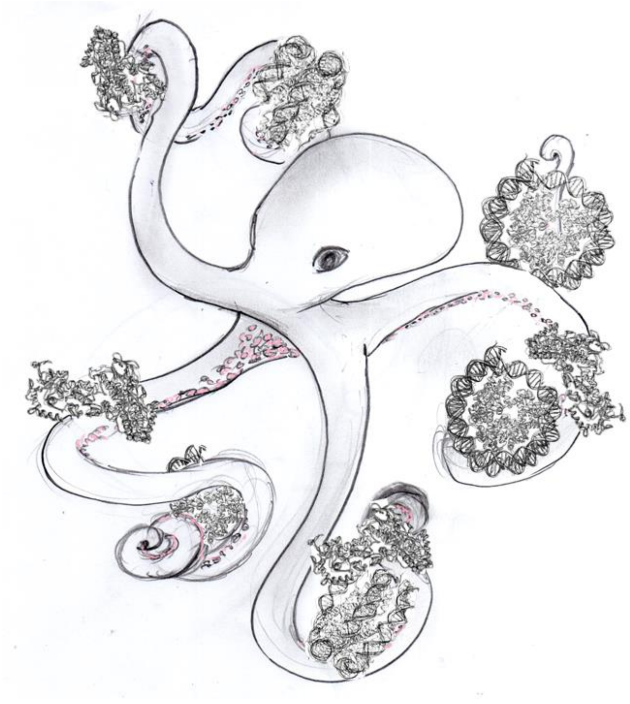
UNIVERSITÀ
DI PAVIA

Department of Biology and Biotechnology “Lazzaro Spallanzani”

Nucleosome modification by histone demethylases



Chiara Marabelli

PhD course on Genetics, Molecular and Cellular Biology

Cicle XXXI - A.A. 2015-2019

UNIVERSITÀ
DI PAVIA

Department of Biology and Biotechnology “Lazzaro Spallanzani”

Nucleosome modification by histone demethylases

Chiara Marabelli

Supervised by Prof. Andrea Mattevi and Prof. Claudia Binda

PhD course on Genetics, Molecular and Cellular Biology

Cicle XXXI - A.A. 2015-2019

Cover image: LSD2/NPAC is a multi-pronged system perfectly tailored for rapid and processive chromatin demethylation in the context of RNA polymerase II. Image by Chiara Marabelli.

Outline of this thesis

This thesis will present my research on LSD1 and LSD2, the only histone demethylases of the flavin class. They share the same substrate and reaction mechanism. Yet, they are recruited by distinct transcriptional macromolecular complexes with opposite effects on the chromatin state.

In **Chapter I**, I will introduce the reader to the world of epigenetics, with a particular emphasis on nucleosome recognition by nuclear players. In the last six years, nucleosome-chromatin remodeler complex structures, obtained by X-ray crystallography and single particle cryo-EM provided deep insights into the structural basis of nucleosome recognition. However, structural characterization of interactions between flexible or dynamic regions has remained challenging. Combination of multiple and "hybrid" structural and biochemical approaches allowed characterization of the mechanism employed by LSD1 for nucleosome recognition, which necessitates DNA binding by the protein partner CoREST1.

In **Chapter II**, I will present my research on the characterization of the relevant biomedical aspects of LSD1/CoREST. In particular, I analyzed the biochemical and structural effects of three pathological LSD1 single-residue mutations. It turned out that enzyme catalysis is only partially affected, and the structure not at all, whereas diminished binding of the H3 tail and of other protein factors through the substrate-binding pocket might be the cause of such severe neurological and physiological symptoms. Here I attached the two published papers to which I contributed dissecting the effect of the pathological mutations on the enzyme folding and on histone H3 binding and catalysis, as well as on the recruitment of non-substrate regulatory proteins and transcription factors such as SNAIL1 and p53.

The technical "know-how" I acquired during the characterization of LSD1/CoREST, greatly helped the structural and mechanistic investigation of its homolog LSD2 with its partner NPAC, illustrated in **Chapter III**. Yet, the LSD2/NPAC project turned out to be a more challenging and exciting team effort than ever expected. Particular technical and methodological choices greatly helped us to complete a solid description of LSD2/nucleosome complex formation mechanism. On this basis, I decided to describe in detail the rationale behind particular methodological choices, as well as the implementation and optimization of the protocols. Given that most of the data have been published (Marabelli *et al.*, 2019), I will extensively present only those experiments I personally participated to. I am also including few unpublished results giving further insights on the biological function of the LSD2/NPAC system, whose chapter have been marked by an asterisk (*). The mechanism for nucleosome recognition by LSD2 is different than that of the previously characterized LSD1. LSD2/NPAC does not contact the core of the nucleosome, which is recruited by a tail-only mechanism. The short NPAC-linker module is extremely efficient in regulating the substrate histone tail processing, whereas other domains of NPAC also affect the avidity and processivity of the demethylase system. In accordance with literature, this machinery seems to be perfectly tailored to support the work of RNA-Polymerase II.

The effort on studying the dehydrogenase domain of NPAC, led us to a new line of research, regarding the changing role of this enzyme during evolution: it appeared indeed that a single-point mutation might have shifted NPAC from being a cytosolic enzyme to a nuclear auxiliary subunit. This study has now been published and it is also reported in **Chapter III**, given its deep connection with the LSD2/NPAC story.

In conclusion, this thesis will show how LSD1 and LSD2 employ different mechanisms for nucleosome recruitment, despite sharing identical catalytic properties. The key role of the demethylase catalytic domain and the associated flavin cofactor, an aspect highly conserved in both enzymes, gives us a stimulating example of how a cell can employ the same tool for different purposes. Finally, biochemical and structural characterization of LSD2/NPAC mechanism of action will now provide a basis to the understanding of the biological significance and the biomedical implications of this enzymatic complex, similarly to what happened for LSD1.

