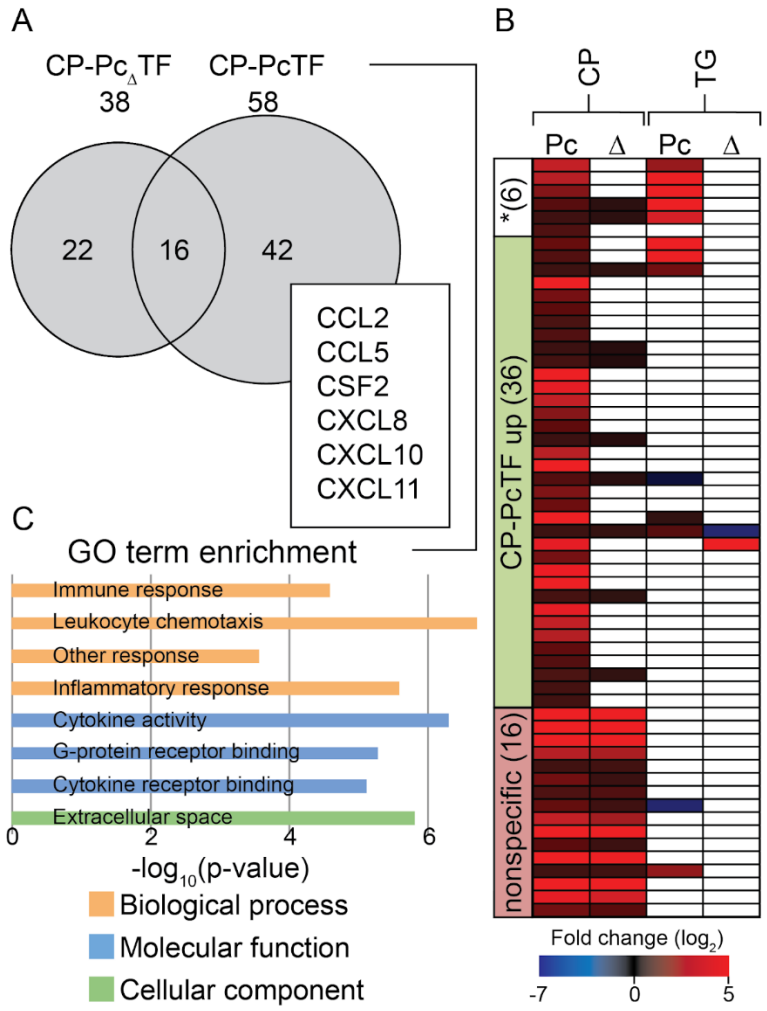


Figure 4.3. Regulation of genes by CP-PcTF. (A) The Venn diagram shows shared and distinct genes with significant upregulation (FC ≥ 2, $p < 0.05$) for cells treated with CP-PcTF or the truncation control CP-Pc$_\Delta$TF. Out of 103 CP-PcTF upregulated genes, a set of genes (box) showed statistically significant gene ontology term enrichment (bar chart). (B) Heatmap of regulated genes. * = Cytokines, white boxes indicate a p value >0.05 or no signal.

3D cells treated with CP-PcTF show delayed growth

In order to investigate the ability of CP-PcTF to penetrate cells in a 3D tumor environment, we delivered protein to U-2 OS spheroids. We generated spheroids by

culturing cells in low attachment plates with constant rotation for several days. We treated the spheroids with CP-PcTF for 24 hours and used fluorescent microscopy and flow cytometry to visualize and quantify cell penetration (Figure 4.4A). We observed punctate RFP signal in the 3D spheroids (Figure 4.4A) similar to the 2D delivery (Figure 4.4B). However, we were unable to image the center of the spheroid where CP-PcTF might be excluded. Therefore, we dissociated the tumoroids into a single cell suspension and determined the fraction of RFP positive (RFP+) cells using a flow cytometer (Figure 4.4B). Compared to the 2D delivery assay, the 3D cells contained fewer RFP+ cells at the same concentrations. 100 nM CP-PcTF was only 8.2% RFP+, followed by 200 nM with 20.6%, 300 nM with 21.2%, 500 nM with 30.1%, and 1 μM at 45.4% RFP+ (Figure 4.4B). This lower efficiency suggests CP-PcTF is unable to penetrate all cells in spheroids.

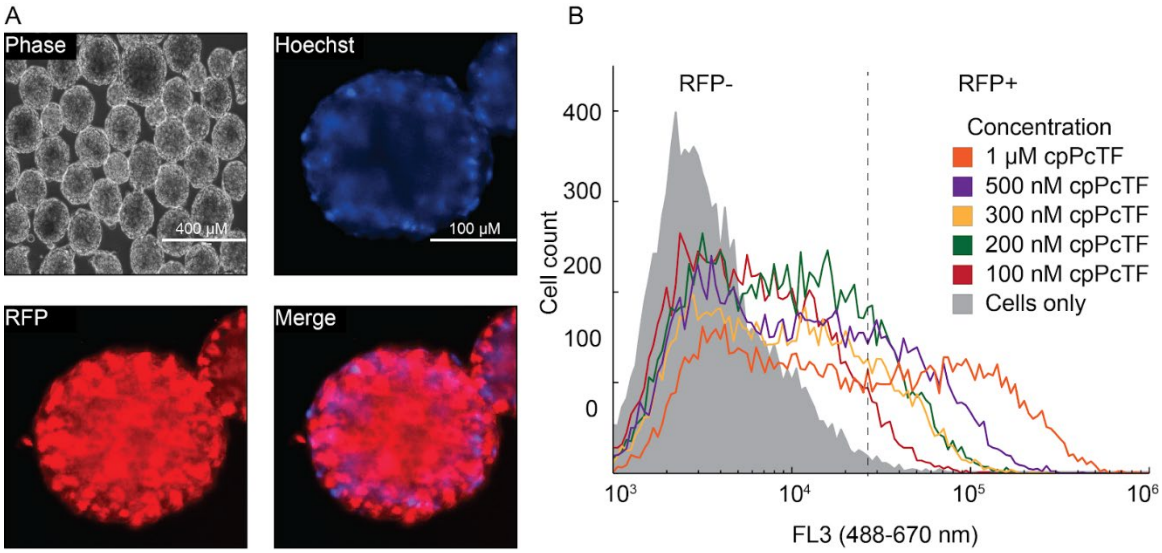


Figure 4.4. Incomplete uptake of CP-PcTF in U-2 OS spheroids. (A) Treatment with 200 nM CP-PcTF shows cell penetration into spheroids (B) Flow cytometry analysis indicates an incomplete delivery of CP-PcTF to cells in spheroid.

To determine how CP-PcTF might affect the growth of tumor tissue, we treated U-2 OS spheroids with CP-PcTF and performed a time course monitoring spheroid growth. Since an immune response can be caused by cellular senescence, and CP-PcTF showed activation of immune response genes (30), we hypothesized that CP-PcTF might be able to influence the growth of the cells by inducing a senescence like state. This was previously demonstrated in a 2D environment with transgenic U-2 OS epressing PcTF (31). Therefore, we treated the tumoroids with protein for 24 hours followed by CP-PcTF washout and incubation in normal growth media. Each day, phase microscopy images were taken for spheroid size analysis. We measured cell size in Adobe Illustrator by measuring the diameter of spheroids (Figure 4.5A).

We observed a significant difference in tumoroid size (p <0.001) by the fourth day between the CP-PcTF treated and untreated cells (Figure 4.5A). As a control, we also treated spheroids with CP-Pc$_{\Delta}$TF, but saw no significant reduction (p =0.33) in size between untreated and cp$_{\Delta}$PcTF treated cells after 4 days (Figure 4.5A). As expected, we observed a loss of mCherry signal over time in the cells treated with protein as seen by RFP negative cells on the outside of spheroids (Figure 4.5B). These results suggest that CP-PcTF is able to induce senescence in a U2-OS cancer line. This is consistent with previous work that demonstrated PcTF induces a senescence like state in 2D U-2 OS models (31).

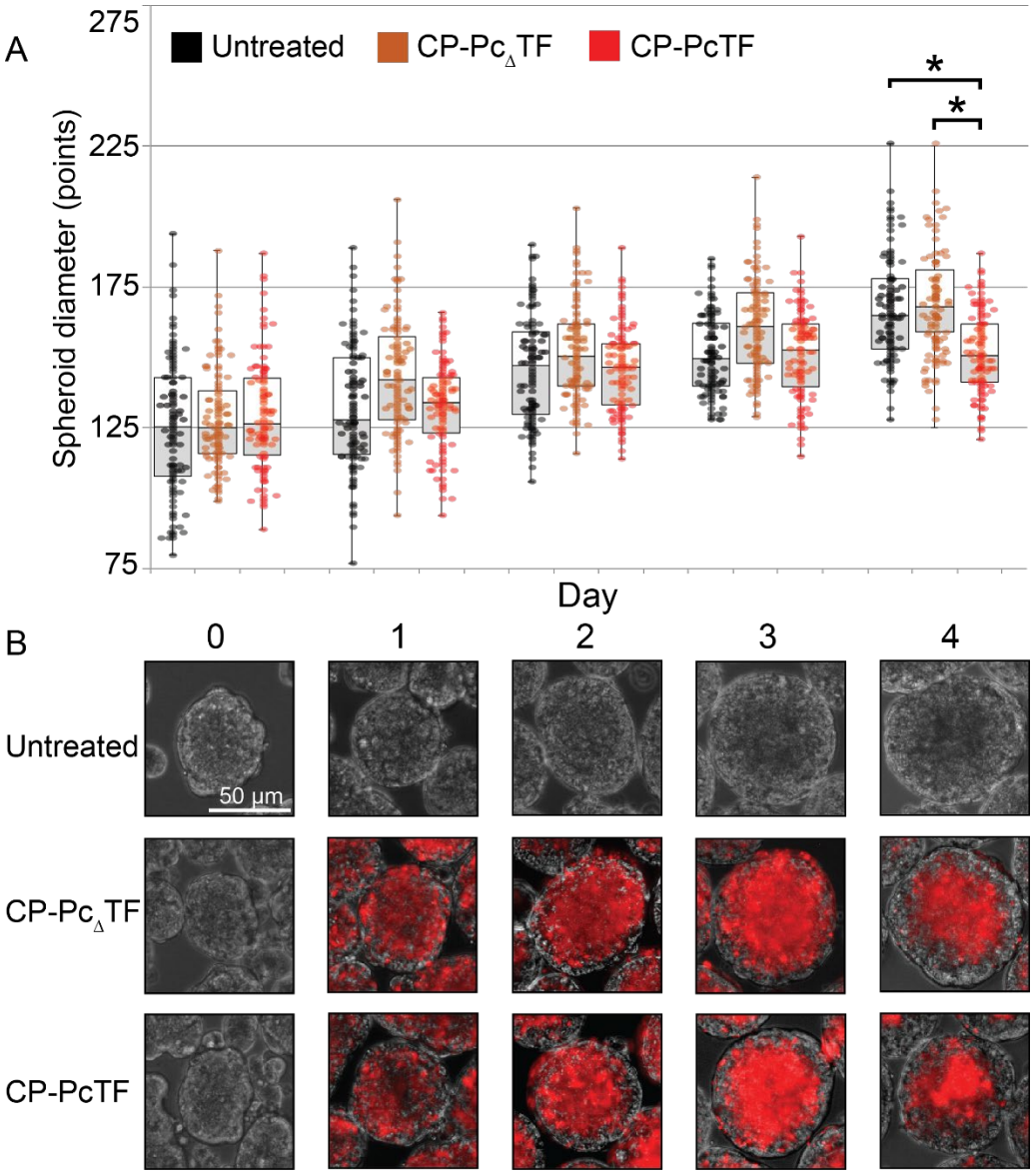


Figure 4.5. Comparisons of spheroid growth over time in CP-PcTF-treated samples versus controls. (A) Spheroid diameter (points) of 3D spheroid U-2 OS over time. Sizes measured in illustrator. (B) Representative images of U-2 OS tumoroid during each day of treatment. * denotes a *p* value <0.001. Significance was determined using a student's two tailed t test.

Discussion

This work presents the first characterization of a synthetic, cell penetrating epigenetic effector. Our previous work demonstrated the *in vitro* function of an H3K27me3 binding activator fusion PcTF using inducible systems and transient expression. In this study, we fuse a cell penetrating peptide tag to PcTF to generate CP-PcTF. We express and purify CP-PcTF, and characterize its effect on 2D and 3D cells.

Here, we demonstrated that CP-PcTF is delivered to 2D U-2 OS cells in a time and concentration dependent manner. At just 200 nM CP-PcTF, we detected mCherry in over 95% of the cells after 24 hours of incubation, indicating that relatively low concentrations of TAT-fused epigenetic regulators can be robustly delivered to U-2 OS cells. However, 24 hours of incubation is not a likely possibility for protein based therapies, so future work optimizing minimal delivery time and possible microinjections into tumoroids should be explored. For example, fusing a receptor ligand to a protein has enabled cell specific delivery (32).

Whole transcriptome analysis of CP-PcTF treated U-2 OS cells revealed a set of genes specifically upregulated compared to a nonbinding control (CP-Pc$_{\Delta}$TF) treated line. Very few of these genes matched the results of a transgenic PcTF expressing line (23), but the differences could be attributed to timing or differing modes of delivery. qPCR should be used to validate this set of genes.

Additional protein incubation time points followed by RNAseq will help elucidate the function of CP-PcTF. This work used a 24 hour time point based on previous work (23), however, there might be a different result based on transient transfection versus transient protein delivery in terms of gene response.

In the context of a 3D tumor environment, CP-PcTF is inefficiently delivered. When treated with 1 uM CP-PcTF, RFP signal was detected in less than 50% of cells. This is consistent with previous reports that show better delivery efficiency in 2D monolayers compared to 3D cultures (33).

Future work will confirm RNAseq and cell growth assay results. Quantitative PCR (qPCR) assaying genes identified from the RNAseq results will validate CP-PcTF-dependent responses. WST1 and B-galactosidase assays will confirm CP-PcTF induced cellular senescence. To further improve function, engineering of the PcTF binding domain can lead to a more robust gene response. For example, a multivalent CP-Pc_2TF might have enhanced function compared to CP-PcTF, as seen previously in transiently transfected cells (34).

Protein therapeutics have become increasingly popular. The ability to robustly deliver proteins to cell targets will advance precision medicine and enable more advanced disease research. Understanding design rules for recombinant delivery of therapeutic synthetic gene regulators can help engineers design optimal fusions that overcome the problems associated with small molecule epigenetic inhibitors. In conclusion, this work demonstrates that the synthetic epigenetic effector PcTF can be delivered to 2D and 3D cells, and provides a possible solution for delivery of engineered chromatin proteins.

Methods

Peptide synthesis:

The peptide sequences were synthesized on a resin using standard Fmoc chemistries. Analysis of the peptides by analytical HPLC and MALDI-TOF confirmed that the peptides had the correct expected masses.

Plasmid constructs:

Plasmids were assembled by consecutive PCRs using pET28(Pc$_{\Delta}$TF) and pET28(PcTF) ((https:// benchling.com/hayneslab/f_/rmSYkAAU-synthetic-chromatinactuators-
2-0). A 3' NLS was added to PcTF PCR using FP 5' ATACAGGATCCATGGAGCTTTCAGCGG and RP 5' ACAATAAAGCTTTCATACCTTGCGCTTTTTCTTGGG. A 3' TAT was assembled onto PcTF-NLS by PCR using FP: 5' ATCAGGATCCATGGAGCTTTCAGCGG and RP 5' CGTTTTTTACGACCATATACCTTGCGCTTTTTC, followed by another PCR with the same forward primer and the RP 5' ATGCTAAGCTTCTAACGACGACGCTGACGACGTTTTTTACGACCA. PCR products were digested with BamHI and HindIII, followed by ligation into a pET28 previously digested with BamHI and HindIII, followed by gel purification. cpPc$_{\Delta}$TF was assembled by PCR amplifying CP-PcTF with 5'GGTCGCGGATCCGTGAGCAAGGGCGAGGAG and 3' ATGCTAAGCTTCTAACGACGACGCTGACGACGTTTTTTACGACCA and cloning into pET28 as described above.

Expression and purification of recombinant proteins:

Proteins were expressed, purified, and quantified as described in Tekel et al 2018 (34).

Cell culture:

U-2 OS cells (ATCC HTB-96) were maintained in McCoy's 5A (Modified) Medium supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin-

streptomycin (all from ThermoFisher). Cells were maintained in a 37°C incubator with 5% CO_2 and passaged once ~80% confluent. 3-D spheroids were generated by seeding 2.0 E6 cells per well of a 6-well cell repellent plate (VWR) in a final volume of 4 mL complete media and placed on an orbital shaker at 95 RPM for 24 hours. Half media change was performed every 24 hours. Spheroids were cultured for a minimum of 48 hours before treatment. BJ fibroblasts (ATCC CRL-2522) were maintained in Dulbecco's modified eagle medium supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin-streptomycin (ThermoFisher). Cells were maintained in a 37°C incubator with 5% CO2 and passaged once ~80% confluent.

