

# Applying Peptide and Protein Synthesis to Study Post-translational Modifications in Epigenetics and Beyond

Beat Fierz\*

**Abstract:** Epigenetics research focuses on the study of heritable gene regulatory mechanisms that do not involve changes of the DNA sequence. Such mechanisms include post-translational modifications of histone proteins that organize the genome in the nucleus into a nucleoprotein complex called chromatin, and which are of key importance in development and disease. Chemical biology tools as developed by my group, in particular synthetic peptide and protein chemistry, have been critical to elucidate epigenetic signaling mechanisms. As outlined below, they allow the reconstitution of chromatin carrying defined modifications and thus the elucidation of detailed molecular mechanisms.

**Keywords:** Chromatin · Epigenetics · Protein ligation · Synthetic histones

## Chromatin, a Key Organizer of the Genome

Eukaryotic cells, containing large genomes, have to contend with a packing problem: How to organize the meters-long strands of DNA (a diploid human cell contains about 2 m of DNA) in the micrometer-sized cell nucleus? Complicating the matter, the DNA has to remain accessible for the cellular machinery, *i.e.* for gene transcription, DNA replication or, in case of damage, DNA repair. Nature has solved these challenges by organizing the genomic DNA into an intricate nucleoprotein complex called chromatin (Fig. 1a): short stretches of DNA are wrapped around spindles made of histone proteins. The complex of 150 bp of genomic DNA with this protein core, containing two copies of each histone H3, H4, H2A and H2B, is called the nucleosome<sup>[1]</sup> (Fig. 1b). Nucleosomes are then lined up like beads-on-a-string (forming a ‘chromatin fiber’), covering the whole genome and thereby compact and organize the DNA. Within this fiber, nucleosomes can interact with each other and impart higher-order organization, *e.g.* by the formation of helical structures, loops or compact nucleosome clusters.<sup>[2]</sup> Whole chromosomes are finally formed by chromatin loops of increasing size, and occupy specific volumes within the nucleus. These chromosome territories are conserved across different cell types, indicating the non-random packaging of the genome.<sup>[3]</sup>

## A Chemical Language to Control Gene Expression

To access specific genes within chromatin, the cell relies on a number of chemical and biochemical mechanisms. Post-translational modifications (PTMs) of histone proteins play a key role, as they serve as a chemical address system to mark genomic regions that are actively expressed or temporarily, as well as permanently, shut down.<sup>[4]</sup> Histones contain unstructured, lysine-rich tail domains which protrude from nucleosomes (Fig. 1b), and these tails are the carriers of the aforementioned PTMs. The modifications are diverse in nature, encompassing the acetylation or methylation of lysine residues, methylation of arginines, phosphorylation of serines or the attachment of whole protein domains (*i.e.* the small protein ubiquitin) to specific lysines<sup>[5]</sup> (Fig. 1c). PTMs are deposited or selectively removed by a set of

‘writer’ and ‘eraser’ enzymes, and site-specifically read out by multivalent ‘readers’, *i.e.* proteins containing several PTM binding domains. Simultaneous interactions then enable these reader proteins to bind chromatin and change its local structure and function as a reaction on the PTM landscape.<sup>[6]</sup> Histone PTMs have thus been proposed to form a code, specifying the activity state of the underlying genome.<sup>[4]</sup> Indeed, particular PTMs are associated with active chromatin (*e.g.* H3 tri-methylated at lysine 4, written as the short-hand H3K4me3) or transcriptionally-repressed chromatin (*e.g.* H3K9me3 or H3K27me3) (Fig. 1a). This system enables the cell to control gene expression, *e.g.* in different cell types, without changes to the underlying DNA sequence and gene structure. Importantly, certain chromatin activity states are heritable across cell generations or even through the germline. Such heritable changes are thus termed ‘epigenetic’ (meaning ‘outside genetics’) and the field of their studies is ‘epigenetics’. Epigenetic regulation is central to organismal development and is involved in the development of many complex diseases, in particular cancer<sup>7</sup>, and the involved writer, eraser and reader proteins are important drug targets.<sup>[8]</sup>

## Exploring Chromatin with Chemical Biology

Due to the involvement of the large number of different PTMs, of different protein factors, and the complex arrangement of chromatin in the cell, unraveling the details of epigenetic regulation has posed a significant challenge to chemists and biologists. In my laboratory at EPFL, we develop methods allowing to quantitatively investigate multivalent chromatin systems, with a focus on dynamic processes. We combine methods for protein semi-synthesis to reconstruct defined chromatin containing patterns of PTMs with single-molecule fluorescence imaging. This enables us to directly observe chromatin processes on the single-molecule level and in real time in chemically defined chromatin environments. As a result, we can provide much needed molecular insight into chromatin regulation with molecular precision, potentially informing downstream research programs towards epigenetic therapeutics. In the following, I will highlight a number of our recent studies targeting mechanisms of gene repression and activation.