List of abbreviations

AT-hook	Adenine-thymidine- hook
β-HAD	β-hydroxiacid-dehydrogenase
Cryo-EM	Transmission electron microscopy at cryogenic temperature
DNA	Deoxy-ribonucleic acid
DH	Dehydrogenase
EDTA	Ethylene-di-amine tetra-acetic acid
FAD	Flavin adenine dinucleotide
GLYR1	Glyoxylate reductase 1 homolog, synonym of NPAC
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H3K4me1/2	Histone H3 lysine 4 mono- and di-methylated
H3K36me3	Histone H3 lysine 36 tri-methylated
H3.1	Histone H3 variant 1
H4	Histone H4
HDAC	Histone deacetylase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria broth medium
KDM	Histone lysine-specific demethylase
KDM1A	LSD1
KDM1B	LSD2
LSD1/2	Lysine-specific histone demethylase 1/2
NAD+/NADH	Nicotinamide adenine dinucleotide
NADP+/NADH	Nicotinamide adenine dinucleotide phosphate
NCP	Nucleosome core particle
NDF	Nucleosome destabilizing factor, alias NPAC
NP60	Nuclear protein 60, alias NPAC

OD_{600nm}	Optical density measured as the 600 nm wavelength absorbance
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PMSF	Phenyl-methyl-sulfonyl fluoride
PTM	Post-translational modification
PWWP	Proline-Tryptophan-Tryptophan-Proline
RNA	Ribonucleic acid
rt	Room temperature
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SHL	Super-helical location
Tris	Tris(hydroxymethyl)aminomethane)
YNB	Yeast nitrogen base with ammonium sulfate without aminoacids
YPD	Yeast extract peptone dextrose medium

Table of contents

Chapter II – Biological and pathological effects of LSD1 37

Chapter I - Introduction

The fate of each cell is continuously designed within the nucleus, where small DNA-containing entities, called nucleosomes, regulate usage of the genetic information. Nuclear factors employ many different mechanisms for nucleosome recognition and modification, whose characterization has only recently become feasible thanks to the newly developed methods in cryo-electron microscopy. Histone demethylases LSD1 and LSD1, despite performing identical reactions, employ different strategies for substrate processing, with totally antithetical effects on gene expression.

Chromatin structure and function

1. The “epigenetics” concept

Since Miescher’s discovery of “nuclein” in 1869, that seemingly inert polymer of sugar, phophoric acid and nitrogen bases, carefully enveloped in the cell nucleus, was known to be key to transmission of the hereditary information. But only after the description of its double-helical structure in 1953 (Crick and Watson, 1953) the huge potential of DNA became evident (Alberts *et al.*, 2002). Because of its complementary duplex nature, the sequence of the nucleotides on one strand can be either replicated into a new identical molecule or transcribed to RNA and successively translated to proteins. The first successful cloning demonstrated that DNA is actually an active master of life processes (Gurdon, 1962), and following experiments with more and more sophisticated techniques only confirmed the “magic” role of the nuclear material in directing the cell destiny (Wilmut *et al.*, 1996). Those exciting years of a worldwide scientific race to unravel the secrets of DNA ended up in the Human Genome Project, which finally ended in 2003 (Noble, 2003).

“Today, we are learning the language in which God created life. […] humankind is on the verge of gaining immense, new power to heal. […] In coming years, doctors increasingly will be able to cure diseases like Alzheimer's, Parkinson's, diabetes and cancer by attacking their genetic roots”
US President Bill Clinton, on the Completion of the First Survey of the Entire Human Genome Project. June 26, 2000. (Clinton, 2000)

It turned out later that the picture is actually not that simple, and that not all phenotypical or pathological traits could be guessed from a given nucleotide sequence. Scientists begun ascribing to “epi-genetics” all those cases in which genetics alone was not sufficient to explain what was going on (Figure 1). Probably the first epigenetic fact one might think about is the specialization of a totipotent zygote cell, which can assume many different identities despite maintaining its genetic information unchanged.

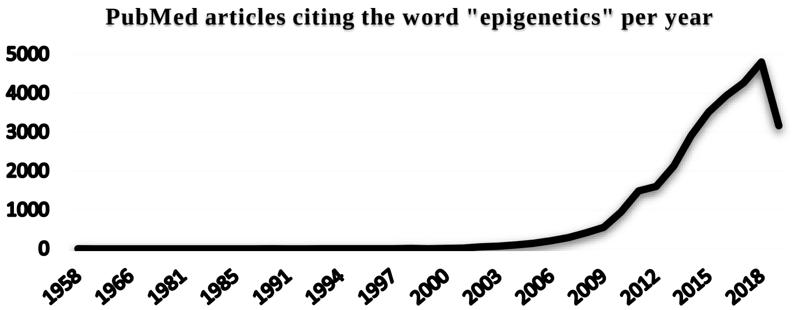


Figure 1. Distribution over time of the scientific publications in PubMed (NCBI) mentioning the word "epigenetics". It is interesting to note that the number of publications increases exponentially since 2000, the same year in which the Human Genome Project was announced to be almost completed. Data for 2019 are incomplete as the site was visited on July 30th, 2019.

2. Chromatin structure regulates gene expression

It became then significant to observe that DNA is never present as a "pure" molecule in the cell nucleus. Actually, DNA is only one third of the incredibly intricated nucleic acid-protein complex named chromatin (Cooper, 2000). Indeed, each chromosome and each chromosome portion is organized in different nucleo-protein architectures, designed to regulate usage of the underlying genetic information. For example, segments of hundreds of kilobases to several million bases in length, are spatially confined within specific nuclear compartments, either alone or together with functionally associated portions of the same or of different chromosomes. Modern techniques of fishing with fluorescent probes allowed visualization of these structures (Bintu *et al.*, 2018), leading to the unexpected discovery that their organization is typically stable across generations and evolutionarily conserved between related species (Dixon, Gorkin and Ren, 2016).