Microscopy and flow cytometry:

Fluorescent images of each well at 10x and 20x were obtained using an EVOS Cell Imaging System using the RFP channel (ex/em: 470/510 nm) at 500 ms exposure time. Flow cytometry was performed to determine protein uptake. Briefly, cells were dissociated using Accutase (ThermoFisher), resuspended in PBS, and analyzed using a BD Accuri C6 cytometer (BD Biosciences). RFP-positive cells were measured using the FL-3 (ex/em: 488/670 nm) channel and gated based on negative control. Data was analyzed using Flowjo.

3D cell growth analysis:

5 unique 10x phase contrast images were taken for each condition on each day using an EVOS. Cell diameter was measured in Adobe Illustrator for 100 cells/image/day.

RNA-seq:

RNAseq was performed by the Bioinformatics core at ASU. Using KAPA's Ribo-Zero RNA HyperPrep Kit with RiboErase (HMR) (Roche #KK8560), total RNA was ribo depleted, sheared to roughly 250 bp, and then converted to cDNA. Illumina-compatible adapters with unique indexes (IDT #00989130v2) was then ligated onto each sample individually. The adapter ligated molecules were then cleaned using AMPure beads (Agencourt Bioscience/Beckman Coulter, A63883), and amplified with Kapa's HIFI enzyme (KK2502). Each library was then analyzed for fragment size on an Agilent Tapestation, and quantified by qPCR (KAPA KK4835) on Thermo Fisher Scientific's Quantstudio 5 before multiplex pooling and sequencing on a 1x75 flow cell on the NextSeq500 platform (Illumina) at the ASU Genomics Core facility.

RNAseq data analysis:

Genes fitting both $p < 0.05$ and log2 fold change >2 from CP-PcTF vs untreated and CP-Pc$_{\Delta}$TF vs untreated were identified by cross referencing gene symbols with the NCBI database. Genes with no known function were removed. Venn diagrams were generated between CP-PcTF and CP-Pc$_{\Delta}$TF with Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/) (35). GOterm analysis was performed by analyzing the genes regulated specifically by CP-PcTF using the following parameters: Analysis Type: PANTHER Statistical Overrepresentation Test (Released 20181113) Annotation Version and Release Date: PANTHER version 14.0 Released 2018-12-03. Analyzed List: Client Text Box Input (Homo sapiens) Reference List: Homo sapiens (all genes in database).

Funding:

RNA sequencing and data analysis was supported by the Illumina and Arizona State University's Genomics Core Innovative Investigators in Arizona Grant. SJT is supported by the Ira A. Fulton School of Engineering at Arizona State University. NB is supported by a fellowship from the International Foundation for Ethical Research (IFER).

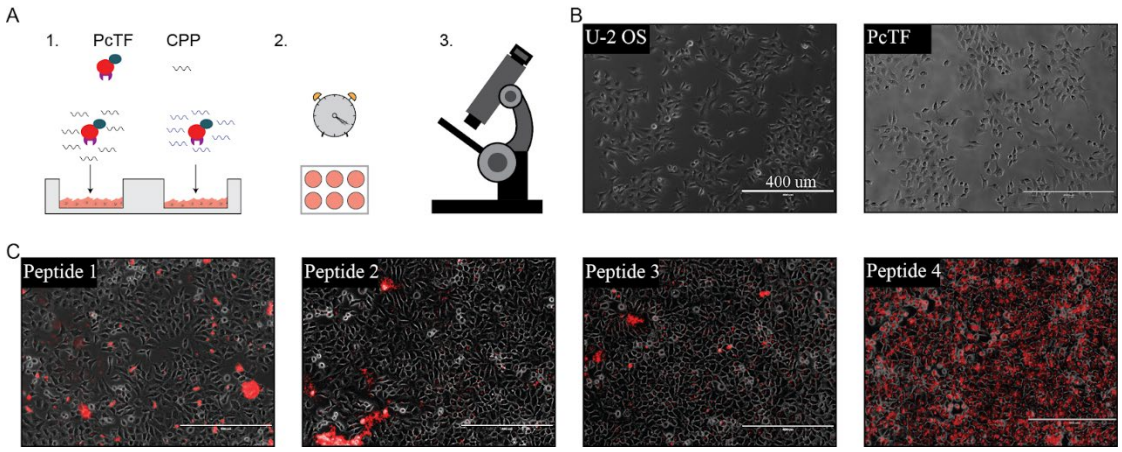
Author contributions:

SJT and KAH conceptualized project. SJT cloned, expressed, purified, and quantified protein constructs. NB performed cell culture. SJT and NB performed microscopy and flow cytometry. KAH oversaw experiments. SJT and KAH generated figures, analyzed data, and wrote the manuscript.

Acknowledgements:

We thank Jason Steele and the Arizona State University Bioinformatics Core for their help with RNA quality analysis, RNA sequencing, and RNAseq data analysis. Peptides were a generous gift from Tara MacCulloch in the Nick Stephanopolous lab. We also thank David Brafman for reagent and supply funding.

Supplemental:



Supplemental Figure 4.S1. Comparisons of cell penetrating peptides co-incubated with PcTF. (A) 1. 2D U-2 OS cells are co-incubated with PcTF and cell penetrating peptides

(CPP), followed by 2. 24-hour incubation. 3. Imaging using fluorescent microscopy. (B) Control wells of cells only and PcTF with no CPP show no detectable mCherry signal. (C) Cells treated with PcTF and CPPs show varying levels of cell penetration.

References:

1. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7,** 21–39 (2008).

2. Wilson,R.C. and Gilbert,L.A. (2017) The Promise and Challenge of in Vivo Delivery for Genome Therapeutics. *ACS Chem. Biol.*, **13**, 376–382.

3. Zelikin,A.N., Ehrhardt,C. and Healy,A.M. (2016) Materials and methods for delivery of biological drugs. *Nat. Chem.*, **8**, 997–1007.

4. Du,S., Liew,S.S., Li,L. and Yao,S.Q. (2018) Bypassing Endocytosis: Direct Cytosolic Delivery of Proteins. *J. Am. Chem. Soc.*, **140**, 15986–15996

5. Rees,H.A., Komor,A.C., Yeh,W.H., Caetano-Lopes,J., Warman,M., Edge,A.S.B. and Liu,D.R. (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.*, **8**, 1–10.

6. Zuris,J.A., Thompson,D.B., Shu,Y., Guilinger,J.P., Bessen,J.L., Hu,J.H., Maeder,M.L., Joung,J.K., Chen,Z.Y. and Liu,D.R. (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.*, **33**, 73–80.

7. Woo,J.W., Kim,J., Kwon,S. Il, Corvalán,C., Cho,S.W., Kim,H., Kim,S.-G., Kim,S.-T., Choe,S. and Kim,J.-S. (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.*, **33**, 1162–1164.

8. Staahl,B.T., Benekareddy,M., Coulon-Bainier,C., Banfal,A.A., Floor,S.N., Sabo,J.K., Urnes,C., Munares,G.A., Ghosh,A. and Doudna,J.A. (2017) Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.*, **35**, 431–434.

9. Wang,H.-X., Song,Z., Lao,Y.-H., Xu,X., Gong,J., Cheng,D., Chakraborty,S., Park,J.S., Li,M., Huang,D., *et al.* (2018) Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide. *Proc. Natl. Acad. Sci.*, **115**, 4903–4908.

10. Kintzing,J.R., Filsinger Interrante,M. V. and Cochran,J.R. (2016) Emerging Strategies for Developing Next-Generation Protein Therapeutics for Cancer Treatment. *Trends Pharmacol. Sci.*, **37**, 993–1008.

11. Bruce,V.J. and McNaughton,B.R. (2017) Inside Job: Methods for Delivering Proteins to the Interior of Mammalian Cells. *Cell Chem. Biol.*, **24**, 924–934.

12. Guidotti,G., Brambilla,L. and Rossi,D. (2017) Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends Pharmacol. Sci.*, **38**, 406–424.

13. Okamoto,Y., Kojima,R., Schwizer,F., Bartolami,E., Heinisch,T., Matile,S., Fussenegger,M. and Ward,T.R. (2018) A cell-penetrating artificial metalloenzyme regulates a gene switch in a designer mammalian cell. *Nat. Commun.*, **9**, 1–7.

14. Krosl,J., Austin,P., Beslu,N., Kroon,E., Humphries,R.K. and Sauvageau,G. (2003) In vitro expansion of hematopoietic stem cells by recombinant TAT-HOXB4 protein. *Nat. Med.*, **9**, 1428–1432.

15. Michiue,H., Tomizawa,K., Wei,F.Y., Matsushita,M., Lu,Y.F., Ichikawa,T., Tamiya,T., Date,I. and Matsui,H. (2005) The NH2 terminus of influenza virus hemagglutinin-2 subunit peptides enhances the antitumor potency of polyarginine-mediated p53 protein transduction. *J. Biol. Chem.*, **280**, 8285–8289.

16. Essafi, M., Baudot,A.D., Mouska,X., Cassuto,J.-P., Ticchioni,M. and Deckert,M. (2011) Cell-Penetrating TAT-FOXO3 Fusion Proteins Induce Apoptotic Cell Death in Leukemic Cells. *Mol. Cancer Ther.*, **10**, 37–46.

17. Kwon,Y. Do, Oh,S.K., Kim,H.S., Ku,S.Y., Kim,S.H., Choi,Y.M. and Moon,S.Y. (2005) Cellular manipulation of human embryonic stem cells by TAT-PDX1 protein transduction. *Mol. Ther.*, **12**, 28–32.

18. Mlambo,T., Nitsch,S., Hildenbeutel,M., Romito,M., Müller,M., Bossen,C., Diederichs,S., Cornu,T.I., Cathomen,T. and Mussolino,C. (2018) Designer epigenome modifiers enable robust and sustained gene silencing in clinically relevant human cells. *Nucleic Acids Res.*, **46**, 4456–4468.

19. Kelly,T.K., De Carvalho,D.D. and Jones,P.A. (2010) Epigenetic modifications as therapeutic targets. *Nat. Biotechnol.*, **28**, 1069–1078.

20. Chevalier,A., Silva,D.-A., Rocklin,G.J., Hicks,D.R., Vergara,R., Murapa,P., Bernard,S.M., Zhang,L., Lam,K.-H., Yao,G., *et al.* (2017) Massively parallel de novo protein design for targeted therapeutics. *Nature*, **550**, 74–79.

21. Tekel, S. J. & Haynes, K. A. Survey and summary: Molecular structures guide the engineering of chromatin. *Nucleic Acids Res.* **45,** 7555–7570 (2017).

22. Tekel, S. J., Barrett, C., Vargas, D. & Haynes, K. A. Design, Construction, and Validation of Histone-Binding Effectors in Vitro and in Cells. *Biochemistry* **57,** 4707–4716 (2018).

23. Nyer, D. B., Daer, R. M., Vargas, D., Hom, C. & Haynes, K. A. Regulation of cancer epigenomes with a histone-binding synthetic transcription factor. *Genomic Med.* **2,** 1 (2017).

24. Wadia, J.S., Stan,R. V and Dowdy,S.F. (2004) Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.*, **10**, 310–315.

25. Lee, Y.J., Erazo-Oliveras,A. and Pellois,J.P. (2010) Delivery of macromolecules into live cells by simple co-incubation with a peptide. *ChemBioChem*, **11**, 325–330.

26. Akishiba, M., Takeuchi,T., Kawaguchi,Y., Sakamoto,K., Yu,H.-H., Nakase,I., Takatani-Nakase,T., Madani,F., Gräslund,A. and Futaki,S. (2017) Cytosolic antibody delivery by lipid-sensitive endosomolytic peptide. *Nat. Chem.*, **9**, 751–761.

27. Lönn,P., Kacsinta,A.D., Cui,X.S., Hamil,A.S., Kaulich,M., Gogoi,K. and Dowdy,S.F. (2016) Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.*, **6**, 32301.

28. Yang,Y., Ma,J., Song,Z. and Wu,M. (2002) HIV-1 TAT-mediated protein transduction and subcellular localization using novel expression vectors. *FEBS Lett.*, **532**, 36–44.

29. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the panther classification system. *Nat. Protoc.* **8,** 1551–1566 (2013).

30. Lasry, A. & Ben-Neriah, Y. Senescence-associated inflammatory responses: Aging and cancer perspectives. *Trends Immunol.* **36,** 217–228 (2015).

31. Haynes, K. A. & Silver, P. A. Synthetic Reversal of Epigenetic Silencing. *J. Biol. Chem.* **286,** 27176–27182 (2011).

32. Rouet, R., Thuma, B.A., Roy, M.D., Lintner, N.G., Rubitski, D.M., Finley, J.E., Wisniewska, H.M., Mendonsa, R., Hirsh, A., De Oñate, L., *et al.* (2018) Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J. Am. Chem. Soc.*, **140**, 6596–6603.

33. Kato,T., Tanaka,M. and Oba,M. (2013) Protein transfection study using multicellular tumor spheroids of human hepatoma Huh-7 cells. *PLoS One*, **8**, 1–8.

34. Tekel, S. J., Vargas, D. A., Song, L., LaBaer, J., Caplan, M. R., and Haynes, K. A. (2018) Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. ACS Synth. Biol. 7 (3), 842−852.

35. Oliveros, J. C. Venny. an interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html (2007–2015).

CHAPTER 5

A NEXTGEN HISTONE POSTTRANSLATION MODIFICATION BINDING ASSAY, CONCLUSIONS, AND FUTURE WORK

Nextgen PTM Binding Analysis: Peptide Arrays

Classic *in vitro* characterization of histone binding domains (HBDs) generally includes protein purification followed by radioactive labeling, low dynamic range dot blots, fluorescent polarization, isothermal calorimetry, or other *in vitro* methods (1-3). Our previous work has described the development and validation of an assay to detect histone posttranslational modification (PTM) binding at a small scale (a few binding variants and a few target peptides) using an enzyme linked immunosorbent-based assay (ELISA), followed by in cell functional validation. While our ELISAs have successfully demonstrated a proof of concept approach to validating binding of novel histone binding fusions, the methods suffer from a few limitations. First, 96 well plate based ELISAs can only contain a few immobilized histone peptide tails per plate, resuling in low specificity throughput. Second, expression and purification of large fusion proteins is time consuming and can be difficult due to solublity of fusions. For example, the bivalent version of PcTF, Pc_2TF, requires overnight, room temperature induction to produce soluble protein (4). This suggests that other large fusion proteins might experience similar solubility problems when recombinantly expressed in *E. coli*. To address these limitations, we set out to develop a next generation (nextgen) protocol to screen novel PTM binding proteins for function and specificity.

In order to address assay length and to develop a screening technique better capable of testing specificity, we approached our assay optimization from two directions. First, to decrease variant protein production time, we utilized TXTL, an *E. coli* based cell free in vitro transcription and translation (IVTT) system (5). Hard to express proteins such as bivalent PcTFs benefit from the lower incubation temperature (30°C) of TXTL by enabling proper protein folding. In addition, we can quickly express up to 96 assay ready protein variants within 16 hours. This system should be expandable to 384 well plates to further increase variant expression throughput. Second, to increase the number of histone PTM peptide variants assayed to reduce assay run time, we used a prefabricated MODified™ Histone Peptide Array (Active Motif) that contains 384 unique modifications immobilized on a microscope slide. Herein, we validate an improved assay for determining function and specificity of the engineered histone binding protein Pc_2TF (Figure 5.1).

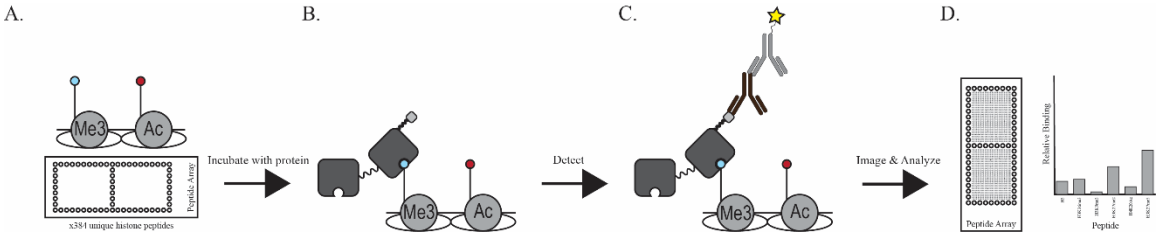


Figure 5.1 General peptide array protocol. (A) An array of pre-immobilized histone peptides containing PTMs on a glass slide. (B) IVTT expressed protein is incubated with slide to allow binding to histone tails. (C) Binding is detected through a primary anti-tag antibody followed by a fluorophore-conjugated secondary antibody. (D) Slide is scanned and analyzed for PTM binding.