\*Correspondence: Prof. B. Fierz, E-mail: beat.fierz@epfl.ch, École Polytechnique Fédérale de Lausanne (EPFL), ISIC, CH-1015 Lausanne, Switzerland

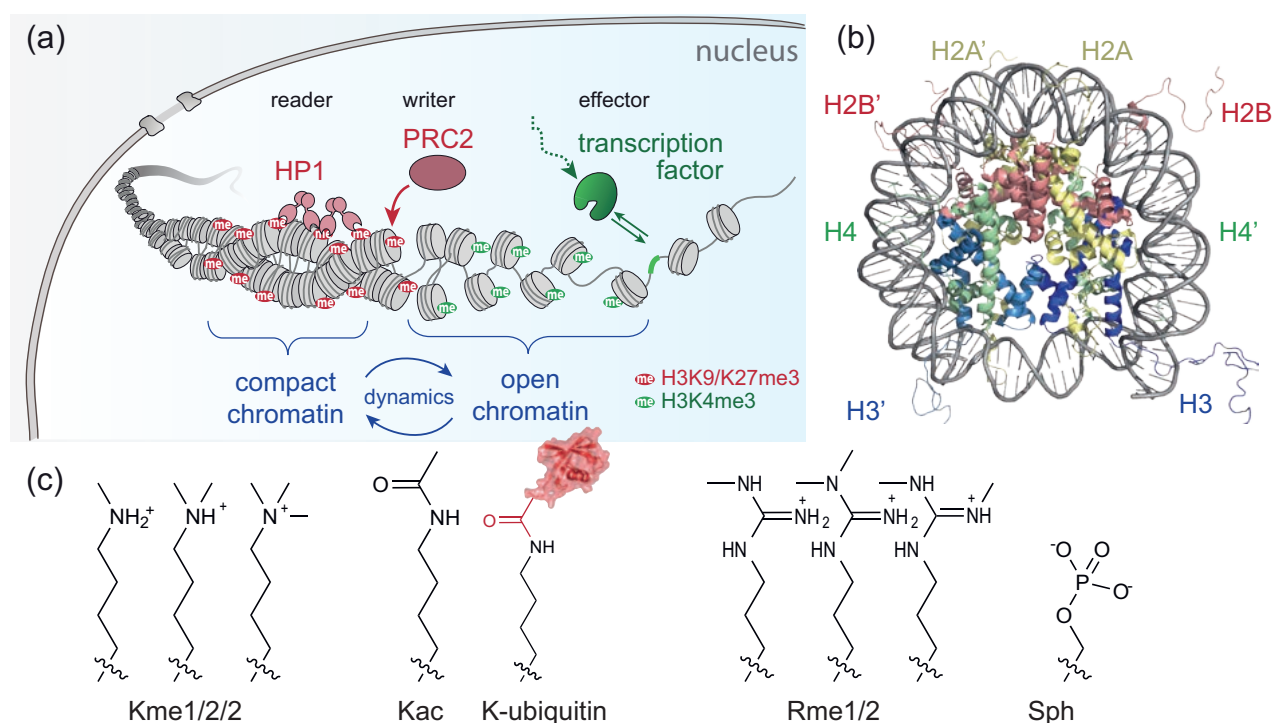


Fig. 1. Chromatin organization: (a) A schematic view of chromatin organization, indicating compact chromatin (heterochromatin) and open chromatin (euchromatin) with associated PTMs. (b) The structure of the nucleosome (PDB code: 1XK5) showing the position of the histone proteins H2A, H2B, H3 and H4. (c) Key histone PTMs, including lysine methylation (Kme1/2/3), acetylation (Kac), ubiquitylation (K-ubiquitin), arginine methylation (Rme1/2) and serine phosphorylation (Sph).