At each level of topological organization, each domain defines the compaction degree of the long DNA molecules and in turn the usage of the genetic information carried by them (Figure 2, left). As the zygote cell acquires a more and more defined identity, the previously totally relaxed DNA molecules fold up into various architectures, ergonomically designed for all the needed levels of transcription. For example, non-coding repetitive sequences are usually put apart in the densest and most inert structures of chromatin, whereas intensely transcribed segments, such as rRNA genes in the nucleoli, have more extended and relaxed conformations easily accessible to DNA- and RNA-Polymerase machineries. Both the sequence of the DNA itself, as well as environmental factors, dictate the final "compaction degree" and the transcription level of the genetic information implicated.

Such a large number of complex decisions on the transcription level of each single gene, gene exon or silent portion of the DNA, is the result of a relatively small number of interplaying factors compared to the infinite possible chromatin states a cell could experience. Which is the rationale behind such a fine handling of the enormous potential of the genetic information?

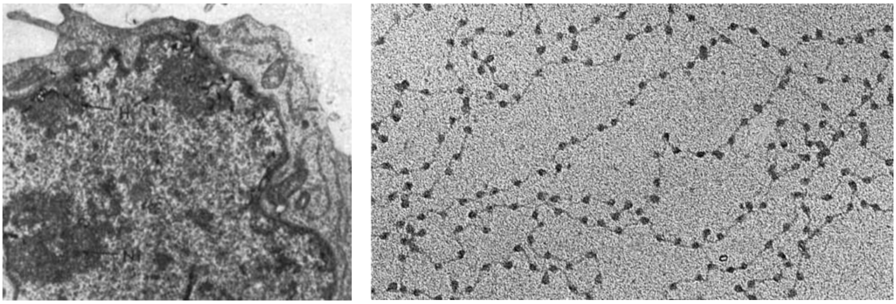


Figure 2. On the left, the very first electron micrograph of a cell nucleus, showing chromatin is organized in regions, characterized by different densities (Milner and Hayhoe, 1968). On the right, a second electron micrograph of a partially digested chicken erythrocyte chromatin (Woodcock, 2010). It is evident the "beads-on-a-string" structure of chromatin in its elementary state. Each "bead" is a nucleosome, linked to two neighbouring ones by linker DNA.

3. The nucleosome is the basic unit of chromatin

Extended DNA digestion of chromatin revealed that DNA is organized in repetitive, building blocks: the nucleosome core particles (Figure 2, Kornberg, 1974; Oudet, Gross-Bellard and Chambon, 1975). Each 147 bp portion of the acidic DNA molecule is bent almost twice in a left-handed superhelix around a core of positively charged proteins named histones (Figure 3).

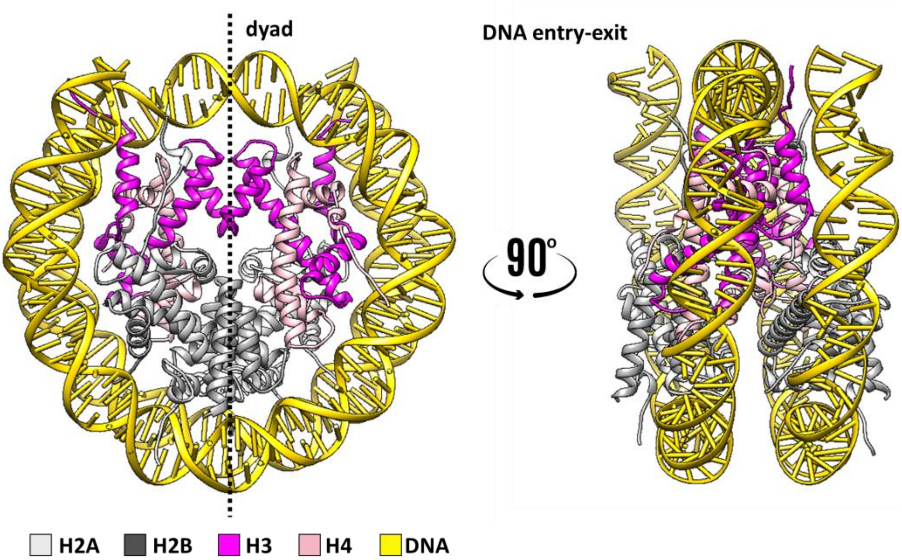


Figure 3. The 147-bp nucleosome core particle structure we obtained from our cryo-EM experiments (Marabelli *et al.*, 2019). The backbone of the 147-bp DNA is in yellow, histones H2A and H2B are light and dark grey respectively, H3 is coloured in magenta and H4 is pink.

The electrostatic interactions guiding the assembly of the nucleosome core particle are extremely conserved. The central 120 bp of DNA are organized around a rigid heterotetramer of two copies each of histones H3 and H4. The N-terminal alpha-helix of H3 binds additional 13 bp on each side of the symmetric unit (Luger *et al*, 1997). Two H2A-H2B dimers also join through multiple interactions with the protein core and further stabilize the DNA on either side. The ordered assembly of the NCP can be easily experienced during *in vitro* reconstitution of the histone octamer and a DNA fragment of the proper length, as the first species to form are the tetrasome and the hexasome (DNA complexes with the H3-H4 tetramer and an additional H2A-H2B dimer for the latter), finally the NCP (hereafter named simply "nucleosome"), which is the most temperature- and salt-resistant species (the T_m value for the recombinant human NCP in 200 mM KCl is 72.5 °C) (Taguchi *et al.*, 2014).