Results and Discussion

To demonstrate feasibility of this nextgen histone PTM binding assay, we first expressed a large quantity of Pc_2TF using TXTL. We used 15 wells of a 96 well plate to produce protein, allowing at least six variants to be produced in one plate for nextgen binding anaylsis. For this work, we decided to use a single slide for one variant to allow measurement of technical replicates for each histone modification (Figure 5.1A). The slides are capable of being partitioned to enable two variants per slide, increasing throughput. However, a single technical replicate for variant binding makes the data less reliable. The histone peptide array was blocked overnight to reduce background binding before being incubated with the protein-lysate mix. Blocking the night before removed the need for blocking the morning of the assay, decreasing assay time. We incubated the slides with unpurified, TXTL expressed Pc_2TF to allow binding (Figure 5.1B). Avoiding protein purification saves a significant amount of time by enabling immediate assaying of expressed protein. Next, we detected Pc_2TF binding through a polyclonal anti-RFP primary antibody followed by signal amplification with a polyclonal anti-primary Cy5-conjugated antibody (Figure 5.1C). Since we were using such a small quantity of unpurified protein, we reasoned that using two antibodies for signal amplification would enable us to maximize our signal: noise. To detect binding, we used a general microscope slide array scanner (Figure 5.1D). A quick, two-step protocol using free software analyzed the data.

The scanned peptide microarray contains two identical subarrays, each containing 384 unique histone peptide modifications (Figure 5.1A). The top subarray incubated with Pc_2TF (Figure 5.2) shows signal at multiple spots, indicating that Pc_2TF binds to more than one PTM. We observe little signal around each spot, indicating that there is low

background signal on the slide (Figure 5.2). This suggests that the TXTL lysate mix does not interfere with fusion protein binding, or fusion binding detection in this assay.

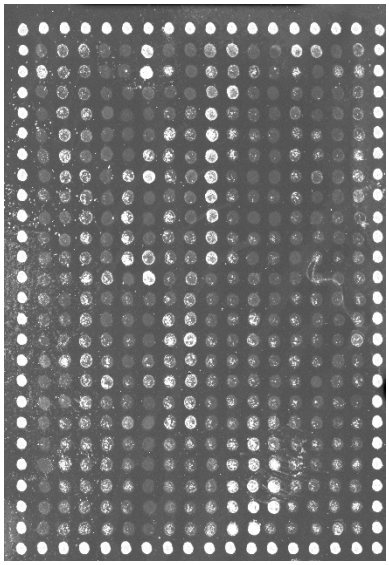


Figure 5.2 Pc_2TF peptide subarray (Top array). IVTT expressed Pc_2TF is incubated with a peptide array and is detected with an anti-mCherry primary followed by a Cy_5 conjugated secondary. Image scanned at 5 µm resolution, 635 nm excitation (Cy_5), and saved as a .TIFF. Top array cropped in Adobe Fireworks. Each spot represents a unique histone tail PTM.

Precision and accuracy are important aspects of a high throughput assay. Since each array contains two identical subarrays, we are able to assess precision by comparing signal intensities of the same PTM on each array. Therefore, we plotted the intenisites of each spot from each array against the other, expecting to see good correlation from a precise array (Figure 5.3A). The Pc_2TF subarrays showed some spots clustering at at the dashed line, however, the R squared value comparing the signal of each array to each

other was only R^2=0.41, indicating that there are some major variances in signal between the two subarrays. This could be due to poor washing, incubation, or detection. Further optimization of these conditions will help reduce the variance. However, the error for each PTM can be independently compared using the raw data, allowing extraction of reliable data from a set with unusable spots. For each individual PTM, a spot error can be calculated by averaging intensities across the single spot. We calculated the spot error range for the Pc_2TF array, with lesss than 5% error for over 250 of the peptides (Figure 5.3B). This suggests that overall; there is low error within a single spot. This error could be reduced by ensuring equal distribution of protein lysate mix across the slide throughout the incubation step.

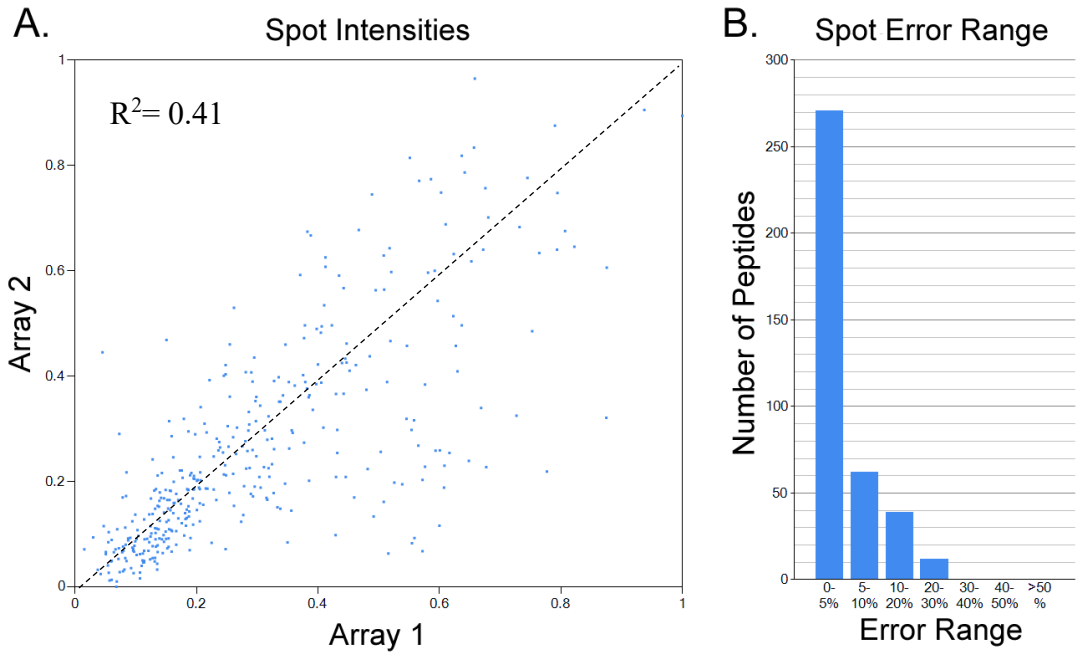


Figure 5.3 Peptide array variance and error. (A) Signal intensity of Array 1 plotted against signal intensity of Array 2. (B) Spot error range for each modified peptide. Error range calculated measuring differences in subarray spot intensities.

We cannot determine the binding affinity of Pc_2TF using this type of microarray-based assay because the post incubation washing steps disrupt the binding equilibrium. However, we are able to determine relative binding and specificity, which provides good information on engineered binding domains. To enable easier interpretation of the relative binding, we set the histone PTM with the highest binding signal equal to one. We then compared the other PTM binding spots with values over the background signal (auto set by software) (Figure 5.4). Not surprisingly, the strongest binding event occured between Pc_2TF and a spot was with our expected target ligand, H3K27me3 (Figure 5.4). We observed additional binding by single or doubly methylated H3K27 (me2/me3), as well as H3K36me1, 2, and 3. The binding to H3K36 is unexpected because it does not share the ARKS sequence motif from H3K27 (1). Surprisingly, we also detected binding to some acetylated histones (Figure 5.4), which are not known to interact with chromodomains at all. In addition, Pc_2TF showed modest binding to H3K14Ac and H2AK13Ac (Figure 5.4). Since there have been no previous report on the interactions of CBX8 fusions on this number of histone tail PTM variants, our binding results might indicate additional uncharacterized protein-histone tail PTM interactions. In order to draw conclusions, additional biological replicates will be needed to ensure accuracy of these binding results. Furthermore, to determine if there are any non-CBX8 related interactions with the histone array, this assay should be performed with the nonbinding domain ($Pc_\Delta TF$) control. Lack of signal on a $Pc_\Delta TF$ array would indicate no off target interactions, and further reinforce the binding results of the Pc_2TF array.

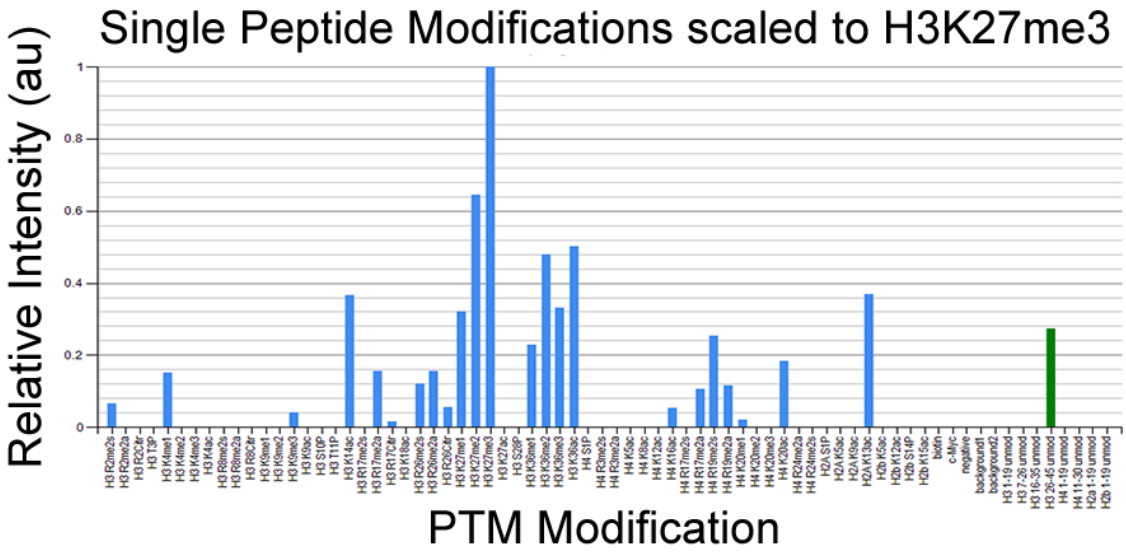


Figure 5.4 Relative binding of Pc_2TF to singly modified peptides. Binding is scaled to the highest binding PTM H3K27me3 (max is 1).

The protocols and results described in this chapter are preliminary and can be improved with a few optimizations. First, optimizing the quantity of TXTL lysate needed for good signal to noise will enable a minimum input of material, saving money and allowing more variants to be expressed at once. Second, optimizing washes, blocking, and incubation time, will maximize throughput by reducing time for each step while simultaneously enabling the highest signal to noise for this assay.

Overall, we are the first to describe a high throughput screening method that combines that ease, speed, and versatility of IVTT expressed protein, coupled with the diversity of prefabricated histone tail microarrays for determining function and specificity of novel histone binding proteins. This assay will enable synthetic biologists to quickly screen novel histone binding domains for *in vitro* function and specicity.

Methods

In vitro transcription and translation (IVTT) was performed as described in chapter 2: Real Time Detection of Cell-Free Transcription-Translation (TXTL). In brief, 9 µL of Sigma 70 Master Mix (myTXTL, Arbor Biosciences), 1 µL of 5 nM P70a-T7 RNA pol vector (Arbor Biosciences), and 2 µL of 50 ng/µL pET28-Pc_2TF in a total volume of 12 µL in one well of a white, 96-well plate. A total of 15 reactions were assembled. The reaction was incubated for 16 hours at 30 °C to allow for transcription and translation of Pc_2TF. After incubation, the samples were pooled to obtain approximately 150 µL of Pc_2TF –TXTL lysate mix.

The peptide array was blocked the day before the assay by incubating the slide in 5% BSA in 0.2% PBST at 4°C overnight with shaking. On the day of the experiment, the slide was washed three times with 3 mL of 0.2% PBST with agitation, 5 minutes each. 150 µL of Pc_2TF –TXTL lysate mix was diluted into 1 mL 5% skim milk in 0.2% PBST and incubated on the slide at 4°C with agitation for 1 hour, ensuring the liquid evenly covered and was flowing across slide. The slide was washed three times in 3 mL of 5% skim milk in 0.2% PBST with agitation, 5 minutes each. Binding was detected by diluting rabbit anti-mCherry (ab167453) 1:3000 in 5% skim milk in 0.2% PBST and incubating the slide for 1 hour at 4°C with agitation. The slide was washed three times with 3 mL of 5% skim milk in 0.2% PBST with agitation, 5 minutes each. Signal was amplified by incubating the slide with polyclonal goat anti-rabbit Cy5 (ab6564) 1:3000 in 5% skim milk in 0.2% PBST and incubating the slide for 1 hour at 4°C with agitation, followed by 5 final washes with 3 mL of 0.2% PBST for 3 minutes each.

The slide was scanned on a GenePix 4000B (Molecular Devices) using standard settings (5µM resolution, dual excitation channels). The output was saved as a .TIFF file.

Analysis was performed in the Array Analyze Software (Active Motif). In brief, a .TIFF was imported into the software and cropped to only display the two subarrays. A left and right bound were set to allow for optimal automatic spot positioning. The software performed automatic background calculation, error, and binding intensity analysis. Data was exported as a .csv for additional analyses.

Rational Design of High Specificity PTM Binding Proteins

In our previous work, we used mutlivalency as a solution to increase binding and function of synthetic HBPs. We took endogenous binding domains and generated the first reported double chromodomain that functions in cells. However, this fusion consists of a CBX8 peptide fragment sourced from nature without any binding optimization (i.e. affinity or specificity). Evolution has selected HBPs to perform a specific task, such as binding to multiple histone marks to maintain differentiations (3), which is not always a desired function in synthetic systems. One might envision a histone binding protein that can bind specifically and with high affinity to only one PTM for a highly specific response. A HBD that could bind to an unnaturally deposited PTM could enable orthogonal epigenetic regulation without interference from natural HBPs. Since many histone-binding domains are promiscuous in binding (6-7) or are part of a complex that can bind to several different modified histone tails (8), finding protein domains directly from nature that can meet our stringent functional expectations can be challenging. Luckily, synthetic biology tools allow us to engineer novel and improved function in proteins (9). Rational design and directed evolution (or a combination of both) are strategies that can lead to histone binding proteins with higher affinity and specificty.

Here, I briefly describe how rational design and directed evolution can be used to generate a modified CBX8 for higher affinity and specificity for H3K27me3 (Figure 5.5).

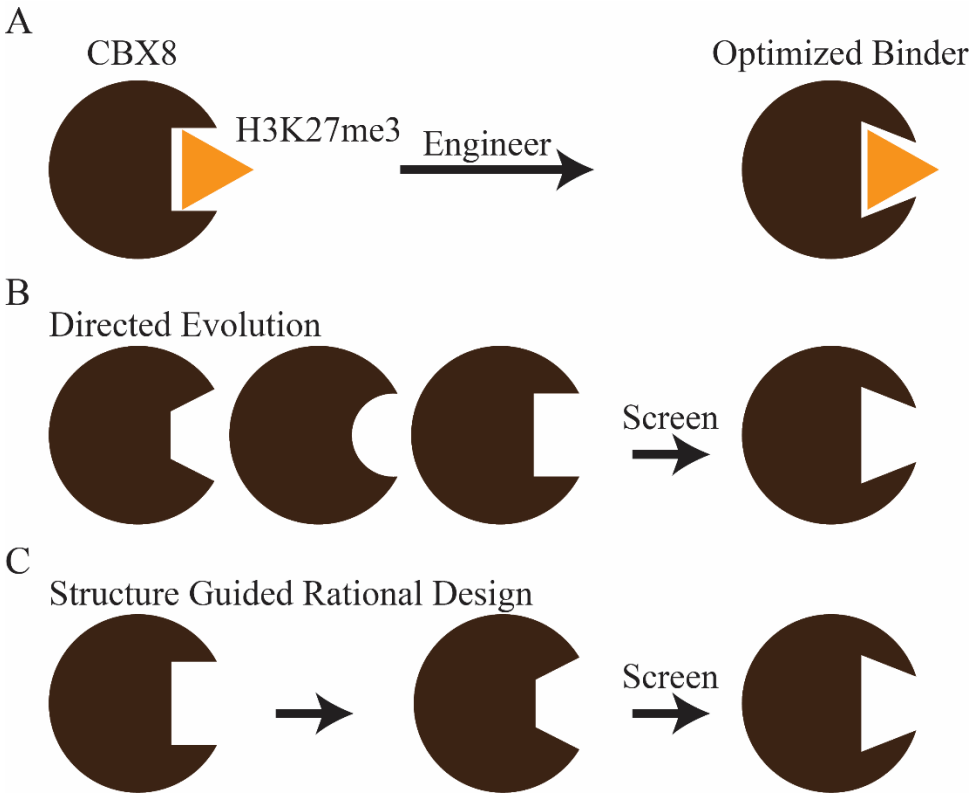


Figure 5.5 Methods to optimize CBX8 binding. (A) Engineering CBX8 for improved binding to H3K27me3. (B) Directed evolution using a random or semi rational library can enrich a library of variants for better binding and specificity. (C) Structure guided rational design can lead to mutations that increase affinity and/or specificity for H3K27me3.