### Silencing Genes by Closing Down Chromatin

Transcriptionally silent chromatin, or heterochromatin (HC), is generally densely packed in the nucleus and exhibits reduced biochemical accessibility.<sup>[9]</sup> Different subtypes of HC can be discriminated by their histone PTM signature and the associated proteins: ‘constitutive HC’ is characterized by H3K9me3 and the presence of heterochromatin protein 1 (HP1).<sup>[10]</sup> In contrast, ‘facultative HC’ contains high levels of H3K27me3, ubiquitylation of histone H2A and is associated with polycomb group proteins (PcG).<sup>[11]</sup> To understand principles of gene repression, we started a research program to recapitulate, *in vitro*, key mechanisms of how these two distinct repressive systems function and how proteins and PTMs work together to establish a silent chromatin state.

### Gene Repression by HP1 Proteins

Initially, we focused on the function of HP1 in constitutive HC formation. HP1 can specifically bind to H3K9me3, dimerize and further oligomerize, enabling multivalent chromatin engagement.<sup>[12]</sup> We wondered how such multivalent interactions drive HP1 chromatin binding and how these interactions alter chromatin structure. To this end, we required methods to reconstitute chromatin carrying defined PTMs. My group employs protein semisynthesis *via* native chemical ligation (NCL)<sup>[13]</sup> or expressed protein ligation (EPL)<sup>[14]</sup> for the traceless synthesis of histone proteins containing defined PTM patterns.<sup>[15]</sup> Indeed, in recent years, we and others have generated specifically modified chromatin containing almost all known PTMs,<sup>[5]</sup> which enabled detailed studies of the function of individual or combined modifications. To investigate HP1 function, we thus synthesized a peptide containing residues 1–14 of histone H3, containing the K9me3 modification. We connected this peptide to the remainder of H3 using EPL to obtain the modified histone (Fig. 2a).<sup>[16]</sup> Using this protein, as well as all the remaining histones and purified DNA (carrying a fluorescent label), we then proceeded to reconstitute nucleosomes or chromatin fibers (Fig. 2b) which

were immobilized in a flow-cell on a microscopy slide. We then injected fluorescently-labeled HP1 proteins into the flow-cell and directly observed how single HP1 proteins bound transiently to chromatin fibers *via* dynamic single-molecule localization microscopy (SMLM, Fig. 2c). These experiments allowed us to extract kinetic parameters for this interaction as a function of the presence of H3K9me3 modifications.<sup>[16]</sup> Moreover, we fixed the oligomeric state of HP1 by a covalent linkage, using a ligand peptide containing two reactive cysteines in a ligand-directed dual ligation strategy (Fig. 2d). This allowed us to directly determine the effect of multivalent chromatin binding by HP1. Together with subsequent detailed kinetic modeling,<sup>[17]</sup> our experiments revealed that higher oligomeric states of HP1 result in faster chromatin recognition and longer residence times, showing that multivalency is a major driver for HC establishment.

We further investigated how the interaction of dimeric HP1 with chromatin alters chromatin structure. Attaching fluorescent dyes to the chromatin DNA enabled us to use the process of fluorescence resonance energy transfer (FRET) to measure chromatin structure and conformational dynamics<sup>[18]</sup> (Fig. 2e). This method revealed that chromatin fibers are always in motion, exhibiting structural fluctuations between locally compact and open states on the microsecond to seconds timescale. Within this dynamic ensemble, HP1 proteins captured transiently closed states and stabilized this compact conformation by cross-bridging nucleosomes.<sup>[18]</sup> Such compacted chromatin impairs biochemical access to DNA and thus is involved in gene silencing.

Together, these results gave detailed insight how a multivalent chromatin protein, *i.e.* HP1, can interact with and reshape chromatin in gene repression. In ongoing research, we are now designing cell-permeable peptides targeting HC loci in cells to visualize and modulate gene silencing. Moreover, in the course of this project we developed a toolbox of methods to probe chromatin processes on the single-molecule scale with chemical precision.