Since the first crystallographic studies led by Aaron Klug (Finch *et al.*, 1977; Richmond *et al.*, 1984) it appeared evident that a rigid core of the octamer folded domains arranges DNA in a precisely defined conformation, whereas disordered histone tails protrude, with seemingly no contribution to the stability of the complex. DNA is so tightly bent around the octamer that a change in the double heical pitch occurs with respect to the standard B conformation, where major and minor grooves show various degrees of supercoiling depending on their nucleosomal location (or super-helical location, SHL). This highly stable, and seemingly rigid conformation of DNA within the nucleosome obviously comes at the price of a significantly lower accessibility to nuclear factors. At the time of the atomic structure of the nucleosome (Luger *et al.*, 1997) it was already known that ATP-dependent machineries able to bind the nucleosome core and unwrap its DNA are necessary to allow the RNA Polymerase II to overcome the nucleosome barrier (Hirschhorn *et al.*, 1992; Kruger *et al.*, 1995). It was implicit in many of the following investigations, that the octamer surface and the nucleosome electrostatics together would generally act as a stable docking platform for various nuclear factors, usually involved in DNA transcription, replication and repair.

Nucleosome dynamics and accessibility

1. Nucleosomal DNA is intrinsically dynamic

Differently from what is suggested by the crystal structure, the nature of the nucleosome has now been proved to be highly dynamic. Regarding the nucleic acid component only, DNA wrapping and unwrapping is actually a physiological process, referred to as “DNA breathing”, for which nucleosomes in the nucleus are at least partially unwrapped 2-10% of their time (Poirier *et al.*, 2009). This implies that certain DNA sites, usually buried within the nucleosome fold, can be spontaneously exposed *in vivo* to DNA-binding proteins and machineries. The structures of partially unwrapped nucleosomes have already been observed through cryo-EM (Bilokapic, Strauss and Halic, 2018a, 2018b).

The rate of DNA breathing is strongly dependent on the nucleotide sequence itself (Lowary and Widom, 1998; Li *et al.*, 2005). Indeed, a tool to regulate accessibility to genetic information, is the chemical modification of the DNA itself. Addition of a methyl group at position 5 of the cytosine base always leads to gene repression (Feil, 2009). These methyl groups indeed project into the double helix major groove and sterically inhibit binding proteins, while providing the binding site for methyl-DNA binders within repressor complexes (Hendrich and Bird, 1998). Not only, it seems that methylated CpG sequences have a higher affinity for the octamer, and thus the DNA wraps more tightly within the nucleosome (Collings, Waddell and Anderson, 2013). DNA methylation is a stable modification that usually persists along the entire life of vertebrate organisms (Suzuki and Bird, 2008).

2. Histone modifications

There are other strategies to regulate DNA breathing, more tunable than nucleosome positioning over the DNA sequence, or its stable modification. The histone-variant composition is one of the strategies for modulation of the nucleosome properties. As it is in the case of DNA methylation, modification of the protein component of the nucleosome not only controls the arrangement of the nucleosome and of nucleosomal arrays, but it also creates new binding sites for several proteins and protein complexes. For example, the histone variant H2A.Z docks onto the H3-H4 tetramer in a looser interaction, and hence detaches more easily. It also exposes a particularly negatively charged area on the surface which is recognized by specific chromatin remodeler complexes in turn (Dechassa *et al.*, 2011). The H3 variant of centromeric nucleosomes lacking the H3 N-terminal helix leads to the accommodation of only 121 bp of DNA (Tachiwana *et al.*, 2011).

The easiest and most rapid way to shape nucleosome properties is the addition or removal of chemical modifications on the protruding protein tails. Indeed the N-terminal tails of all histones and the C-terminal tail of H2A also have a role, despite their highly flexible nature. The interactions between histone lysine and arginine residues with the phosphate groups of DNA further influence the nucleosome stability (Iwasaki *et al.*, 2013) and accessibility by DNA-binding proteins and chromatin-remodelers modules (Zhou, Gaullier and Luger, 2018). The addition of a negatively charged acetyl group onto the lysine of a histone tail generally lowers the electrostatic attraction between DNA and the octamer core. Acetyl marks are indeed mostly found within actively transcribed regions, where the intra- and inter- nucleosomal interactions are more relaxed, and sterically more accessible to transcriptional machineries (Marmorstein and Zhou, 2014).

Not only the type of modification but also its position on the nucleosome surface affects the final outcome. In this view, addition of the same acetyl group onto lysine 9 or lysine 56 of the same histone H3 would make a significant difference. On the one hand, masking of lysine 4 positive charge affects the dynamics of the H3 N-terminus only, which is more prone to explore “open”, “DNA-detached” conformations (Fu *et al.*, 2017). This regulates the binding of other chromatin binders and remodelers in turn (Emma A. Morrison *et al.*, 2018). On the other hand, histone H3 lysine 56 locates at the DNA entry-exit site, where the presence of an acetyl goup increases twofold the rate of DNA unwrapping, and hence the site exposure of DNA internal sites (North *et al.*, 2012).

Differently from acetylation, methylation does not alter the charge of histone proteins and probably does not directly affect nucleosomal interactions. Methylation however occurs on both arginines and lysines of H3 and H4 histone tails, onto which different degrees of methylation are possible, leading to opposite outcomes: trimethylation of lysine residues H3K9, H3K20 and H4K27 is a characteristic of heterochromatin, whereas monomethylation of exactly the same residues is found in activated regions, as well as mono-, di- and tri- methylated lysines of histone H3: K4, K36 and K79 (Barski *et al.*, 2007; Wagner and Carpenter, 2012; Højfeldt, Agger and Helin, 2013). Given the extreme versatility of the methylation mark, its role is most often designed to provide a binding platform for specific chromatin readers, despite significant changes of nucleosome dynamics due to histone methylation marks have been studied (Zhou, Gaullier and Luger, 2018) (Figure 4).