Understanding which mutations in HBDs influence PTM binding can lead to design principles for engineering synthetic chromatin systems. Directed evolution is an approach to optimizing protein binding that works by applying a selective pressure to a library of rationally designed, semirationally designed, or random protein variants (Figure

5.5B). Through iterative rounds of selective pressure, directed evolution enriches a pool of variants for improved function (10-11). One method of directed evolution, mRNA display (Figure 5.6), is a good candidate for evolving CBX8 into a higher affinity binder because the process has a higher chance of success with smaller proteins (less than 110 amino acids) (12). In addition, starting with innate binding or activity for a target ligand greatly improves the chance of successful optimization (13). A round of selection using mRNA display would start with a random or semi rational library of DNA encoding CBX8 variants and associated regulatory units (Figure 5.6). First, the library is transcribed into RNA (Figure 5.61), followed by photo crosslinking of the mRNA (Figure 5.62) to a puromycin moiety. The crosslinked RNA is translated in a cell free system (Figure 5.63), followed by purification of RNA-protein fusions on an oligo-dT column (Figure 5.64). Reverse transcription (Figure 5.65) increases stability of the protein-nucleic acid complex and enables downstream PCR, which is followed by binding fusions on an H3K27me3 immobilized column (Figure 5.66) in the presence of free H3K27me3 competitor. This step selects for better binders by enabling the lower affinity binders to dissociate and bind to free competitor. Including a negative selection by flowing the RNA-peptide fusions over a column immobilized with a similar PTM before this step would help enrich for specificity as well as affinity, since off target binders would be left immobilized on the negative selection column. Finally, the remaining fusions are competitively eluted off the column with high concentrations of H3K27me3 and PCR amplified to enrich the population for better binders before repeating the process. After a number of cycles, the library is sequenced to look for consensus sequences and mutations. Using a strategy such as mRNA display has been successful in

evolving de novo binding proteins, as well as increasing affinity of proteins (14-15), suggesting its plausibility in selecting a CBX8 variants for enhanced H3K27me3 binding. Unfortunately, mRNA display requires expensive and specialized equipment (16) and is very dependent on the quality of the selection strategy for enrichment. Due to these restrictions, we investigated additional approaches to engineer CBX8 for higher affinity.

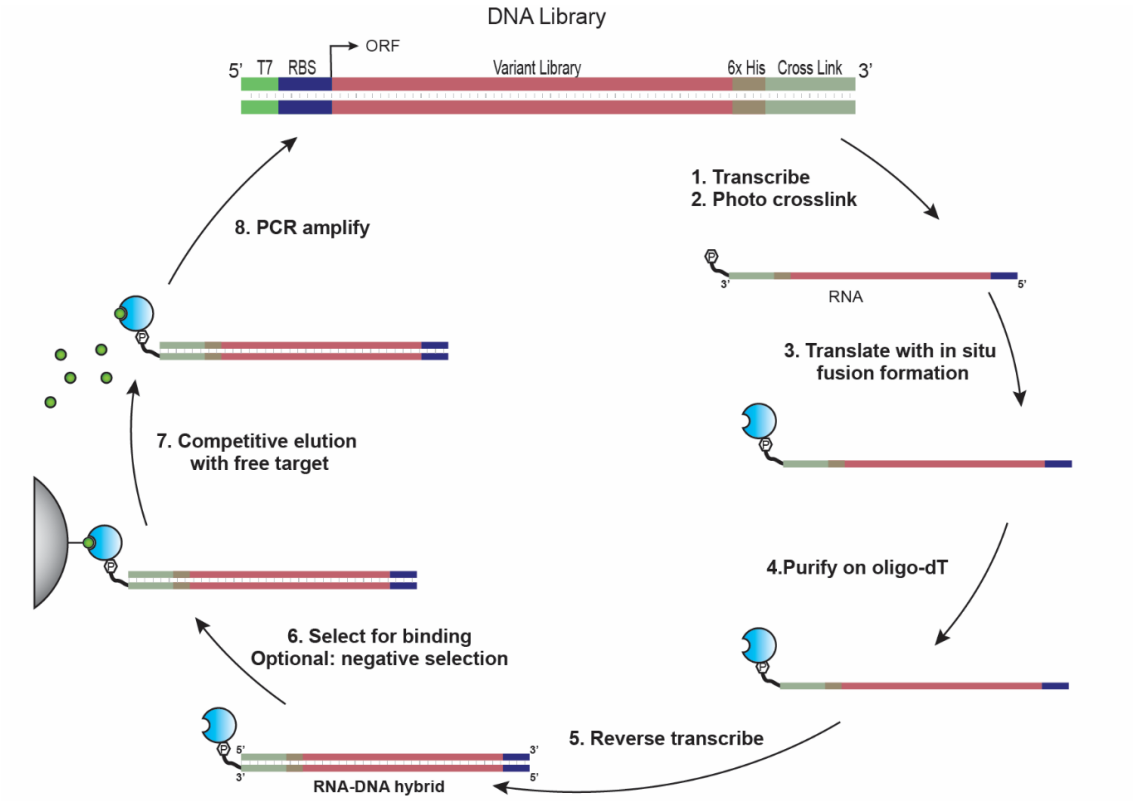


Figure 5.6 Directed evolution of PTM-binding proteins. A DNA variant library (top) is (1) transcribed into RNA followed by (2) RNA photo crosslinking to a puromycin moiety. (3) The cross-linked RNA is translated into protein using a cell free system, followed by purification of protein-RNA fusions on an oligo-dT column. The fusions are eluted from the oligo-dT column and (5) reverse transcribed to form RNA-DNA hybrids

covalently fused to their corresponding protein. (6) A selection (positive, negative, or both) is performed to enrich for binding, followed by (7) competitive elution with free target molecule. The eluted fusions are then (8) PCR amplified to generate templates for the next round of selection.

Coupling structures of protein-ligand interactions with computer modeling can help predict point mutations that can lead to a desired function (Figure 5.5C) (17). Due to the extensive *in vitro* characterization of the CBX histone binding proteins, we set out to use rational design to predict point mutations that might enhance the binding affinity of CBX8 for H3K27me3. We envision a standard workflow to optimize histone-binding proteins *in silico* (Figure 5.7) starting with a HBD that has a solved structure. For example, the structure of CBX8 has been elucidated and is a good proof of concept domain for binding optimization (Figure 5.7A). Alignments of CBX8 with the other CBX orthologs might elucidate mutations that could increase affinity (Figure 5.7B). A more software-based approach couples the structure of a CBX8-H3K27me3 complex with a minimal free energy analysis using Rosetta (18) (Figure 5.7B) to predict single mutations within the protein that might increase binding affinity. The output of either approach (Figure 5.7C) can either be used as an input for further *in silico* characterization or be tested for function in an *in vitro* assay (Figure 5.7D). Any proteins that demonstrate improved function *in vitro* can be further tested as fusions with effector domains for in cell activity (Figure 5.7E). The main caveat of computer-based approaches is the need for programming knowledge and familiarity with modeling software, as well as computers with strong processors, which can be cost prohibitive.

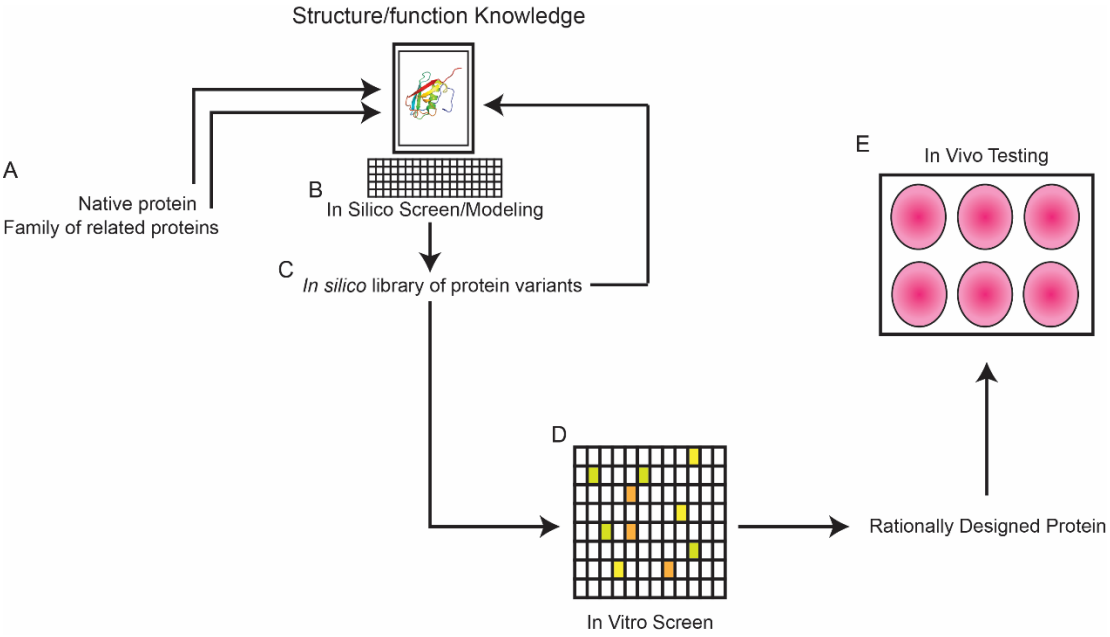


Figure 5.7 Standard workflow to optimize protein binding using structure/function knowledge. (A) Native proteins or homologous sequences that have solved structures act as a starting point for optimization. (B) Structures are fed into a software such as Rosetta (https://www.rosettacommons.org/) to optimize protein-ligand interaction or are aligned to determine possible residues responsible for affinity. (C) A library of sequences is returned with predicted binding to target and can be reused as input for *in silico* modeling. (D) An *in vitro* assay finds higher binding variants. (E) New variants are tested as fusions to effector domains for enhanced function in cells.

Since CBX8 is a relatively small protein (~80 amino acids), we hypothesized that we could rationally design single base mutations using the available *in vitro* binding data of CBX orthologs along with primary structure alignments. The human CBX orthologs that can bind to H3K27me3 have been extensively characterized *in vitro*, with available data on both three dimensional structure and target binding affinity (1-2).

We aligned the amino acid primary structures of the human CBX orthologs CBX 2, 4, 6, 7, and 8, sorting them high to low by their *in vitro* binding affinity to H3K27me3 (Figure 5.8). Interestingly, there is a large variance in reported binding affinities for the CBX orthologs using the same technique (fluorescence polarization), which might be due to the different buffers used in the experiment or other experiment specific parameters. Regardless, the order of the orthologs based on binding affinity remains the same. The similarity of the consensus sequence to each ortholog reinforces the structural similarity of each CBX homolog, and how just a few amino acid changes must be responsible for the large variance in binding affinity (Figure 5.8). For example, the primary structure of CBX7 varies by about 15% when compared to other orthologs, but has an affinity for H3K27me3 up to 15 times greater (1). We hypothesized that one or several of the amino acid differences in the orthologs that are responsible for high affinity binding in CBX7 might be portable into the CBX8 chassis, increasing the binding affinity for H3K27me3. However, several of these higher affinity orthologs (i.e. CBX7 and CBX2) have increased affinities for off target PTMs, such as H3K9me3 (1-2). Therefore, the beneficial mutations that might increase affinity for H3K27me3 have to be weighed against the possibility of acquiring off target binding that was previously absent in CBX8 (1-2).

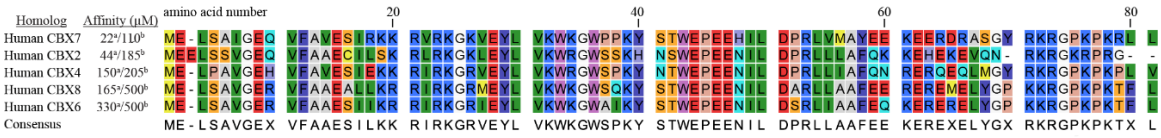


Figure 5.8 CBX sequence alignment based on H3K27me3 affinity. Human CBX homologs are aligned and assembled from highest binding affinity (low K_d) to lowest

binding affinity (high K_d). Sequences obtained from NCBI. Alignments assembled using CLC Main. [a](1), [b](2).

To focus on the orthologs with a higher affinity for H3K27me3 compared to CBX8, we aligned the primary structure of the two orthologs with the highest affinity for H3K27me3, CBX7 and 2, to CBX8. This resulted in a consensus sequence being 78% similar (64/82 amino acids) to wild type CBX8 (Figure 5.9). Of the 18 varying positions, 7 of these residues are in an unconserved region starting after amino acid 61 (Figure 5.9). Of the remaining 11 residues, several are conserved between CBX7 and CBX2 and are not found in CBX8. For example, residues 9 (glutamine), 16 (isoleucine), 25 (lysine), 51 (proline), 60 (lysine), and 70 (arginine) are shared between CBX7 and CBX2, but not CBX8. This suggests that one or a combination of these residues could be responsible for the increased affinity for H3K27me3. In fact, it was previously suggested that CBX8 has a lower affinity for H3K27me3 because of its overall electropositive charge (2). Therefore, removing some of the positive charges by mutating them to an uncharged residue or a less basic residue could improve affinity. A library of CBX8 mutants containing a distribution of the mutations R10Q, L17I, R26K, A52P, R61K, and K71R would be a good starting point for screening (Figure 5.9). To reduce the number of variants to be constructed and assayed, we sought to use structural data to help us rationally elucidate the impact of each mutation on the binding to H3K27me3.

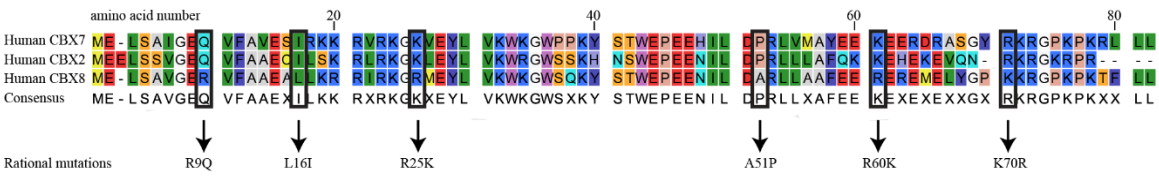


Figure 5.9 Primary structure of CBX8 vs CBX7 and CBX2. Rational mutations for CBX8 are based on the consensus amino acids from CBX orthologs that have a higher affinity for H3K27me3.

There is no available solved structure of a CBX8-H3K27me3 complex, but there is structural data for the ortholog CBX7 complexed with H3K27me3, and CBX8 complexed with H3K9me3. Using these structures, we used Pymol (19) to align the CBX8 crystal structure and the NMR state of CBX7 that most closely resembles the CBX8 structure. We then removed the H3K9me3 and the CBX7, leaving an approximate CBX8-H3K27me3 complex (Figure 5.10). The red residues comprise the conserved hydrophobic core consisting of Phenylalanine 11 (F11), Tryptophan 32 (W32), and Tryptophan 35 (W35) that encapsulates the trimethyl moiety (me3) (Figure 5.10A). The gray peptide represents histone 3 peptides 19-33. This peptide containing H3K27me3 traverses the side of the CBX8, highlighting the protein-ligand interactions (Figure 5.10B). We highlighted the proposed residues for mutation from Figure 5.9 on the structure CBX8 to elucidate how they might interact with H3K27me3. We reasoned that we might be able to determine the contribution of each mutation to binding of H3K27me3.

Two highlighted residues appear to have direct interactions with the H3K27me3 peptide, arginine 25 (R25) and arginine 9 (R9) (Figure 5.10). Kaustov et al previously noted that the positive charge from arginine 9 might associate less efficiently with a basic histone peptide, lowering the overall binding affinity (2). Introducing the R25Q mutation would eliminate the basic charge at that interaction and might allow for higher binding affinity. Introducing R25K in CBX8 would not remove the positive charge, but it does

replace arginine with a slightly less basic residue (pK_a of 12.48 vs 10.79), which might allow for a stronger protein-ligand interaction.

The other predicted mutations (L16I, A51P, and R60K) have a much less obvious impact on binding. Alanine 51 (A51) and arginine 60 (R60) reside in the alpha helix of CBX8 that does not look like it directly interacts with H3K27me3 (Figure 5.10). Perhaps, in addition to arginine 9, mutating this arginine residue to a less basic lysine would help the overall association and binding to histone tails. In addition, leucine 16 (L16) resides in a linker region between a beta sheet and turn (Figure 5.10A). A proline might introduce a more rigid turn that stabilize the protein-ligand interaction.

These less obvious residues could also carry an unknown significance that contributes to binding through overall protein stabilization. For example, Hatley et al demonstrated that a network of amino acids could allosterically mediate binding to GTP (20). Perhaps one of these residues has a more subtle impact on binding that cannot be elucidated in *silico*. Ultimately, these mutations will have to be introduced as a library of CBX8 mutants and screened to determine the impact of each mutation. Cycles of design, build, and test, could result in a set of improved binders that could be further optimized through additional *in silico* or *in vitro* testing (Figure 5.11).