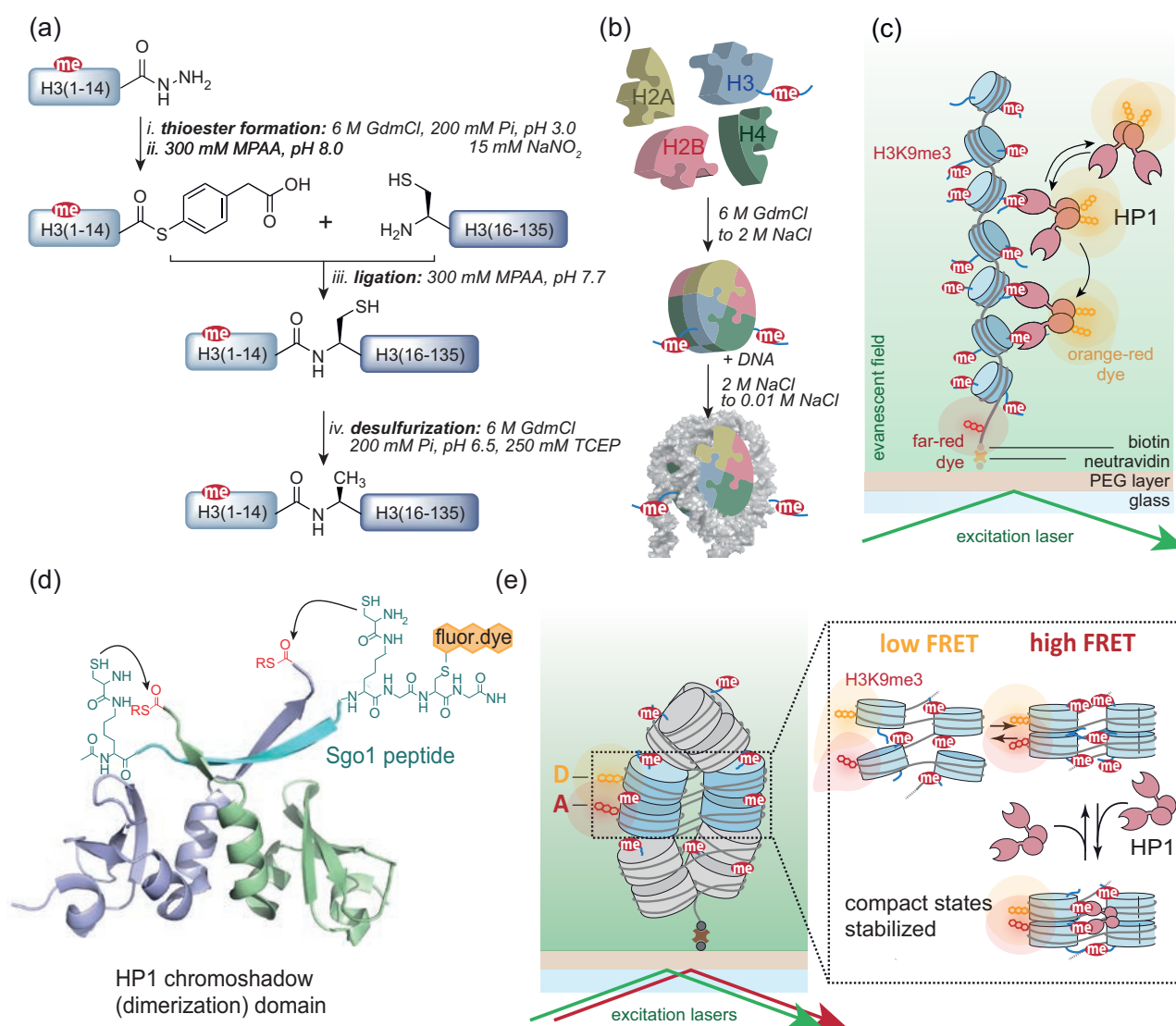


Fig. 2. Heterochromatin establishment by HP1: (a) Semisynthesis of H3K9me3 from a synthetic peptide, H3(1-14) carrying K9me3, and an expressed protein half carrying an N-terminal cysteine, using expressed protein ligation. GdmCl: guanidinium hydrochloride, MPAA: 4-mercaptophenylacetic acid, Pi: phosphate buffer, TCEP: tris(2-carboxyethyl)phosphine. (b) Chromatin reconstitution from unstructured histones and DNA via two dialysis steps. (c) Scheme of SMLM experiment to detect HP1 binding, using total internal reflection fluorescence (TIRF) microscopy. (d) Chemical dimerization of HP1 via a ligand peptide (Sgo1L) carrying two reactive cysteine handles attacking pre-installed C-terminal thioesters at the HP1 upon binding. (e) Scheme of FRET experiment: Energy transfer is measured between a FRET donor (D) and a FRET acceptor (A) dye, reporting on inter-dye distance and thus of chromatin conformational fluctuations, influenced by HP1.