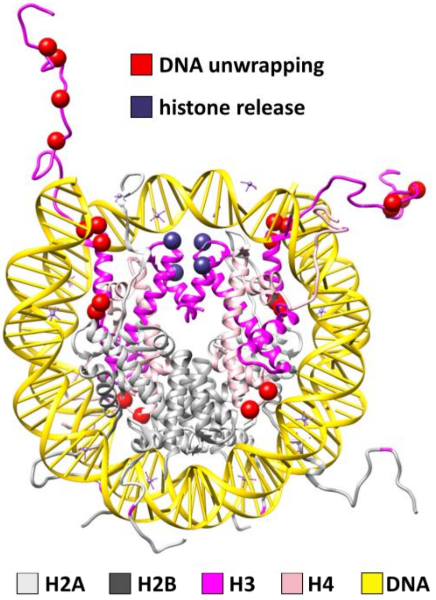


Figure 4. Histone PTMs whose effect on nucleosome stability is known. Many of them are subjected to acetylation: H3K9, H3K14, H3K18, H3K23, H3K56, H3K115, H3K122, H4K77 and H4K79. H3Y41, H3T45 and H3S57 have been studies for their phosphoryl mark, whereas H3R42 for the effect of its demethylation. At the moment, only two effects due to histone PTMs have been described and are indicated in figure: DNA unwrapping and a looser histone-histone packing. From PDB: 1KX5 (Luger *et al.*, 1997).

A plethora of other histone modifications have been characterized over the past twenty years through modern techniques in molecular biology and mass-spectrometry. The most relevant of those include phosphorylation, ADP-rybosilation, ubiquitination and sumoylation, as well as isomerization of proline residues and deamination of arginines (Kouzarides, 2007). Cold Spring harbor laboratories created an extended and detailed list that is available online (Zhao and Garcia, 2015). Multiple chemical modifications can be host simultaneously by the same nucleosome, where they modulate interactions both between DNA and histones (within the same and among different nucleosomes) as well as the recruitment of non-histone proteins and enzymes critical for DNA processes (Patel, 2016). The destiny of each 147 bp strand of DNA is then designed separately, although not independently, and eventually the architecture of chromatin fibers and regions is only the outcome of multiple factors interplaying at the single nucleosome level.

The importance of histone post-translational modifications (PTMs) is not only confined to the life of a single cell, but also to its descendance. During replication, DNA is not completely unwound from the nucleosome and each new copy strand may receive modified histone proteins, which will guide as reference model the modification of the “new” incoming histones, and in turn the epigenetic state of the synthesized DNA molecule (Jablonka and Lamb, 1998; Lind and Spagopoulou, 2018).

3. The nucleosome is substrate to all chromatin processes

A complex pattern of post-translational modifications and histone variants defines nucleosome accessibility and dynamic state, yet this pattern itself is dynamic, and can last up to a few minutes until it is further modified according to the needs of the cell (Patel, 2016). All histone PTMs are indeed reversible because of the concerted action of "writer" and "eraser" enzymes, each one specifically recognizing its substrate nucleosomal epitope: histone acetyl-transferases and their deacetylase counterparts (HATs and HDACs), histone lysine and arginine methyl-transferases and demethylases (HMTs and KDMs), ubiquitin and sumo proteases along with other types of hydrolases, isomerases and phosphatases, etc..

Histone chaperones which exchange histone components, and other chromatin remodelers can dramatically change the nucleosome conformation. Through sliding nucleosomal DNA, they modulate its accessibility by usually giant machineries involved in DNA replication, translation and repair. For example, FACT is the necessary partner of RNA polymerase II. Indeed the transcription machinery is able to only partially disrupt the histone-DNA interactions at a few sites on the nucleosome (Farnung, Vos and Cramer, 2018; Kujirai *et al.*, 2018). FACT is necessary the proper sliding of RNA Pol-II over the nucleosomal DNA without pausing. There are many chromatin remodelers catalyzing quite different non-covalent modifications of the nucleosome. Nonetheless, all of them share a conserved DNA-binding, ATP-ase domain for DNA recognition and its energetically expensive translocation (Narlikar, Sundaramoorthy and Owen-Hughes, 2013; Clapier *et al.*, 2017). They all have a greater affinity for nucleosomes over free DNA and are classified according to the nucleosome elements positively recognized for catalysis (McGinty and Tan, 2015; Gamarra *et al.*, 2018).

All histone modifying and chromatin remodeling enzymes, are strictly regulated in time and localization by the pattern of PTMs on the substrate nucleosome as well as by non-histone protein partners, chromatin-readers and multi-enzymatic protein complexes. A complex scenario appears, in which each nucleosome is the point where many different pathways continuously converge and integrate their information (Figure 5). Subtle variations in the intricate network of nuclear players can significantly affect global levels of gene expression.