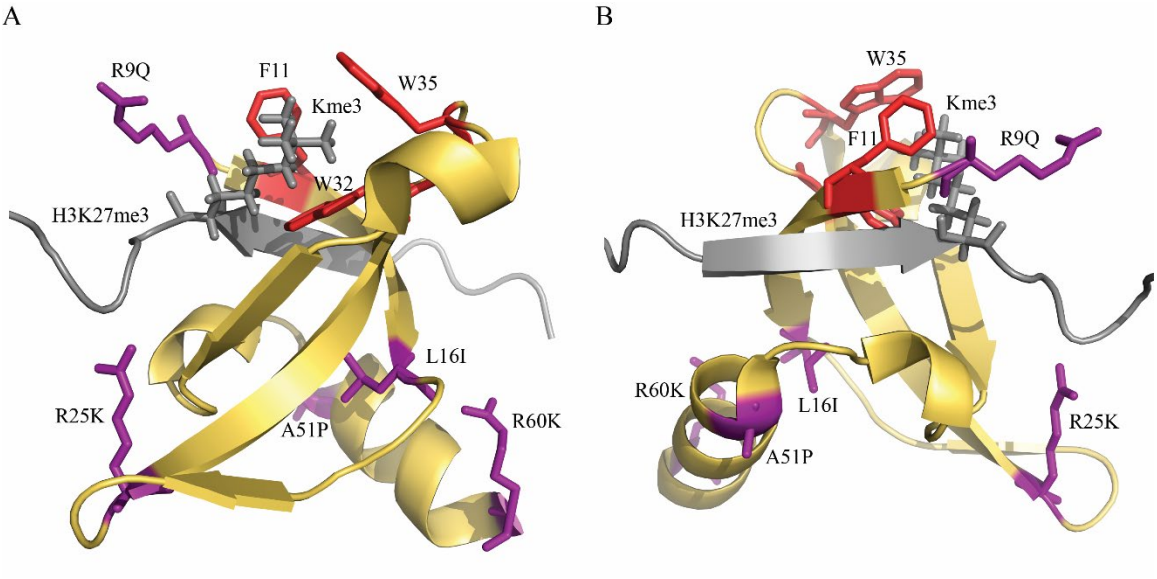


Figure 5.10 3D structure of CBX8 highlighting residues hypothesized to impact binding affinity based on primary structure alignment of higher affinity orthologs. CBX8 (PDB 3I91) complexed with H3K9me3 is aligned with CBX7 (PDB 2L1B) complexed with H3K27me3. CBX7 and H3K9me3 are removed to approximate a CBX8-H3K27me3 interaction (A) View of complex showing hydrophobic core interaction with trimethyl lysine (Kme3). (B) View of complex highlighting H3K27me3 interaction along CBX8 backbone. Yellow residues indicate CBX8 backbone. Purple sticks highlight human wild type CBX8 residues that would be mutated as shown. Gray chain shows H3K27me3 peptide residues 19-33. Red residues indicate hydrophobic cage for trimethyl (me3) moiety. Figure generated with PyMol (19).

Expanding H3K27me3 binding protein variant library size by including orthologs from other species can increase the likelihood of finding a better binder. Both fruit fly (drosophila melanogaster) and zebrafish (danio rerio) have polycomb proteins with differing primary structure (21), which could mean they have differing affinities for their target ligands. However, many polycomb proteins are expressed during development (22-

23), and each species has a different optimal developmental temperature (24-25) which might be required for optimal binding and function of the endogenous polycomb proteins. This study focused on human polycomb proteins because it allowed us to minimize the potential for misfolding or functional differences due to native expression temperature and conditions when used in human cells. In addition, structural data is not available for every ortholog from every species, but modeling software such as Rosetta (https://www.rosettacommons.org/), or additional *in vitro* structural studies can help fill in those gaps. Iterative processes of design, build, and test (Figure 5.11) can be applied to other histone binding proteins such as bromodomains or PHD domains that target different PTMs, resulting in an expanded synthetic histone binding protein toolbox for engineering biological systems.

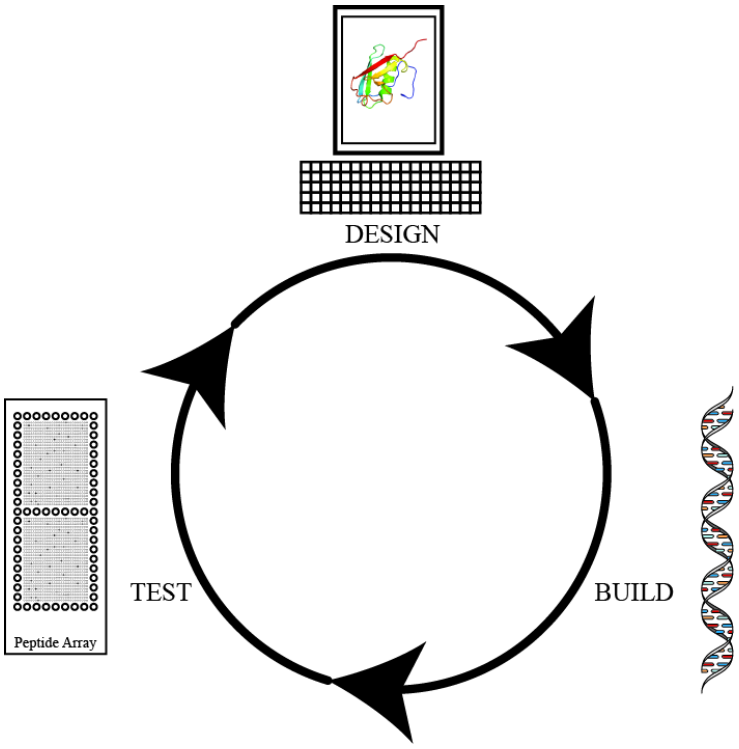


Figure 5.11 Design, build, test. Structural and biochemical data guides design of protein variants. High throughput DNA synthesis enables facile library assembly. A series of *in vitro* and in cell assays determines the function and specificity of the variants. The variants with desired functions are used as input for further optimization.

Future Work

Significant work is still required to elucidate all of the rules for designing synthetic histone-binding proteins. Since new technology has enabled fast DNA library synthesis, building constructs is no longer a bottleneck in the design, build, and test cycle. Histone PTM binding assays developed in this thesis will suffice for testing novel fusion protein functionality, but need to be further optimized to test as many variants in parallel as possible. In addition to computer modeling, a directed evolution approach should be taken to test a larger library of variants for function. This process could enrich for expected active site mutations, or allosteric mutations that contribute to affinity that rational design might miss. Additional ration design rules that arise from these experiments can help synthetic biologists engineer other histone binding domains. The best mutants from these screens should be tested in live cells as fusions with effector domains to assay for in cell function. Cycles of design, test, and build can be expanded to other PTM binding proteins that have characterization data available to engineer additional synthetic histone binding effectors.

Further optimization of protein delivery to live cells can result in a set of principles for designing proteins for recombinant delivery. This optimization can occur through a two-step process. First, by expanding the variety of different cell penetration tags tested, we can determine the best delivery method using the least amount of protein,

enabling a lower overall dose of protein while maximizing function. Second, delivering recombinant bivalent proteins might reinforce the *in vitro* data that demonstrates enhanced *in vitro* binding and in cell function compared to monovalent versions. This knowledge will expand our understanding of mutlivalency in epigenetics, enabling scientists to harness engineered histone binding proteins for other applications such as personalized medicine, agriculture, and the biotechnology industry.

Thesis Summary and Conclusions

In chapter 1, we introduced the concept of engineering chromatin proteins. We outlined the composition of chromatin and how each component has previously been engineered. We described how structural and biochemical information could be used to help rationally design new chromatin binding proteins. We then concluded with a proposed pipeline for *in vitro* screening of novel histone PTM binding proteins.

In chapter 2, we described and outlined our pipeline for *in vitro* screening of novel histone PTM binding proteins. As proof of concept for our pipeline, we assayed a small library of bivalent CBX8 fusions varying by species of CBX8 binding domain. We demonstrated an easy cloning strategy to assemble the constructs, followed by a cell free *in vitro* transcription and translation system that produced enough functional protein for assaying without need for further purification. Next, we performed ELISAs to determine relative binding to on (H3K27me3) and off (other me3 peptides) target histone peptides. We ported the highest relative binding variants into a mammalian expression vectors and transfected them into a cell line containing a silenced reporter gene under the control of H3K27me3. A three in one microplate assay determined the fusion protein expression and function in cells. In addition to demonstrating the feasibility of our pipeline, we

outlined recommended human cell lines for testing other PTM binding proteins using endogenous genes as output and described our best practices for chromatin immunoprecipitation (ChIP) to validate that the protein is binding to the expected PTM. Overall, we successfully demonstrated the functionality of our pipeline to test synthetic chromatin proteins for binding.

Chapter 3 investigated the impact of bivalency on PTM binding *in vitro* and in cells, expanding the available functional tests for engineered histone binding proteins. We characterized Pc_2TF, a bivalent version of PcTF that contains two CBX8 domains linked by a short, flexible $(GGGGS)_4$ linker. We started by screening a small set of Pc_2TF fusions varying in linker characteristics for binding to H3K27me3, resulting in Pc_2TF. We then recombinantly expressed and purified $Pc_{\Delta}TF$, PcTF, and Pc_2TF in order to determine their apparent binding affinity (K_d^{app}). We concluded that Pc_2TF shows enhanced avidity for H3K27me3 over PcTF. We then ported these constructs into mammalian constructs, and tested their ability to reactivate a synthetically silenced reporter locus. Compared to PcTF, Pc_2TF was able to reactivate transcription of an H3K27me3 silenced luciferase by a factor of two. This chapter presented the first *in vitro* and in cell characterization of a synthetic, multivalent chromatin-derived histone-binding protein.

In Chapter 4, we investigated cell delivery mechanisms for the recombinant, synthetic histone binding effector PcTF. We screened a variety of cell penetrating peptides co incubated and fused to PcTF. Our most promising candidate contained a short, cell penetrating peptide fused to the C terminus of PcTF (CP-PcTF). To determine characteristics of CP-PcTF delivery to cells, we compared varying concentrations of CP-

PcTF over varying time points to optimize nuclear delivery. We then demonstrated how incubating CP-PcTF enabled time dependent uptake into 2D and 3D U-2 OS cells. In order to determine the gene response to CP-PcTF, we performed RNA sequencing on both U-2 OS and foreskin fibroblasts treated with either CP-PcTF or CP-Pc$_{\Delta}$TF. We outlined the genes significantly up and downregulated by CP-PcTF, and explored plausible reasons for the specific responses. Finally, we demonstrated that U-2 OS cells treated with CP-PcTF experience a delayed growth compared to an untreated control or CP-Pc$_{\Delta}$TF. This work ourlines a foundation for testing newly designed histone binding protein fusions for function in cells without the complications associated with DNA or viral transfection.

Chapter 5 explored steps toward optimizing a PTM binding specificity assay, and outlined how structural data enables rational design of proteins. We started our assay optimization by expressing a large quantity of Pc$_2$TF using TXTL as a proof of concept. We directly incubated the protein-lysate mix with a peptide array consisting of 384 unique histone PTMs. Instead of the classic HRP detection, we detected the array with a fluorophore-conjugated antibody, followed by imaging on a generic array scanner. The preliminary results indicated that TXTL can express enough protein to detect binding on this array with our new detection method, and that Pc$_2$TF has the strongest binding interaction with H3K27me3. This protocol has improved the throughput of screening histone-binding proteins by enabling assay ready protein production *in vitro* and screening hundreds of peptides for binding at once.

Next, we hypothesized that we could leverage structural and biochemical data to design synthetic histone proteins as another approach to engineering higher affinity

binders. Since our work focused on engineering CBX8 fusions, we used the CBX8-H3K27me3 interaction as an example. We proposed a generic outline for optimizing binding using computer software, as well as using structural data coupled with biochemical binding affinity data. We discussed using directed evolution to increase affinity and specificity, along with an overview of mRNA display as a possible approach. Based on primary structure alignments, we rationally proposed several residues of interest for mutating in CBX8, and explored their possible contribution to overall binding affinity. Overall, this chapter explored steps to improve histone PTM binding assays, and two additional approaches to engineer synthetic chromatin effectors.

The work outlined in this thesis only begins to unravel the potential of synthetic epigenetics. We elucidate design rules for engineering histone binding proteins by characterizing novel HBDs and their functional delivery to cells. The work in this thesis will enable synthetic biologists to be able to engineer novel histone binding proteins and effectors using cycles of design, build, and test.

References

1. Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., and Allis, C. D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. 26, 2560−2569.

2. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G.A., Li, Z., Vedadi, M., Schapira, M. *et al.* (2011) Recognition and specificity determinants of the human Cbx chromodomains. *J. Biol. Chem.*, **286**, 521–529.

3. 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**, 983–94.

4. Tekel, S. J., Vargas, D. A., Song, L., LaBaer, J., Caplan, M. R., and Haynes, K. A. (2018) Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. ACS Synth. Biol. 7 (3), 842−852.

5. Shin,J. and Noireaux,V. (2012) An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.*, **1**, 29–41.

6. Filippakopoulos, P. and Knapp, S. (2012) The bromodomain interaction module. *FEBS Lett.*, **586**, 2692–2704.

7. Su, Z., Wang, F., Lee, J.-H., Stephens, K.E., Papazyan, R., Voronina, E., Krautkramer, K.A., Raman, A., Thorpe, J.J., Boersma, M.D., *et al.* (2016) Reader domain specificity and lysine demethylase-4 family function. *Nat. Commun.*, **7**, 13387.

8. 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., *et al.* (2011) Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell*, **145**, 692–706.

9. Rosenfeld, L., Heyne, M., Shifman, J.M. and Papo, N. (2016) Protein Engineering by Combined Computational and In Vitro Evolution Approaches. *Trends Biochem. Sci.*, **41**, 421–433.

10. Packer, M.S. and Liu, D.R. (2015) Methods for the directed evolution of proteins. *Nat. Rev. Genet.*, **16**, 379–394.

11. Arnold ,F.H. (2018) Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie Int. Ed.*, **57**, 4143–4148.

12. Takahashi, T.T., Austin, R.J. and Roberts, R.W. (2003) mRNA display: Ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.*, **28**, 159–165.

13. Hammer, S.C., Knight, A.M. and Arnold, F.H. (2017) Design and evolution of enzymes for non-natural chemistry. *Curr. Opin. Green Sustain. Chem.*, **7**, 23–30.

14. Keefe, A.D. and Szostak, J.W. (2001) Functional proteins from a random-sequence library. *Nature*, **410**, 715–718.

15. Chaput,J .C. and Szostak, J.W. (2004) Evolutionary Optimization of a Nonbiological ATP Binding Protein for Improved Folding Stability. *Chem. Biol.*, **11**, 865–874.

16. Seelig, B. (2011) MRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat. Protoc.*, **6**, 540–552.

17. Dydio, P., Key, H.M., Nazarenko, A., Rha, J.Y.E., Seyedkazemi, V., Clark, D.S. and Hartwig, J.F. (2016) An artificial metalloenzyme with the kinetics of native enzymes. *Science,* **354**, 102–106.

18. Kaufmann, K.W., Lemmon, G.H., Deluca, S.L., Sheehan, J.H. and Meiler, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry*, **49**, 2987–98.

19. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

20. Hatley, M.E., Lockless, S.W., Gibson, S.K., Gilman, A.G. and Ranganathan, R. (2003) Allosteric determinants in guanine nucleotide-binding proteins. *Proc. Natl. Acad. Sci.*, **100**, 14445–14450.

21. Haynes, K.A. and Silver, P.A. (2011) Synthetic Reversal of Epigenetic Silencing. *J. Biol. Chem.*, **286**, 27176–27182.

22. Margueron, R. and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343–9.

23. Santanach, A., Blanco, E., Jiang, H., Molloy, K.R., Sansó, M., LaCava, J., Morey, L. and Di Croce, L. (2017) The Polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity. *Nat. Commun.*, **8**, 1–11.

24. Gibert, P., Huey, R.B. and Gilchrist, G.W. (2001) Locomotor Performance of Drosophila melanogaster : Interactions among Developmental and Adult Temperatures, Age, and Geography. Evolution, **55**, 205–209.

25. Scott, G.R. and Johnston, I.A. (2012) Temperature during embryonic development has persistent effects on thermal acclimation capacity in zebrafish. *Proc. Natl. Acad. Sci.*, **109**, 14247–14252.

REFERENCES

1. Akishiba, M., Takeuchi,T., Kawaguchi,Y., Sakamoto,K., Yu,H.-H., Nakase,I., Takatani-Nakase,T., Madani,F., Gräslund,A. and Futaki,S. (2017) Cytosolic antibody delivery by lipid-sensitive endosomolytic peptide. *Nat. Chem.*, **9**, 751–761.

2. Amabile,A., Migliara,A., Capasso,P., Biffi,M., Cittaro,D., Naldini,L. and Lombardo,A. (2016) Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell*, **167**, 219–232.

3. Andrews,F.H., Strahl,B.D. and Kutateladze,T.G. (2016) Insights into newly discovered marks and readers of epigenetic information. *Nat. Chem. Biol.*, **12**, 662–668.

4. Arnold ,F.H. (2018) Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie Int. Ed.*, **57**, 4143–4148.

5. 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.

6. Baker, T., Nerle, S., Pritchard, J., Zhao, B., Rivera, V. M., Garner, A., and Gonzalvez, F. (2015) Acquisition of a single EZH2 D1 domain mutation confers acquired resistance to EZH2-targeted inhibitors. Oncotarget 6, 32646−32655.

7. Bannister,A.J. and Kouzarides,T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.

8. Bannister,A.J., Zegerman,P., Partridge,J.F., Miska,E.a, Thomas,J.O., Allshire,R.C. and Kouzarides,T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, **410**, 120–124.

9. Bauer,M., Trupke,J. and Ringrose,L. (2016) The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*, **125**, 471–496.