### Establishment of Facultative Heterochromatin

In contrast, gene silencing in facultative HC involves PcG proteins, in particular the two complexes polycomb repressive complex 1 (PRC1) and PRC2. PRC2 is a methyltransferase that tri-methylates H3K27 and binds specifically to this PTM. PRC1 also binds H3K27me3, it further ubiquitylates H2A and compacts chromatin.<sup>[11]</sup> Here, we applied our chemical biology approaches to systematically explore PRC2 enzymatic regulation and chromatin targeting. Of particular interest to us was the fact that embryonic stem cells (as well as some cancer cells) contain a particular chromatin state right on the balance of repression and activation.<sup>[19]</sup> This chromatin state, called ‘bivalent chromatin’ is characterized by the co-existence of the repressive PTM H3K27me3 and the active PTMs H3K4me3 or H3K36me3 on the same nucleosome but on separate copies of H3 in an asymmetric fashion.<sup>[20]</sup> We wondered how PRC2 is regulated by pre-existing PTMs on substrate nucleosomes and if the generation of bivalent chromatin is an intrinsic property of the enzyme.

We thus developed a general method to assemble nucleosomes containing an asymmetric distribution of PTMs<sup>[21]</sup> (Fig. 3a). In

this approach, the differently modified histones, one containing H3K27me3 and the other H3K4me3 (or any other combination of modified or unmodified histones) are synthesized *via* EPL. Moreover, the histones contain a protease-cleavable N-terminal tag containing a single cysteine residue (the *lnc*-tag, Fig. 3b). This cysteine is then subsequently used to transiently crosslink two differently modified histone proteins, forcing them into the same nucleosome upon refolding. Finally, the *lnc*-tag is removed using TEV protease, leaving no trace of the whole procedure. Using this approach<sup>[21]</sup> (and variants thereof targeting H4,<sup>[22]</sup> Fig. 3c) we synthesized the whole ensemble of nucleosomes carrying asymmetric modifications on H3 and H4 which were found in stem cells.<sup>[20]</sup> Using this library we could show that PRC2 methyltransferase activity is activated by pre-existing H3K27me3 in asymmetric nucleosomes across the nucleosome surface, allowing the enzyme to propagate this PTM along the chromatin fiber. In contrast, PRC2 activity was found to be locally inhibited by pre-existing H3K4me3 or H3K36me3, limiting its encroachment into euchromatin<sup>[21,23]</sup> (Fig. 3c). Importantly, SMLM experiments further demonstrated that H3K27me3 (and surprisingly also H3K36me3)