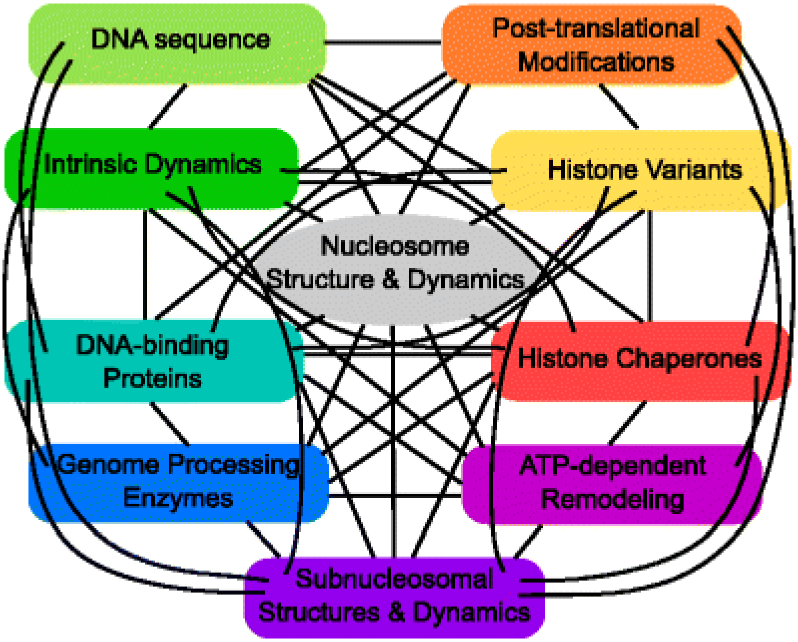


Figure 5. Nucleosome structure and dynamics are affected by many different factors: sequence and methylation state of the DNA (light green) and DNA binders (green blue), histone variants (yellow) and PTMs (orange), the action of histone chaperones (red) and nucleosome remodelers (magenta), with their associated machineries for chromatin architecture organization (violet) and genome processing (dark blue). All these mechanisms are coherent and act in a concerted manner (black lines). Image from Ordu, Lusser and Dekker, 2016.

Nucleosome recognition by chromatin factors

1. The nucleosome is a switchable docking module

Most of the chromatin factors recognize nucleosomes as their substrate, and not naked DNA. Moreover, the huge number of different nucleosome binders, each with its own specificity, suggests that every single chemical component of the nucleosome, even those less exposed on its surface, is recognized by a chromatin factor and thus involved in a chromatin pathway. This means that the nucleosome cannot be considered a “barrier” to genetic information any more, but as an active player in regulation of gene expression instead. The number of possible interactors for a given epigenetic state can be further amplified by its varied dynamic conformations. Moreover its surface is extremely variable because of the DNA and histone modification which provide docking modules for selected subsets of interactors (McGinty and Tan, 2015; Zhou, Gaullier and Luger, 2018; Ricketts *et al.*, 2019).

2. Nucleosomal DNA recognition

The tightly bent conformation assumed by DNA offers the framework for nucleotide-protein interactions to occur, in particular at the aligned grooves. Supercoiling and stretching, plus the conformation of each groove and strand of the double helix, depends on the nucleotide sequence itself. By sensing the sequence-specific conformation adopted, viral integrases can recognize their target insertion point (Maskell *et al.*, 2015). Another well-known binding site provided by DNA only, is at the dyad. Here the H1 histone binds the entry-exit DNA segments, "locking" the nucleosome in a rigid state (Bednar *et al.*, 2017) (Figure 6, left). Human transcription factors FoxA1 and GATA4 (Cirillo and Zaret, 2007) can target DNA sites on heterochromatin nucleosomes, where they displace linker histones, to create a nucleosomal configuration recognizable by other factors (Iwafuchi-Doi *et al.*, 2016). Pioneer factors solve the "chicken and egg" problem of how specific DNA sequences embedded within seemingly unaccessible heterochromatin assemblies are "switched on" to their functional state (Zaret, Lerner and Iwafuchi-Doi, 2016). Also chromatin remodeler Snf2 has been recently demonstrated to specifically recognize nucleosomal DNA at SHL±2 and SHL±5 (Figure 6, right). Interestingly, the cryo-EM structure of the complex also revealed a DNA stretching at the bound locations (Li *et al.*, 2019), which implies that particular positions of nucleosomal DNA have sufficient freedom to acquire further new conformations to allow the catalytically productive interaction with ATP-dependent chromatin remodelers.

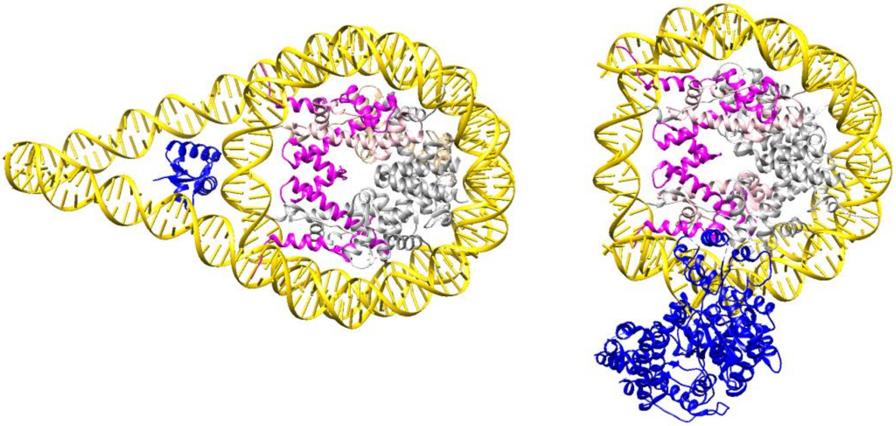


Figure 6. Cryo-EM structures of the nucleosome in complex with linker histone H1 (left, PDB 5NL0) or Snf1 chromatin remodeler (right, PDB 5X0X). The colour scheme of the nucleosome components follows that of figures 3 and 4.