10. Berman,H.M.,Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H. and Shindyalov,I.N. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

11. Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., and Allis, C. D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. 26, 2560−2569.

12. Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., and Allis, C. D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. 26, 2560−2569.

13. Bernstein,E., Duncan,E.M., Masui,O., Gil,J., Heard,E. and Allis,C.D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.*, **26**, 2560–2569.

14. Berry,D. (2003) Molecular Visualizations of DNA. http://www.wehi.edu.au/wehi-tv/molecular-visualisations-dna.

15. Bintu, L., Yong, J., Antebi, Y. E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., and Elowitz, M. B. (2016) Dynamics of Epigenetic Regulation at the Single-Cell Level. Science 351 (6274), 720−724.

16. Black,J.C., Van Rechem,C. and Whetstine,J.R. (2012) Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell*, **48**, 491–507.

17. Bramswig,N.C., Everett,L.J., Schug,J., Dorrell,C., Liu,C., Luo,Y., Streeter,P.R., Naji,A., Grompe,M. and Kaestner,K.H. (2013) Epigenomic plasticity enables human pancreatic _ to _ cell reprogramming. *J. Clin. Invest.*, **123**, 1275–1284.

18. Bruce,V.J. and McNaughton,B.R. (2017) Inside Job: Methods for Delivering Proteins to the Interior of Mammalian Cells. *Cell Chem. Biol.*, **24**, 924–934.

19. Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., et al. (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin. Chem. 55 (4), 611−622.

20. Campos,E.I. and Reinberg,D. (2009) Histones: annotating chromatin. *Annu. Rev. Genet.*, **43**, 559–599.

21. Cano-Rodriguez, D., and Rots, M. G. (2016) Epigenetic Editing: On the Verge of Reprogramming Gene Expression at Will. Curr. Genet. Med. Rep. 4, 170−179.

22. Caschera, F., and Noireaux, V. (2014) Synthesis of 2.3 mg/mL of protein with an all Escherichia coli cell-free transcription−translation system. Biochimie 99, 162−168.

23. Cavalli,G. and Paro,R. (1998) The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell*, **93**, 505–518.

24. Cavalli,G. and Paro,R. (1999) Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science*, **286**, 955–958.

25. Chaput,J .C. and Szostak, J.W. (2004) Evolutionary Optimization of a Nonbiological ATP Binding Protein for Improved Folding Stability. *Chem. Biol.*, **11**, 865–874.

26. Chen, X., Zaro, J. L., and Shen, W.-C. (2013) Fusion protein linkers: property, design and functionality. Adv. Drug Delivery Rev. 65, 1357−1369.

27. Cheung,K., Lu,G., Sharma,R., Vincek,A., Zhang,R., Plotnikov,A.N., Zhang,F., Zhang,Q., Ju,Y., Hu,Y. *et al.* (2017) BET N-terminal bromodomain inhibition selectively blocks Th17 cell differentiation and ameliorates colitis in mice. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 2952–2957.

28. Chevalier,A., Silva,D.-A., Rocklin,G.J., Hicks,D.R., Vergara,R., Murapa,P., Bernard,S.M., Zhang,L., Lam,K.-H., Yao,G., *et al.* (2017) Massively parallel de novo protein design for targeted therapeutics. *Nature*, **550**, 74–79.

29. Cuddapah,S., Roh,T.-Y., Cui,K., Jose,C.C., Fuller,M.T., Zhao,K. and Chen,X. (2012) A novel human Polycomb binding site acts as a functional Polycomb response element in Drosophila. *PLoS One*, **7**, e36365.

30. Daer, R. M., Cutts, J. P., Brafman, D. A., and Haynes, K. A. (2017) The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. ACS Synth. Biol. 6, 428−438.

31. David,Y., Vila-Perell ’o,M., Verma,S. and Muir,T.W. (2015) Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nat. Chem.*, **7**, 394–402.

32. Dejardin,J. and Cavalli,G. (2004) Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *EMBO J.*, **23**, 857–868.

33. Dekker,F.J., van den Bosch,T. and Martin,N.I. (2014) Small molecule inhibitors of histone acetyltransferases and deacetylases are potential drugs for inflammatory diseases. *Drug Discov. Today*, **19**, 654–660.

34. Delachat, A. M.-F., Guidotti, N., Bachmann, A. L., Meireles- Filho, A. C. A., Pick, H., Lechner, C. C., Deluz, C., Deplancke, B., Suter, D. M., and Fierz, B.

(2018) Engineered Multivalent Sensors to Detect Coexisting Histone Modifications in Living Stem Cells. Cell Chem. Biol. 25, 51.

35. Delmore,J.E., Issa,G.C., Lemieux,M.E., Rahl,P.B., Shi,J., Jacobs,H.M., Kastritis,E., Gilpatrick,T., Paranal,R.M., Qi,J. *et al.* (2011) BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*, **146**, 904–917.

36. 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.

37. Du,S., Liew,S.S., Li,L. and Yao,S.Q. (2018) Bypassing Endocytosis: Direct Cytosolic Delivery of Proteins. *J. Am. Chem. Soc.*, **140**, 15986–15996

38. Dydio, P., Key, H.M., Nazarenko, A., Rha, J.Y.E., Seyedkazemi, V., Clark, D.S. and Hartwig, J.F. (2016) An artificial metalloenzyme with the kinetics of native enzymes. *Science,* **354**, 102–106.

39. Eissenberg,J.C. (2012) Structural biology of the chromodomain: Form and function. *Gene*, **496**, 69–78.

40. Elasser,S.J., Ernst,R.J., Walker,O.S. and Chin,J.W. (2016) Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat. Methods*, **13**, 158–164.

41. Engler, C., and Marillonnet, S. (2011) Generation of Families of Construct Variants Using Golden Gate Shuffling. Methods Mol. Biol. 729, 167−181.

42. Engreitz,J.M., Ollikainen,N. and Guttman,M. (2016) Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.*, **17**, 756–770.

43. Essafi, M., Baudot,A.D., Mouska,X., Cassuto,J.-P., Ticchioni,M. and Deckert,M. (2011) Cell-Penetrating TAT-FOXO3 Fusion Proteins Induce Apoptotic Cell Death in Leukemic Cells. *Mol. Cancer Ther.*, **10**, 37–46.

44. Falkenberg,K.J. and Johnstone,R.W. (2014) Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat. Rev. Drug Discov.*, **13**, 673–691.

45. Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempl, P. M., Schmidl, C., et al. (2016) Comprehensive Genome and Epigenome Characterization of CHO Cells in Response to Evolutionary Pressures and over Time. Biotechnol. Bioeng. 113 (10), 2241−2253.

46. 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., and Knapp, S. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149, 214−231.

47. Filippakopoulos, P. and Knapp, S. (2012) The bromodomain interaction module. *FEBS Lett.*, **586**, 2692–2704.

48. Filippakopoulos,P., Qi,J., Picaud,S., Shen,Y., Smith,W.B., Fedorov,O., Morse,E.M., Keates,T., Hickman,T.T., Felletar,I. *et al.* (2010) Selective inhibition of BET bromodomains. *Nature*, **468**, 1067–1073.

49. Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003) Molecular Basis for the Discrimination of Repressive Methyl-Lysine Marks in Histone H3 by Polycomb and HP1 Chromodomains. Genes Dev. 17 (15), 1870−1881.

50. Flynn, E. M., Huang, O. W., Poy, F., Oppikofer, M., Bellon, S. F., Tang, Y., and Cochran, A. G. (2015) A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications. Structure 23 (10), 1801−1814.

51. Fujisawa,T. and Filippakopoulos,P. (2017) Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.*, **18**, 246–262.

52. Futterer,a, Raya,A., Llorente,M., Izpisu'a-Belmonte,J.C., de la Pompa,J.L., Klatt,P. and Mart'ınez-A,C. (2012) Ablation of Dido3 compromises lineage commitment of stem cells in vitro and during early embryonic development. *Cell Death Differ.*, **19**, 132–143.

53. Gao,Z., Zhang,J., Bonasio,R., Strino,F., Sawai,A., Parisi,F., Kluger,Y. and Reinberg,D. (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell*, **45**, 344–356.

54. Gatchalian, J., Fütterer, A., Rothbart, S. B., Tong, Q., Rincon- Arano, H., Sánchez de Diego, A., Groudine, M., Strahl, B. D., Martínez-A, C., van Wely, K. H. M., and Kutateladze, T. G. (2013) Dido3 PHD modulates cell differentiation and division. Cell Rep. 4, 148−158.

55. Gelato,K.A., Tauber,M., Ong,M.S., Winter,S., Hiragami-Hamada,K., Sindlinger,J., Lemak,A., Bultsma,Y., Houliston,S., Schwarzer,D. *et al.* (2014) Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. *Mol. Cell*, **54**, 905–919.

56. Gibaja, V., Shen, F., Harari, J., Korn, J., Ruddy, D., Saenz-Vash, V., Zhai, H., Rejtar, T., Paris, C. G., Yu, Z., Lira, M., King, D., Qi, W., Keen, N., Hassan, A. Q., and Chan, H. M. (2016) Development of secondary mutations in wild-type and mutant EZH2 alleles cooperates to confer resistance to EZH2 inhibitors. Oncogene 35, 558−566.

57. Gibert, P., Huey, R.B. and Gilchrist, G.W. (2001) Locomotor Performance of Drosophila melanogaster : Interactions among Developmental and Adult Temperatures, Age, and Geography. Evolution, **55**, 205–209.

58. Guidotti,G., Brambilla,L. and Rossi,D. (2017) Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends Pharmacol. Sci.*, **38**, 406–424.

59. Greisman, H. A., and Pabo, C. O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science 275, 657−661.

60. Hacisuleyman,E., Shukla,C.J., Weiner,C.L. and Rinn,J.L. (2016) Function and evolution of local repeats in the Firre locus. *Nat. Commun.*, **7**, 11021.

61. Hammer, S.C., Knight, A.M. and Arnold, F.H. (2017) Design and evolution of enzymes for non-natural chemistry. *Curr. Opin. Green Sustain. Chem.*, **7**, 23–30.

62. Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008) A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. 10, 1291−1300.

63. Hasegawa,Y., Brockdorff,N., Kawano,S., Tsutui,K., Tsutui,K. and Nakagawa,S. (2010) The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell*, **19**, 469–476.

64. Hatley, M.E., Lockless, S.W., Gibson, S.K., Gilman, A.G. and Ranganathan, R. (2003) Allosteric determinants in guanine nucleotide-binding proteins. *Proc. Natl. Acad. Sci.*, **100**, 14445–14450.

65. Hattori, N., Niwa, T., Kimura, K., Helin, K., and Ushijima, T. (2013) Visualization of Multivalent Histone Modification in a Single Cell Reveals Highly Concerted Epigenetic Changes on Differentiation of Embryonic Stem Cells. Nucleic Acids Res. 41 (15), 7231−7239.

66. Hauenschild,A., Ringrose,L., Altmutter,C., Paro,R. and Rehmsmeier,M. (2008) Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. *PLoS Biol.*, **6**, 2130–2143. 148. Gilman,J. and Love,J. (2016) Synthetic promoter design for new microbial chassis. *Biochem. Soc. Trans.*, **44**, 731–737.

67. Haumaitre,C., Lenoir,O. and Scharfmann,R. (2008) Histone deacetylase inhibitors modify pancreatic cell fate determination and amplify endocrine progenitors. *Mol. Cell. Biol.*, **28**, 6373–6383.

68. Haynes, K.A. and Silver, P.A. (2011) Synthetic Reversal of Epigenetic Silencing. *J. Biol. Chem.*, **286**, 27176–27182.

69. Haynes,K.A. and Silver,P.A. (2009) Eukaryotic systems broaden the scope of synthetic biology. *J. Cell Biol.*, **187**, 589–596.

70. Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I. A., Zenn, H. M., Kost, N., Pohl, W., Chernev, A., Schwarzer, D., Jenuwein, T., Lorincz, M., Zimmermann, B., Walla, P. J., Neumann, H., Baubec, T., Urlaub, H., and Fischle, W. (2016) Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. Nat. Commun. 7, 11310.

71. He,S., Bauman,D., Davis,J.S., Loyola,A., Nishioka,K., Gronlund,J.L., Reinberg,D., Meng,F., Kelleher,N. and McCafferty,D.G. (2003) Facile synthesis of site-specifically acetylated and methylated histone proteins: Reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12033–12038.

72. Hendrickson,D.G., Kelley,D.R., Tenen,D., Bernstein,B. and Rinn,J.L. (2016) Widespread RNA binding by chromatin-associated proteins. *Genome Biol.*, **17**, 28.

73. Hilton,I.B., D'Ippolito,A.M., Vockley,C.M., Thakore,P.I., Crawford,G.E., Reddy,T.E. and Gersbach,C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.*, **33**, 510–517.

74. Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I. A., Zenn, H. M., Kost, N., Pohl, W., et al. (2016) Dynamic and Flexible H3K9me3 Bridging via HP1β Dimerization Establishes a Plastic State of Condensed Chromatin. Nat. Commun. 7, 11310.

75. Hiragami-Hamada,K., Soeroes,S., Nikolov,M., Wilkins,B., Kreuz,S., Chen,C., De La Rosa-Vel'azquez,I.A., Zenn,H.M., Kost,N., Pohl,W. *et al.* (2016) Dynamic and flexible H3K9me3 bridging via HP1_ dimerization establishes a plastic state of condensed chromatin. *Nat. Commun.*, **7**, 11310.

76. Horton,J.R., Liu,X., Gale,M., Wu,L., Shanks,J.R., Zhang,X., Webber,P.J., Bell,J.S.K., Kales,S.C., Mott,B.T. *et al.* (2016) Structural basis for KDM5A

histone lysine demethylase inhibition by diverse compounds. *Cell Chem. Biol.*, **23**, 769–781.

77. Houbaviy,H.B., Usheva,A., Shenk,T. and Burley,S.K. (1996) Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 13577–13582.

78. Huang,H., Lin,S., Garcia,B.A. and Zhao,Y. (2015) Quantitative proteomic analysis of histone modifications. *Chem. Rev.*, **115**, 2376–2418.

79. Huang,H., Zhang,J., Shen,W., Wang,X., Wu,J., Wu,J. and Shi,Y. (2007) Solution structure of the second bromodomain of Brd2 and its specific interaction with acetylated histone tails. *BMC Struct. Biol.*, **7**, 57.

80. Huh,J., Wu,J., Lee,C., Yun,M., Gilada,D., Brautigam,C.A. and Li,B. (2012) Multivalent di-nucleosome recognition enables the Rpd3S histone deacetylase complex to tolerate decreased H3K36 methylation levels. *EMBO J.*, **31**, 3564–3574.

81. Jacobs,S.A. (2002) Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science*, **295**, 2080–2083.

82. Jacobson,R.H., Ladurner,A.G., King,D.S. and Tjian,R. (2000) Structure and function of a human TAF II 250 double bromodomain module. *Science*, **288**, 1422–1425.

83. James,L.I., Barsyte-Lovejoy,D., Zhong,N., Krichevsky,L., Korboukh,V.K., Herold,J.M., MacNevin,C.J., Norris,J.L., Sagum,C.A., Tempel,W. *et al.* (2013) Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat. Chem. Biol.*, **9**, 184–191.

84. Jantz, D., and Berg, J. M. (2010) Probing the DNA-binding affinity and specificity of designed zinc finger proteins. Biophys. J. 98, 852−860.

85. Jeyaprakash,A.A., Basquin,C., Jayachandran,U. and Conti,E. (2011) Structural basis for the recognition of phosphorylated histone H3 by the surviving subunit of the chromosomal passenger complex. *Structure*, **19**, 1625–1634.

86. Jones, P. A., Issa, J.-P. J., and Baylin, S. (2016) Targeting the Cancer Epigenome for Therapy. Nat. Rev. Genet. 17 (10), 630−641.

87. Jung, M., Philpott, M., Müller, S., Schulze, J., Badock, V., Eberspächer, U., Moosmayer, D., Bader, B., Schmees, N., Fernández- Montalván, A., et al. (2014) Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the

Histone H4 Tail and the Small Molecule Inhibitor JQ1. J. Biol. Chem. 289 (13), 9304−9319.

88. Kallappagoudar,S., Dammer,E.B., Duong,D.M., Seyfried,N.T. and Lucchesi,J.C. (2013) Expression, purification and proteomic analysis of recombinant histone H4 acetylated at lysine 16. *Proteomics*, **13**, 1687–1691.

89. Kamakaka,R.T. (2005) Histone variants: deviants? *Genes Dev.*, **19**, 295–316.

90. Kassis,J.A. and Brown,J.L. (2013) Polycomb group response elements in Drosophila and vertebrates. In: *Advances in Genetics*. Elsevier Inc., Vol. **81**, pp. 83–118.

91. Kato,T., Tanaka,M. and Oba,M. (2013) Protein transfection study using multicellular tumor spheroids of human hepatoma Huh-7 cells. *PLoS One*, **8**, 1–8.

92. Kaufmann, K.W., Lemmon, G.H., Deluca, S.L., Sheehan, J.H. and Meiler, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry*, **49**, 2987–98.

93. Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G.A., Li, Z., Vedadi, M., Schapira, M. *et al.* (2011) Recognition and specificity determinants of the human Cbx chromodomains. *J. Biol. Chem.*, **286**, 521–529.

94. Kearns,N.a, Pham,H., Tabak,B., Genga,R.M., Silverstein,N.J., Garber,M. and Maehr,R. (2015) Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat. Methods*, **12**, 401–403.

95. Keefe, A.D. and Szostak, J.W. (2001) Functional proteins from a random-sequence library. *Nature*, **410**, 715–718.

96. Kelly,A.E., Ghenoiu,C., Xue,J.Z., Zierhut,C., Kimura,H. and Funabiki,H. (2010) Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase aurora B. *Science*, **330**, 235–239.

97. Kelly,T.K., De Carvalho,D.D. and Jones,P.A. (2010) Epigenetic modifications as therapeutic targets. *Nat. Biotechnol.*, **28**, 1069–1078.

98. Keung,A.J., Bashor,C.J., Kiriakov,S., Collins,J.J. and Khalil,A.S. (2014) Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell*, **158**, 110–120.

99. Khan,S. and Jena,G. (2014) Sodium butyrate, a HDAC inhibitor ameliorates eNOS, iNOS and TGF-_1-induced fibrogenesis, apoptosis and DNA damage in the kidney of juvenile diabetic rats. *Food Chem. Toxicol.*, **73**, 127–139.

100. Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N.C., Joung, J. K., and Collins, J. J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. Cell 150, 647−658.

101. Khan,S. and Jena,G.B. (2014) Protective role of sodium butyrate, a HDAC inhibitor on beta-cell proliferation, function and glucose homeostasis through modulation of p38/ERK MAPK and apoptotic pathways: Study in juvenile diabetic rat. *Chem. Biol. Interact.*, **213**, 1–12.

102. Khuong, M. T., Fei, J., Cruz-Becerra, G., and Kadonaga, J. T. (2017) A Simple and Versatile System for the ATP-Dependent Assembly of Chromatin. J. Biol. Chem. 292 (47), 19478−19490.

103. Kim,C.H., Kang,M., Kim,H.J., Chatterjee,A. and Schultz,P.G. (2012) Site-specific incorporation of _-N-crotonyllysine into histones. *Angew. Chem. Int. Ed.*, **51**, 7246–7249.

104. Kim,H. and Bae,S. (2011) Histone deacetylase inhibitors: molecular mechanisms of action and clinical trials as anti-cancer drugs. *Am. J. Transl. Res.*, **3**, 166–179.

105. Kim,S., Natesan,S., Cornilescu,G., Carlson,S., Tonelli,M., McClurg,U.L., Binda,O., Robson,C.N., Markley,J.L., Balaz,S. *et al.* (2016) Mechanism of histone H3K4me3 recognition by the plant homeodomain of inhibitor of growth 3. *J. Biol. Chem.*, **291**, 18326–18341.

106. Kimura,H. and Cook,P.R. (2001) Kinetics of core histones in living human cells. *J. Cell Biol.*, **153**, 1341–1354.

107. Kintzing,J.R., Filsinger Interrante,M. V. and Cochran,J.R. (2016) Emerging Strategies for Developing Next-Generation Protein Therapeutics for Cancer Treatment. *Trends Pharmacol. Sci.*, **37**, 993–1008.

108. Konermann,S., Brigham,M.D., Trevino,A., Hsu,P.D., Heidenreich,M., Le,Cong, Platt,R.J., Scott,D.a, Church,G.M. and Zhang,F. (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, **500**, 472–476.

109. Konze,K.D., Ma,A., Li,F., Barsyte-Lovejoy,D., Parton,T., MacNevin,C.J., Liu,F., Gao,C., Huang,X.-P., Kuznetsova,E. *et al.* (2013) An orally bioavailable

chemical probe of the lysine methyltransferases EZH2 and EZH1. *ACS Chem. Biol.*, **8**, 1324–1334.

110. Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.

111. Kowarz, E., Löscher, D., and Marschalek, R. (2015) Optimized Sleeping Beauty Transposons Rapidly Generate Stable Transgenic Cell Lines. Biotechnol. J. 10 (4), 647−653.

112. Krosl,J., Austin,P., Beslu,N., Kroon,E., Humphries,R.K. and Sauvageau,G. (2003) In vitro expansion of hematopoietic stem cells by recombinant TAT-HOXB4 protein. *Nat. Med.*, **9**, 1428–1432.

113. Kungulovski,G., Nunna,S., Thomas,M., Zanger,U.M., Reinhardt,R. and Jeltsch,A. (2015) Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenet. Chromatin*, **8**, 12.

114. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 16, 2893−2905.

115. Kwaks,T.H.J., Sewalt,R.G.A.B., van Blokland,R., Siersma,T.J., Kasiem,M., Kelder,A. and Otte,A.P. (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *J. Biotechnol.*, **115**, 35 46.

116. Kwon,Y. Do, Oh,S.K., Kim,H.S., Ku,S.Y., Kim,S.H., Choi,Y.M. and Moon,S.Y. (2005) Cellular manipulation of human embryonic stem cells by TAT-PDX1 protein transduction. *Mol. Ther.*, **12**, 28–32.

117. Lasry, A. & Ben-Neriah, Y. Senescence-associated inflammatory responses: Aging and cancer perspectives. *Trends Immunol.* **36,** 217–228 (2015).

118. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7,** 21–39 (2008).

119. Lee, C.-H., Wu, J., and Li, B. (2013) Chromatin Remodelers Fine-Tune H3K36me-Directed Deacetylation of Neighbor Nucleosomes by Rpd3S. Mol. Cell 52 (2), 255−263.

120. Lee, Y.J., Erazo-Oliveras,A. and Pellois,J.P. (2010) Delivery of macromolecules into live cells by simple co-incubation with a peptide. *ChemBioChem*, **11**, 325–330.

121. Lever,M.a, Th'ng,J.P., Sun,X. and Hendzel,M.J. (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, **408**, 873–876.

122. Li,B., Gogol,M., Carey,M., Lee,D., Seidel,C. and Workman,J.L. (2007) Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science*, **316**, 1050–1054.

123. Li,H., Fischle,W., Wang,W., Duncan,E.M., Liang,L., Murakami-Ishibe,S., Allis,C.D. and Patel,D.J. (2007) Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell*, **28**, 677–691.

124. Li,H., Ilin,S., Wang,W., 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.

125. Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S., and Jacobsen, S. E. (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J. 23, 4146−4155.

126. Lönn,P., Kacsinta,A.D., Cui,X.S., Hamil,A.S., Kaulich,M., Gogoi,K. and Dowdy,S.F. (2016) Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.*, **6**, 32301.

127. Lungu, C., Pinter, S., Broche, J., Rathert, P., and Jeltsch, A. (2017) Modular Fluorescence Complementation Sensors for Live Cell Detection of Epigenetic Signals at Endogenous Genomic Sites. Nat. Commun. 8 (1), 649.

128. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal Structure of the Nucleosome Core Particle at 2.8 A Resolution. Nature 389 (6648), 251−260.

129. Lupo,A., Cesaro,E., Montano,G., Zurlo,D., Izzo,P. and Costanzo,P. (2013) KRAB-zinc finger proteins: a repressor family displaying multiple biological functions. *Curr. Genomics*, **14**, 268–278.

130. Lyubitelev,A. V, Nikitin,D. V, Shaytan,A.K., Studitsky,V.M. and Kirpichnikov,M.P. (2016) Structure and functions of linker histones. *Biochem.*, **81**, 213–223.

131. Ma,A.-N., Wang,H., Guo,R., Wang,Y.-X., Li,W., Cui,J., Wang,G., Hoffman,A.R. and Hu,J.-F. (2014) Targeted gene suppression by inducing de

novo DNA methylation in the gene promoter. *Epigenetics and Chromatin*, **7**, 20.

132. Margueron, R. and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343–9.

133. McDaniel,S.L., Fligor,J.E., Ruan,C., Cui,H., Bridgers,J.B., DiFiore,J.V., Guo,A.H., Li,B. and Strahl,B.D. (2016) Combinatorial histone readout by the dual plant homeodomain (PHD) fingers of Rco1 mediates Rpd3S chromatin recruitment and the maintenance of transcriptional fidelity. *J. Biol. Chem.*, **291**, 14796–14802.

134. McGinty,R.K., Kim,J., Chatterjee,C., Roeder,R.G. and Muir,T.W. (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature*, **453**, 812–816.

135. Meckler, J. F., Bhakta, M. S., Kim, M.-S., Ovadia, R., Habrian, C. H., Zykovich, A., Yu, A., Lockwood, S. H., Morbitzer, R., Elsäesser, J., Lahaye, T., Segal, D. J., and Baldwin, E. P. (2013) Quantitative analysis of TALE−DNA interactions suggests polarity effects. Nucleic Acids Res. 41, 4118−4128.

136. Mendenhall,E.M.,Williamson,K.E., Reyon,D., Zou,J.Y., Ram,O., Joung,J.K. and Bernstein,B.E. (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.*, **31**, 1133–1136.

137. Mertz,J.A., Conery,A.R., Bryant,B.M., Sandy,P., Balasubramanian,S.,Mele,D.A., Bergeron,L. and Sims,R.J. (2011) Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16669–16674.

138. Merutka, G., Shalongo, W., and Stellwagen, E. (1991) A model peptide with enhanced helicity. Biochemistry 30, 4245−4248.

139. Messmer, S., Franke, A., and Paro, R. (1992) Analysis of the functional role of the Polycomb chromo domain in Drosophila melanogaster. Genes Dev. 6, 1241−1254.

140. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the panther classification system. *Nat. Protoc.* **8,** 1551–1566 (2013).

141. Michiue,H., Tomizawa,K., Wei,F.Y., Matsushita,M., Lu,Y.F., Ichikawa,T., Tamiya,T., Date,I. and Matsui,H. (2005) The NH2 terminus of influenza virus hemagglutinin-2 subunit peptides enhances the antitumor potency of

polyarginine-mediated p53 protein transduction. *J. Biol. Chem.*, **280**, 8285–8289.

142. Milosevich,N., Gignac,M.C., McFarlane,J., Simhadri,C., Horvath,S., Daze,K.D., Croft,C.S., Dheri,A., Quon,T.T.H., Douglas,S.F. *et al.* (2016) Selective inhibition of CBX6: a methyllysine reader protein in the polycomb family. *ACS Med. Chem. Lett.*, **7**, 139–144.

143. Min,J. (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.*, **17**, 1823–1828.

144. Mlambo,T., Nitsch,S., Hildenbeutel,M., Romito,M., Müller,M., Bossen,C., Diederichs,S., Cornu,T.I., Cathomen,T. and Mussolino,C. (2018) Designer epigenome modifiers enable robust and sustained gene silencing in clinically relevant human cells. *Nucleic Acids Res.*, **46**, 4456–4468.

145. Muller,J. and Kassis,J.A. (2006) Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.*, **16**, 476–484.

146. Musselman,C.A. and Kutateladze,T.G. (2011) Handpicking epigenetic marks with PHD fingers. Nucleic Acids Res., 39, 9061–9071.

147. Musselman,C.A., Lalonde,M.-E., Cˆot'e,J. and Kutateladze,T.G. (2012) Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.*, **19**, 1218–1227.

148. 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**, 153–163.

149. Nicodeme,E., Jeffrey,K.L., Schaefer,U., Beinke,S., Dewell,S., Chung,C., Chandwani,R., Marazzi,I., Wilson,P., Coste,H. *et al.* (2010) Suppression of inflammation by a synthetic histone mimic. *Nature*, **468**, 1119–1123.

150. Niedzialkowska,E., Wang,F., Porebski,P.J., Minor,W., Higgins,J.M.G. and Stukenberg,P.T. (2012) Molecular basis for phosphospecific recognition of histone H3 tails by Survivin paralogues at inner centromeres. *Mol. Biol. Cell*, **23**, 1457–1466.

151. Nielsen,P.R., Nietlispach,D.,Mott,H.R., Callaghan,J., Bannister,A., Kouzarides,T., Murzin,A.G., Murzina,N.V. and Laue,E.D. (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature*, **416**, 103–107.

152. Norwood,L.E.,Moss,T.J., Margaryan,N. V., Cook,S.L., Wright,L., Seftor,E.A., Hendrix,M.J.C., Kirschmann,D.A. and Wallrath,L.L. (2006) A requirement for dimerization of HP1Hs in suppression of breast cancer invasion. *J. Biol. Chem.*, **281**, 18668–18676.

153. Nyer, D. B., Daer, R. M., Vargas, D., Hom, C., and Haynes, K. A. (2017) Regulation of Cancer Epigenomes with a Histone-Binding Synthetic Transcription Factor. npj Genomic Med. 2, 1.

154. Oike,T., Ogiwara,H., Amornwichet,N., Nakano,T. and Kohno,T. (2014) Chromatin-regulating proteins as targets for cancer therapy. *J. Radiat. Res.*, **55**, 613–628.

155. Okamoto,Y., Kojima,R., Schwizer,F., Bartolami,E., Heinisch,T., Matile,S., Fussenegger,M. and Ward,T.R. (2018) A cell-penetrating artificial metalloenzyme regulates a gene switch in a designer mammalian cell. *Nat. Commun.*, **9**, 1–7.

156. Oliver, S. S., Musselman, C. A., Srinivasan, R., Svaren, J. P., Kutateladze, T. G., and Denu, J. M. (2012) Multivalent recognition of histone tails by the PHD fingers of CHD5. Biochemistry 51, 6534−6544.

157. Oliveros, J. C. Venny. an interactive tool for comparing lists with Venn's diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html (2007–2015).

158. Owen,D.J. (2000) The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J.*, **19**, 6141–6149.

159. Packer, M.S. and Liu, D.R. (2015) Methods for the directed evolution of proteins. *Nat. Rev. Genet.*, **16**, 379–394.

160. Park,M., Keung,A.J. and Khalil,A.S. (2016) The epigenome: the next substrate for engineering. *Genome Biol.*, **17**, 183.

161. PDB ID: 4XBF, Luka,Z., Loukachevitch,L.V., Martin,W.J., Wagner,C. and Reiter,N.J. (2016) Structure of LSD1:CoREST in complex with ssRNA

162. Pena,P. V., Davrazou,F., Shi,X., Walter,K.L., Verkhusha,V. V, Gozani,O., Zhao,R. and Kutateladze,T.G. (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature*, **442**, 100–103.

163. Pengue,G. and Lania,L. (1996) Kr· ·uppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 1015–1020.

164. Pengue,G., Calabr'o,V., Bartoli,P.C., Pagliuca,A. and Lania,L. (1994) Repression of transcriptional activity at a distance by the evolutionary conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Res.*, **22**, 2908–2914.

165. Phillips, I., and Silver, P. (2006) A New Biobrick Assembly Strategy Designed for Facile Protein Engineering. http://dspace.mit. edu/handle/1721.1/32535.

166. Pick,H., Kilic,S. and Fierz,B. (2014) Engineering chromatin states: chemical and synthetic biology approaches to investigate histone modification function. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1839**, 644–656.

167. Platero,J.S., Hartnett,T. and Eissenberg,J.C. (1995) Functional analysis of the chromo domain of HP1. *EMBO J.*, **14**, 3977–3986.

168. Portoso,M., Ragazzini,R., Brencic, Z., Moiani,A., Michaud,A., Vassilev,I., Wassef,M., Servant,N., Sargueil,B. and Margueron,R. (2017) PRC2 is dispensable for HOTAIR -mediated transcriptional repression. *EMBO J.*, **36**, 981–994.

169. Qin,J.Y., Zhang,L., Clift,K.L., Hulur,I., Xiang,A.P., Ren,B.-Z. and Lahn,B.T. (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One*, **5**, e10611.

170. Quadri, S. M. S. (2015) Parafilm-M®, An Available Cost- Effective Alternative for Immuno-blot Pouches. Methods Mol. Biol. 1314, 313−323.

171. Rahman,S., Sowa,M.E., Ottinger,M., Smith,J.A., Shi,Y., Harper,J.W. and Howley,P.M. (2011) The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol. Cell. Biol.*, **31**, 2641–2652.

172. Rees,H.A., Komor,A.C., Yeh,W.H., Caetano-Lopes,J., Warman,M., Edge,A.S.B. and Liu,D.R. (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.*, **8**, 1–10.

173. Ren, C., Morohashi, K., Plotnikov, A. N., Jakoncic, J., Smith, S. G., Li, J., Zeng, L., Rodriguez, Y., Stojanoff, V., Walsh, M., and Zhou, M.-M. (2015) Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. Chem. Biol. 22, 161−168.

174. Ren,C., Smith,S.G., Yap,K., Li,S., Li,J., Mezei,M., Rodriguez,Y., Vincek,A., Aguilo,F., Walsh,M.J. *et al.* (2016) Structure-guided discovery of selective

antagonists for the chromodomain of polycomb repressive protein CBX7. *ACS Med. Chem. Lett.*, **7**, 601–605.

175. Ringrose,L., Rehmsmeier,M., Dura,J. and Paro,R. (2003) Genome-wide prediction of polycomb/trithorax response elements in Drosophila melanogaster. *Dev. Cell*, **5**, 759–771.