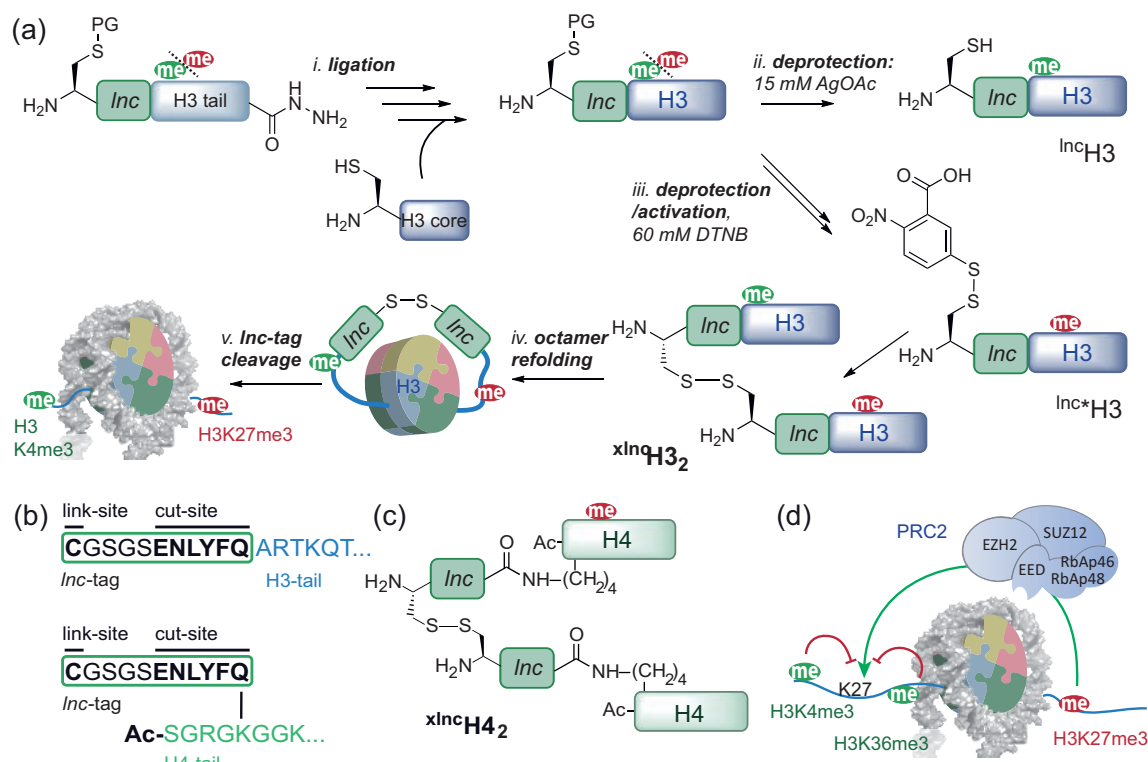


Fig. 3. Probing PRC2 regulation: (a) Semisynthesis of nucleosomes asymmetrically modified by H3K4me3 (green) and H3K27me3 (red), via disulfide-linked, modified H3 ( $\text{xInc}^{\text{me}}\text{H3}_2$ ). *Inc*: the Inc-tag (for the sequence see panel (b)), DTNB: 5,5-dithio-bis-(2-nitrobenzoic acid (Ellmann's reagent)). (b) Sequences of *Inc*-tag at H3 (top) and H4 (bottom). The cut-site can be cut with TEV protease, removing the tag without leaving a mark. (c) For the assembly of nucleosomes asymmetrically modified at H4 (here H4K20me3) a similar strategy is employed as in (a), but the *Inc*-tag was introduced at an internal lysine (K5) to allow the native acetylation at the H4 N-terminus. (d) Regulation of PRC2 by pre-existing modifications on asymmetric nucleosomes. Green arrows mark stimulation, whereas red lines indicate repression of enzymatic activity.

enhanced PRC2 chromatin interactions, thereby providing a basis for dynamic propagation of this silent chromatin state.<sup>[23,24]</sup>

Current ongoing efforts in the laboratory are now focused on developing chemical and biophysical methods to decipher the activity of PRC1 complexes, which read out the H3K27me3 chromatin state, compact and silence the regions and ubiquitylate H2A. With these studies we aim to provide a dynamic and mechanistic picture how chromatin inhibits gene expression on the molecular level.

### Mechanisms of Chromatin Invasion

Structural opening of chromatin is instrumental to yield biochemical access to the genomic DNA. In particular the DNA binding proteins involved in transcription regulation, *i.e.* transcription factors, need to enact local chromatin opening to fulfill their regulatory role. The mechanisms involved in these processes are complex and require collaboration between different molecular players. Putting our chemical and biophysical tools to use, we recently gained novel insights into this dynamic process. We studied the transcription factor Rap1, which is a critical first DNA binder (a 'pioneer factor') paving the way for a cascade of subsequent DNA binding proteins to regulate ribosomal gene expression in budding yeast.<sup>[25]</sup> Based on nucleosome positioning data, we reconstituted the yeast promoter chromatin architecture *in vitro*, and observed how Rap1 can invade chromatin structure using SMLM and smFRET<sup>[26]</sup> (Fig. 4a). We observed that Rap1 can directly bind to compact chromatin DNA using its flexible DNA binding domains and locally pry open chromatin structure. Then, Rap1 collaborates with a veritable molecular machine, the remodeling factor RSC,<sup>[27]</sup> which uses chemical energy in the form of ATP to displace or wholly remove nucleosomes. Rap1 biases the action of RSC such that its target sites are cleared of chromatin,<sup>[26]</sup> which enables the binding of further factors and of gene transcription at

these loci (Fig. 4b). Our laboratory is now exploring the mechanistic details of how DNA binding factors and chromatin remodelers collaborate in diverse environments and contexts.