3. The acidic patch is a stabilizing docking point

On the octamer surface, the ‘acidic patch’ on H2A–H2B, was the first inter-nucleosomal interaction site to be described in 1997 (Luger *et al.*, 1997; Kalashnikova *et al.*, 2013). This pocket comprises eight negatively charged amino acid residues (H2A: Glu56, Glu61, Glu64, Asp90, Glu91, Glu92; H2B: Glu105 and Glu113). Crystallographic studies on viral nucleosome-binding proteins revealed a common strategy for anchoring of this patch through an arginine residue (Figure 7, Makde *et al.*, 2010; Kalashnikova *et al.*, 2013; McGinty and Tan, 2015). It is interesting to note that until present, crystallographic approaches to chromatin complexes were successful only for complexes in which the acidic patch was recognized by the nucleosome binding protein. The fact that well-diffracting crystals require that the repeating unit is conformationally stable, plus the seemingly conserved “arginine anchor” approach to the acidic patch, suggests that this site is generally employed for stable docking of the chromatin factor onto the nucleosome surface. Modification of the critical H2A and H2B residues shaping the acidic patch can change the accessibility of the nucleosome by all these factors as well as histone H4 tails from neighboring nucleosomes in packed chromatin (McGinty and Tan, 2015; Zhou, Gaullier and Luger, 2018).

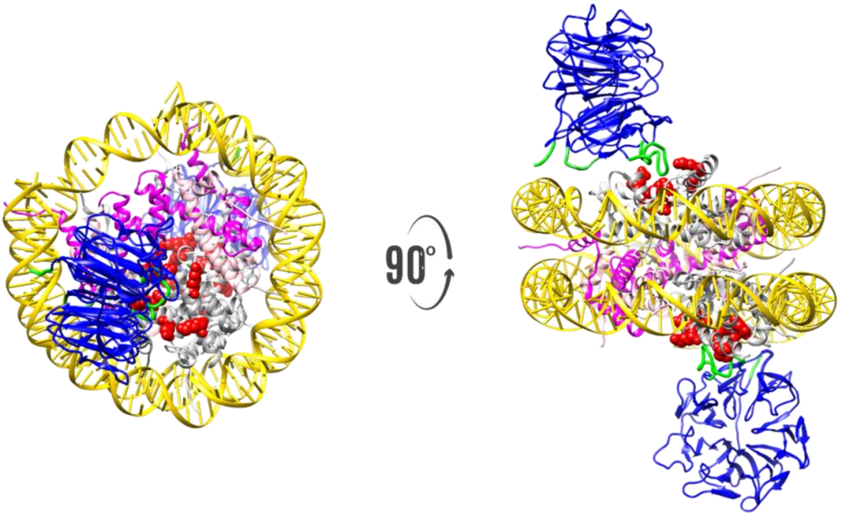


Figure 7. Crystal structure of the RCC1-nucleosome complex (PDB entry 3MVD) (Makde *et al.*, 2010). The H2A and H2b alpha-carbons of the acidic patch residues are evidenced by red spheres. Two RCC1 subunits recognize both the acidic patches of the same nucleosome in a symmetric fashion, as clearly seen from the nucleosome side point of view, on the right. The colour scheme of the nucleosome components follows that of figures 3 and 4.

4. Histone tails recognition

The flexible N-termini of all four histones and the C-terminus of H2A are the most variable nucleosomal component, and offer plenty of recognition modules because of histone variants, the resident PTMs and their conformations with respect to DNA and the acidic patch as well. Indeed, they can be employed as “signals“of the epigenetic state of a particular nucleosome to direct chromatin remodelers and other DNA-processing enzymes onto their appropriate target. For example, DNA methyltransferases 3A and 3B both need recognition of the H3K36me3 mark for their activation (Rondelet *et al.*, 2016), and DNA- and RNA-polymerase machineries rely on multiple histone reader modules as well (Wen *et al.*, 2014; Fei *et al.*, 2018). On the same trend, all histone modifier enzymes till now characterized do not recognize only their substrate histone residue in a specific way. They bind indeed at least to a second nucleosomal epitope, whose recognition triggers allosteric regulatory mechanisms for the enzymatic activity(Zhou, Gaullier and Luger, 2018).

Bromo-domains, PHD fingers and PWWPs are small chromatin reader modules specifically devoted to recognition of a particular epigenetic state of a target histone tail particularly when in close proximity to DNA (Weaver, Morrison and Musselman, 2018). It has been demonstrated for many of them indeed, that the sum of their affinities for the isolated histone tail peptides and DNA oligos cannot explain the very strong binding of the same components within the nucleosome (Savitsky *et al.*, 2016; Weaver, Morrison and Musselman, 2018). Cooperative binding of distinct nucleosome components can increase the affinity up to three orders of magnitude as in the case of the PWWP domain of human Lens-Epitelium-Derived factor (LEDGF) (Eidahl *et al.*, 2013; Van Nuland *et al.*, 2013). In addition, it has already been discussed how cooperative binding between linker DNA and a particular conformation of a histone tail, can also direct a DNA-binding protein to its specific nucleosomal location (Emma A Morrison *et al.*, 2018).

Not only DNA, but also the flat surface of the octamer can contribute to recognition of particular histone tail motifs. Human 53BP1, a reader of H2AK15ub and H4K20me2 involved in double-strand break (DSB) repair, also specifically contacts H2B C-terminal helix and the nearby acidic patch (Wilson *et al.*, 2016) (Figure 8, left). A second example of this dual recognition mode of the nucleosome is the Polycomb repressive complex (PRC1), a H2AK119 ubiquitinase. The crystal structure of its ubiquitination module in complex with the nucleosome shows several interactions other than the ones between the substrate H2A C-terminus and the active cleft of the enzyme. The acidic patch, the C-terminal end of H3, as well as DNA, all support the stabilization of the enzyme onto its nucleosomal substrate, allowing proper orientation of the enzyme with respect to the nucleosome and thus the specific recruitment of the target tail (McGinty, Henrici and Tan, 2014) (Figure 8, right).