176. Rosca, E. V., Gillies, R. J., and Caplan, M. R. (2009) Glioblastoma targeting via integrins is concentration dependent. Biotechnol. Bioeng. 104, 408−417.

177. Rosenfeld, L., Heyne, M., Shifman, J.M. and Papo, N. (2016) Protein Engineering by Combined Computational and In Vitro Evolution Approaches. *Trends Biochem. Sci.*, **41**, 421–433.

178. Rossetto,D., Avvakumov,N. and Cˆot'e,J. (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, **7**, 1098–1108.

179. Rouet, R., Thuma, B.A., Roy, M.D., Lintner, N.G., Rubitski, D.M., Finley, J.E., Wisniewska, H.M., Mendonsa, R., Hirsh, A., De Oñate, L., *et al.* (2018) Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J. Am. Chem. Soc.*, **140**, 6596–6603.

180. 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**, 983–94.

181. 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. *et al.* (2011) Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell*, **145**, 692–706.

182. Sanchez, O. F., Mendonca, A., Carneiro, A. D., and Yuan, C. (2017) Engineering Recombinant Protein Sensors for Quantifying Histone Acetylation. ACS Sens 2 (3), 426−435.

183. Sanchez,R. and Zhou,M.-M. (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.*, **36**, 364–372.

184. Santanach, A., Blanco, E., Jiang, H., Molloy, K.R., Sansó, M., LaCava, J., Morey, L. and Di Croce, L. (2017) The Polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity. *Nat. Commun.*, **8**, 1–11.

185. Sarma,K. and Reinberg,D. (2005) Histone variants meet their match. *Nat. Rev. Mol. Cell Biol.*, **6**, 139–149.

186. Schmitges,F.W., Prusty,A.B., Faty,M., Stutzer,A., Lingaraju,G.M., Aiwazian,J., Sack,R., Hess,D., Li,L., Zhou,S. *et al.* (2011) Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell*, **42**, 330–341.

187. Schwaiger,M., Schonauer,A., Rendeiro,A.F., Pribitzer,C., Schauer,A., Gilles,A.F., Schinko,J.B., Renfer,E., Fredman,D. and Technau,U. (2014) Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.*, **24**, 639–650.

188. Scott, G.R. and Johnston, I.A. (2012) Temperature during embryonic development has persistent effects on thermal acclimation capacity in zebrafish. *Proc. Natl. Acad. Sci.*, **109**, 14247–14252.

189. Seelig, B. (2011) MRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat. Protoc.*, **6**, 540–552.

190. Senthilkumar, R., and Mishra, R. K. (2009) Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. BMC Genomics 10, 549.

191. Shen, Y., Chen, Y., Wu, J., Shaner, N. C., and Campbell, R. E. (2017) Engineering of mCherry variants with long Stokes shift, redshifted fluorescence, and low cytotoxicity. PLoS One 12, e0171257.

192. Shewmake, T. A., Solis, F. J., Gillies, R. J., and Caplan, M. R. (2008) Effects of Linker Length and Flexibility on Multivalent Targeting. Biomacromolecules 9, 3057−3064.

193. Shin,J. and Noireaux,V. (2012) An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.*, **1**, 29–41.

194. Shogren-Knaak,M. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, **311**, 844–847.

195. Shogren-Knaak,M.A., Fry,C.J. and Peterson,C.L. (2003) A native peptide ligation strategy for deciphering nucleosomal histone modifications. *J. Biol. Chem.*, **278**, 15744 15748.

196. Simhadri,C., Daze,K.D., Douglas,S.F., Quon,T.T.H., Dev,A., Gignac,M.C., Peng,F., Heller,M., Boulanger,M.J., Wulff,J.E. *et al.* (2014) Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J. Med. Chem.*, **57**, 2874–2883.

197. 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**, 1003–1012.

198. Sing,A., Pannell,D., Karaiskakis,A., Sturgeon,K., Djabali,M., Ellis,J., Lipshitz,H.D. and Cordes,S.P. (2009) A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell*, **138**, 885–897.

199. Smith,E. and Shilatifard,A. (2010) The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol. Cell*, **40**, 689–701.

200. Snowden,A.W., Gregory,P.D., Case,C.C. and Pabo,C.O. (2002) Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. *Curr. Biol.*, **12**, 2159–2166.

201. Solomon, M. J., Larsen, P. L., and Varshavsky, A. (1988) Mapping proteinDNA Interactions in Vivo with Formaldehyde: Evidence That Histone H4 Is Retained on a Highly Transcribed Gene. Cell 53 (6), 937−947.

202. Somarowthu,S., Legiewicz,M., Chill on,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell*, **58**, 353–361.

203. Staahl,B.T., Benekareddy,M., Coulon-Bainier,C., Banfal,A.A., Floor,S.N., Sabo,J.K., Urnes,C., Munares,G.A., Ghosh,A. and Doudna,J.A. (2017) Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.*, **35**, 431–434.

204. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62−67.

205. Strahl,B.D. and Allis,C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41–45.

206. Stuckey,J.I., Dickson,B.M., Cheng,N., Liu,Y., Norris,J.L., Cholensky,S.H., Tempel,W., Qin,S., Huber,K.G., Sagum,C. *et al.* (2016) A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nat. Chem. Biol.*, **12**, 180–187.

207. Su, X., Zhu, G., Ding, X., Lee, S. Y., Dou, Y., Zhu, B., Wu, W., and Li, H. (2014) Molecular Basis Underlying Histone H3 Lysine- Arginine Methylation Pattern Readout by Spin/Ssty Repeats of Spindlin1. Genes Dev. 28 (6), 622−636.

208. Su, Z., and Denu, J. M. (2016) Reading the Combinatorial Histone Language. ACS Chem. Biol. 11, 564−574.

209. Su, Z., Wang, F., Lee, J.-H., Stephens, K.E., Papazyan, R., Voronina, E., Krautkramer, K.A., Raman, A., Thorpe, J.J., Boersma, M.D., *et al.* (2016) Reader domain specificity and lysine demethylase-4 family function. *Nat. Commun.*, **7**, 13387.

210. Suganuma,T. and Workman,J.L. (2008) Crosstalk among histone modifications. *Cell*, **135**, 604–607.

211. Suganuma,T. and Workman,J.L. (2011) Signals and combinatorial functions of histone modifications. *Annu. Rev. Biochem.*, **80**, 473–499.

212. Swigut, T., and Wysocka, J. (2007) H3K27 demethylases, at long last. Cell 131, 29−32.

213. Tabet,S., Douglas,S.F., Daze,K.D., Garnett,G.A.E., Allen,K.J.H., Abrioux,E.M.M., Quon,T.T.H.,Wulff,J.E. and Hof,F. (2013) Synthetic trimethyllysine receptors that bind histone 3, trimethyllysine 27 (H3K27me3) and disrupt its interaction with the epigenetic reader protein CBX7. *Bioorg. Med. Chem.*, **21**, 7004–7010.

214. Taguchi,H., Xie,Y., Horikoshi,N., Maehara,K., Harada,A., Nogami,J., Sato,K., Arimura,Y., Osakabe,A., Kujirai,T. *et al.* (2017) Crystal structure and characterization of novel human histone H3 variants, H3.6, H3.7, and H3.8. *Biochemistry*, **56**, 2184–2196.

215. Takahashi, T.T., Austin, R.J. and Roberts, R.W. (2003) mRNA display: Ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.*, **28**, 159–165.

216. Talbert,P.B. and Henikoff,S. (2010) Histone variants — ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.*, **11**, 264–275.

217. Tan,M., Luo,H., Lee,S., Jin,F., Yang,J.S., Montellier,E., Buchou,T., Cheng,Z., Rousseaux,S., Rajagopal,N. *et al.* (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016–1028.

218. Taverna,S.D., Li,H., Ruthenburg,A.J., Allis,C.D. and Patel,D.J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.*, **14**, 1025–1040.

219. Tekel, S. J., and Haynes, K. A. (2017) Molecular Structures Guide the Engineering of Chromatin. Nucleic Acids Res. 45 (13), 7555−7570.

220. Tekel, S. J., Vargas, D. A., Song, L., LaBaer, J., Caplan, M. R., and Haynes, K. A. (2018) Tandem Histone-Binding Domains Enhance the Activity of a Synthetic Chromatin Effector. ACS Synth. Biol. 7 (3), 842−852.

221. Tekel, S. J., Barrett, C., Vargas, D. & Haynes, K. A. Design, Construction, and Validation of Histone-Binding Effectors in Vitro and in Cells. *Biochemistry* **57,** 4707–4716 (2018).

222. Thakore, P. I., Black, J. B., Hilton, I. B., and Gersbach, C. A. (2016) Editing the Epigenome: Technologies for Programmable Transcription and Epigenetic Modulation. Nat. Methods 13 (2), 127−137.

223. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

224. Trembecka-Lucas,D.O., Trembecka-Lucas,D.O., Szczurek,A.T., Szczurek,A.T., Dobrucki,J.W. and Dobrucki,J.W. (2013) Dynamics of the HP1_-PCNA-containing complexes in DNA replication and repair. *Nucleus*, **4**, 74–82.

225. Tsai,M.-C., Manor,O., Wan,Y., Mosammaparast,N., Wang,J.K., Lan,F., Shi,Y., Segal,E. and Chang,H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689–693.

226. Turner,B.M. (2005) Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat. Struct. Mol. Biol.*, **12**, 110–112.

227. Umehara,T., Nakamura,Y., Jang,M.K., Nakano,K., Tanaka,A., Ozato,K., Padmanabhan,B. and Yokoyama,S. (2010) Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J. Biol. Chem.*, **285**, 7610–7618.

228. van Ingen,H., van Schaik,F.M.A., Wienk,H., Ballering,J., Rehmann,H., Dechesne,A.C., Kruijzer,J.A.W., Liskamp,R.M.J., Timmers,H.T.M. and Boelens,R. (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure*, **16**, 1245–1256.

229. VanDemark,A.P., Kasten,M.M., Ferris,E., Heroux,A., Hill,C.P. and Cairns,B.R. (2007) Autoregulation of the Rsc4 tandem bromodomain by Gcn5 acetylation. *Mol. Cell*, **27**, 817–828.

230. Vasanthi,D., Nagabhushan,A.,Matharu,N.K. and Mishra,R.K. (2013) A functionally conserved Polycomb response element from mouse HoxD complex responds to heterochromatin factors. *Sci. Rep.*, **3**, 3011.

231. Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967−980.

232. Verma,S.K., Tian,X., LaFrance,L. V., Duquenne,C., Suarez,D.P., Newlander,K.A., Romeril,S.P., Burgess,J.L., Grant,S.W., Brackley,J.A. *et al.* (2012) Identification of potent, selective, cell-active inhibitors of the histone lysine methyltransferase EZH2. *ACS Med. Chem. Lett.*, **3**, 1091–1096.

233. Voigt, P., and Reinberg, D. (2011) Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. Chem- BioChem 12, 236−252.

234. Voigt, P., LeRoy, G., Drury, W. J., 3rd, Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A., and Reinberg, D. (2012) Asymmetrically modified nucleosomes. Cell 151, 181−193.

235. Wadia, J.S., Stan,R. V and Dowdy,S.F. (2004) Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.*, **10**, 310–315.

236. Wakamori, M., Fujii, Y., Suka, N., Shirouzu, M., Sakamoto, K., Umehara, T., and Yokoyama, S. (2015) Intra- and inter-nucleosomal interactions of the histone H4 tail revealed with a human nucleosome core particle with genetically-incorporated H4 tetra-acetylation. Sci. Rep. 5, 17204.

237. Wan, L., Wen, H., Li, Y., Lyu, J., Xi, Y., Hoshii, T., Joseph, J. K., Wang, X., Loh, Y.-H. E., Erb, M. A., et al. (2017) ENL Links Histone Acetylation to Oncogenic Gene Expression in Acute Myeloid Leukaemia. Nature 543 (7644), 265−269.

238. Wang, R., and You, J. (2015) Mechanistic analysis of the role of bromodomain-containing protein 4 (BRD4) in BRD4-NUT oncoprotein-induced transcriptional activation. J. Biol. Chem. 290, 2744−2758.

239. Wang,F., Dai,J., Daum,J.R., Niedzialkowska,E., Banerjee,B., Stukenberg,P.T., Gorbsky,G.J. and Higgins,J.M.G. (2010) Histone H3 Thr-3 phosphorylation by haspin positions aurora B at centromeres in mitosis. *Science*, **330**, 231–235.

240. Wang,G.G., Song,J., Wang,Z., Dormann,H.L., Casadio,F., Li,H., Luo,J., Patel,D.J. and Allis,C.D. (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature*, **459**, 847–851.

241. Wang,R. and You,J. (2015) Mechanistic analysis of the role of bromodomain-containing protein 4 (BRD4) in BRD4-NUT oncoprotein-induced transcriptional activation. *J. Biol. Chem.*, **290**, 2744–2758.

242. Wang,X., Wei,X., Pang,Q. and Yi,F. (2012) Histone deacetylases and their inhibitors: molecular mechanisms and therapeutic implications in diabetes mellitus. *Acta Pharm. Sin. B*, **2**, 387–395.

243. Wang,H.-X., Song,Z., Lao,Y.-H., Xu,X., Gong,J., Cheng,D., Chakraborty,S., Park,J.S., Li,M., Huang,D., *et al.* (2018) Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide. *Proc. Natl. Acad. Sci.*, **115**, 4903–4908.

244. West,A.C. and Johnstone,R.W. (2014) New and emerging HDAC inhibitors for cancer treatment. J Clin Invest., 124, 30–39.

245. Whitehead,L., Dobler,M.R., Radetich,B., Zhu,Y., Atadja,P.W., Claiborne,T., Grob,J.E., McRiner,A., Pancost,M.R., Patnaik,A. *et al.* (2011) Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. *Bioorg. Med. Chem.*, **19**, 4626–4634.

246. Wilkins,B.J., Hahn,L.E., Heitmuller,S., Frauendorf,H., Valerius,O., Braus,G.H. and Neumann,H. (2015) Genetically encoding lysine modifications on histone H4. *ACS Chem. Biol.*, **10**, 939–944.

247. Wilson,R.C. and Gilbert,L.A. (2017) The Promise and Challenge of in Vivo Delivery for Genome Therapeutics. *ACS Chem. Biol.*, **13**, 376–382.

248. Woo,C.J., Kharchenko,P. V, Daheron,L., Park,P.J. and Kingston,R.E. (2013) Variable requirements for DNA-binding proteins at Polycomb-dependent repressive regions in human HOX clusters. *Mol. Cell. Biol.*, **33**, 3274–3285.

249. Woo,J.W., Kim,J., Kwon,S. Il, Corvalán,C., Cho,S.W., Kim,H., Kim,S.-G., Kim,S.-T., Choe,S. and Kim,J.-S. (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.*, **33**, 1162–1164.

250. Wriggers, W., Chakravarty, S., and Jennings, P. A. (2005) Control of protein functional dynamics by peptide linkers. Biopolymers 80, 736−746.

251. Wu,L., Murat,P., Matak-Vinkovic,D., Murrell,A. and Balasubramanian,S. (2013) Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. *Biochemistry*, **52**, 9519–9527.

252. Wysocka,J., Swigut,T., Xiao,H., Milne,T.A., Kwon,S.Y., Landry,J., Kauer,M., Tackett,A.J., Chait,B.T., Badenhorst,P. *et al.* (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, **442**, 86–90.

253. Xiao,H., Xuan,W., Shao,S., Liu,T. and Schultz,P.G. (2015) Genetic incorporation of _-*N*-2-hydroxyisobutyryl-lysine into recombinant histones. *ACS Chem. Biol.*, **10**, 1599–1603.

254. Xu,W.S., Parmigiani,R.B. and Marks,P.A. (2007) Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene*, **26**, 5541–5552.

255. Yan, W., Imanishi, M., Futaki, S., and Sugiura, Y. (2007) α-Helical Linker of an Artificial 6-Zinc Finger Peptide Contributes to Selective DNA Binding to a Discontinuous Recognition Sequence. Biochemistry 46, 8517−8524.

256. Yang,Y., Ma,J., Song,Z. and Wu,M. (2002) HIV-1 TAT-mediated protein transduction and subcellular localization using novel expression vectors. *FEBS Lett.*, **532**, 36–44.

257. Yun,M., Wu,J., Workman,J.L. and Li,B. (2011) Readers of histone modifications. *Cell Res.*, **21**, 564–578.

258. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.

259. Zelikin,A.N., Ehrhardt,C. and Healy,A.M. (2016) Materials and methods for delivery of biological drugs. *Nat. Chem.*, **8**, 997–1007.

260. Zeng,L., Zhang,Q., Li,S., Plotnikov,A.N., Walsh,M.J. and Zhou,M. (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*, **466**, 258–262.

261. Zhang,T., Cooper,S. and Brockdorff,N. (2015) The interplay of histone modifications—writers that read. *EMBO Rep.*, **16**, 1467–1481.

262. Zhu,P. and Li,G. (2016) Structural insights of nucleosome and the 30-nm chromatin fiber. *Curr. Opin. Struct. Biol.*, **36**, 106–115.