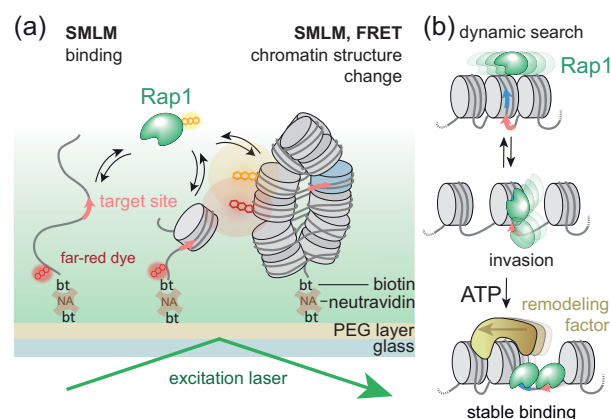


Fig. 4. Chromatin invasion by a pioneer transcription factor. (a) Scheme of Rap1 DNA, nucleosome and chromatin binding experiments. Indicated in red: Rap1 target sequence. (b) Chromatin invasion mechanism by Rap1 through local chromatin invasion and collaboration with the RSC remodeling factor.

### Conclusions and Outlook

By developing synthetic protein and peptide chemistry we can reconstitute complex chromatin systems, carrying defined patterns of PTMs and mimicking the naturally occurring states. This allows us to decipher mechanisms of genome regulation in molecular detail in real time using single-molecule and super-resolution microscopy methods. While these *in vitro* systems are

becoming ever so complex, we are now pushing towards implementing complementary techniques in living cells – enabling a chemical control of the chromatin state, or designing engineered probes to image changes in the chromatin states.<sup>[28]</sup> Moreover, we are expanding our methods into areas of DNA repair or innate immunity function within chromatin.<sup>[29]</sup>

However, beyond chromatin and epigenetics there exist other cellular systems which are regulated by PTM patterns, and we are just starting to develop the methodologies to probe their function. One system of particular interest to us is the cytoskeleton: The proteins making up the microtubules, tubulins, contain unstructured tails decorated by PTMs, which have been proposed to form a ‘tubulin code’ involved in health and disease.<sup>[30]</sup> Chemical protein synthesis methods, paired with imaging and biophysics will be required to crack this protein code, providing further challenges for dedicated chemical biologists.

### Acknowledgements

The author thanks all his current and former team members. Research support: Swiss National Science Foundation (SNSF) Project funding, 31003A\_149789 and 31003A\_173169, NCCR Chemical Biology, European Research Council (ERC) (grant 724022) and EPFL.

Received: April 8, 2021

- [1] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251, <https://doi.org/10.1038/38444>
- [2] K. Luger, M. L. Dechassa, D. J. Tremethick, *Nat. Rev. Mol. Cell. Biol.* **2012**, *13*, 436, <https://doi.org/10.1038/nrm3382>
- [3] D. Zink, H. Bornfleth, A. Visser, C. Cremer, T. Cremer, *Exp. Cell Res.* **1999**, *247*, 176, <https://doi.org/10.1006/excr.1998.4311>
- [4] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41, <https://doi.org/10.1038/47412>
- [5] M. M. Muller, T. W. Muir, *Chem. Rev.* **2015**, *115*, 2296, <https://doi.org/10.1021/cr5003529>
- [6] I. O. Torres, D. G. Fujimori, *Curr. Opin. Struct. Biol.* **2015**, *35*, 68, <https://doi.org/10.1016/j.sbi.2015.09.007>
- [7] P. Chi, C. D. Allis, G. G. Wang, *Nat. Rev. Cancer* **2010**, *10*, 457, <https://doi.org/10.1038/nrc2876>
- [8] K. Helin, D. Dhanak, *Nature* **2013**, *502*, 480, <https://doi.org/10.1038/nature12751>
- [9] C. Beisel, R. Paro, *Nat. Rev. Genet.* **2011**, *12*, 123, <https://doi.org/10.1038/nrg2932>
- [10] F. Hediger, S. M. Gasser, *Curr. Opin. Genet. Dev.* **2006**, *16*, 143, <https://doi.org/10.1016/j.gde.2006.02.013>
- [11] T. Cheutin, G. Cavalli, *Curr. Opin. Genet. Dev.* **2014**, *25*, 30, <https://doi.org/10.1016/j.gde.2013.11.016>
- [12] D. Canzio, E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar, B. Al-Sady, *Mol. Cell* **2011**, *41*, 67, <https://doi.org/10.1016/j.molcel.2010.12.016>
- [13] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776, <https://doi.org/10.1126/science.7973629>
- [14] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705, <https://doi.org/10.1073/pnas.95.12.6705>
- [15] B. Fierz, T. W. Muir, *Nat. Chem. Biol.* **2012**, *8*, 417, <https://doi.org/10.1038/nchembio.938>
- [16] S. Kilic, A. L. Bachmann, L. C. Bryan, B. Fierz, *Nat. Commun.* **2015**, *6*, 7313, <https://doi.org/10.1038/ncomms8313>
- [17] L. C. Bryan, D. R. Weilandt, A. L. Bachmann, S. Kilic, C. C. Lechner, P. D. Odermatt, G. E. Fantner, S. Georgeon, O. Hantschel, V. Hatzimankatis, B. Fierz, *Nucleic Acids Res.* **2017**, *45*, 10504, <https://doi.org/10.1093/nar/gkx697>
- [18] S. Kilic, S. Felekyan, O. Doroshenko, I. Boichenko, M. Diura, H. Vardanyan, L. C. Bryan, G. Arya, C. A. M. Siedel, B. Fierz, *Nat. Commun.* **2018**, *9*, 235, <https://doi.org/10.1038/s41467-017-02619-5>
- [19] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, E. S. Lander, *Cell* **2006**, *125*, 315, <https://doi.org/10.1016/j.cell.2006.02.041>
- [20] P. Voigt, G. LeRoy, W. J. Drury, 3rd, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia, D. Reinberg, *Cell* **2012**, *151*, 181, <https://doi.org/10.1016/j.cell.2012.09.002>
- [21] C. C. Lechner, N. D. Agashe, B. Fierz, *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 2903, <https://doi.org/10.1002/anie.201510996>
- [22] N. Guidotti, C. C. Lechner, B. Fierz, *Chem. Commun.* **2017**, *53*, 10267, <https://doi.org/10.1039/C7CC06180C>
- [23] N. Guidotti, C. C. Lechner, A. L. Bachmann, *ChemBioChem* **2019**, *20*, 1124, <https://doi.org/10.1002/cbic.201800744>
- [24] J. Choi, A. L. Bachmann, K. Tauscher, C. Benda, B. Fierz, J. Müller, *Nat. Struct. Mol. Biol.* **2017**, *24*, 1039, <https://doi.org/10.1038/nsmb.3488>
- [25] B. Knight, S. Kubik, B. Ghosh, M. J. Bruzzone, M. Geertz, V. Martin, N. Dénevaud, P. Jacquet, B. Ozkan, J. Rougement, S. J. Maerkl, F. Naef, B. Fierz, *Genes Dev.* **2014**, *28*, 1695, <http://www.genesdev.org/cgi/doi/10.1101/gad.244434.114>
- [26] M. Mivelaz, A. M. Cao, S. Kubik, S. Zencir, R. Hovius, I. Boichenko, A. M. Stachowicz, C. F. Kurat, D. Shore, B. Fierz, *Mol. Cell* **2020**, *77*, 488, <https://doi.org/10.1016/j.molcel.2019.10.025>
- [27] Y. Ye, H. Wu, K. Chen, C. R. Clapier, N. Verma, W. Zhang, H. Deng, B. R. Cairns, N. Gao, Z. Chen, *Science* **2019**, *366*, 838, <https://doi.org/10.1126/science.aay0033>
- [28] A. M. Delachat, N. Guidotti, A. L. Bachmann, A. C. A. Meireles-Filho, H. Pick, C. C. Lechner, C. Deluz, B. Deplancke, D. M. Suter, B. Fierz, *Cell Chem. Biol.* **2018**, *25*, 51, <https://doi.org/10.1016/j.chembiol.2017.10.008>
- [29] B. Guey, M. Wischniewski, A. Decout, K. Makasheva, M. Kaynak, M. S. Sakar, B. Fierz, A. Ablasser, *Science* **2020**, *369*, 823, <https://doi.org/10.1126/science.aaw6421>
- [30] C. Janke, M. M. Magiera, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 307, <https://doi.org/10.1038/s41580-020-0214-3>

### License and Terms



This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (<http://chimia.ch/component/sppagebuilder/?view=page&id=12>).

The definitive version of this article is the electronic one that can be found at <https://doi.org/10.2533/chimia.2021.484>