The cooperative binding of both histone tails and a second site of the nucleosome, either DNA or the octamer surface, is a general strategy employed by histone readers and modifier enzymes. Usually, the enzymatic module of histone modifiers and of chromatin remodelers is accompanied by distinct auxiliary domains conferring specific properties to the nucleosome-recruiting complex (Ricketts *et al.*, 2019). The same enzyme can be endowed with multiple domains or organized in different subunits, so that many different complexes can direct the same enzymatic activity onto different nucleosomal substrates. Cooperative recognition of multiple nucleosomal sites seems then to be a valid strategy to regulate nucleosome processing itself at various interdependent levels: the epigenetic state of the nucleosome selects the decorating non-histone proteins, and in turn dictates the composition of the resulting complex. Histone tails are thus like switches of the nucleosome control panel, and their epigenetic state can determine the chromatin remodeling activities of the nucleosome itself and of the nearby nucleic acid.

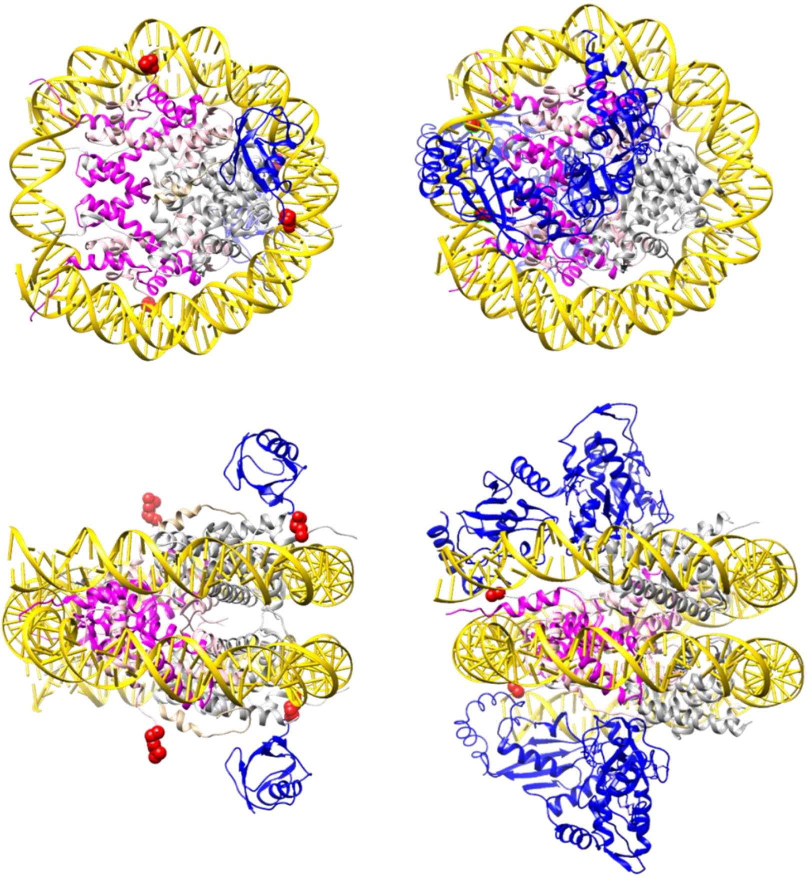


Figure 8. Cryo-EM nucleosome structure of 53BP1 bound to the nucleosome (left, PDB code 5KGF), and crystal structure of PRC1, also bound to its substrate nucleosome (right, PDB code 4R8P). Both nucleosome factors 53BP1 and PRC1 have been coloured blue. The colour scheme of nucleosomal components follows that of figures 3 and 4. The histone substrate residues (H2AK15 and H4K20 for 53BP1, and H2A118-119 for PRC1 have been coloured red. In the latter case, the choice to evidence the additional non-substrate residue H2A118, is due to the fact that one of the substrate residues of the symmetric unit is not visible in the PDB. It is evident from these images that histone binders and modifiers have a preference for stable docking onto the octamer disk surface, rather than the single-point interactions with their target residues.

Nucleosome recognition by LSD1 and LSD2

The same catalytic tool for opposite functions

Among all histone modifiers, histone methyl-transferases and demethylases surely are the most variegate group, where each enzyme needs to be specific not only for the substrate histone residue, but for its degree of methylation as well. It interesting to note that all lysine and arginine methyltransferases employ the same S-Adenosyl methionine (SAM) cofactor and most of the histone demethylases rely on Fe(II) and α-ketoglutarate, whereas only two demethylases use flavin adenine dinucleotide (FAD) (Mosammaparast and Shi, 2010).

The puzzling role of the latter two, named lysine-specific demethylases 1 and 2 (LSD1 and LSD2), becomes even more intriguing given the fact that they demethylate the same substrate H3K4me1/2 through exactly the same mechanism. Not only, despite the strikingly similarity of their catalytic domains, their biological effects are opposite: LSD1 silences gene promoters, whereas LSD2 sustains transcriptional elongation. The important role of auxiliary subunits is evident in this case: Recognition of the substrate tail by LSD1 is triggered by its DNA-binding partner CoREST (Kim *et al.*, 2015; Pilotto *et al.*, 2015), whereas LSD2 nucleosomal activity is regulated by NPAC (Chen *et al.*, 2013; Fang *et al.*, 2013; Fei *et al.*, 2018; Marabelli *et al.*, 2019).

From a methodological point of view, structural characterization of the interaction between an histone binder/modifier and the flexible histone tail within the nucleosome remains challenging. LSD1 has been successfully investigated mainly through combined multiple approaches including fluorescence polarization, analytical SEC and mutational analysis.

In the following pages (20-29), a review by our group (Marabelli *et al.*, 2016) is attached to further describe in more detail the state of the art about LSD1, and to highlight the open questions about the different biological roles of LSD1 and LSD2.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

The growing structural and functional complexity of the LSD1/KDM1A histone demethylase

Chiara Marabelli, Biagina Marrocco and Andrea Mattevi

LSD1 was the first discovered histone demethylase. Using a flavin-dependent oxidative mechanism, LSD1 demethylates the N-terminal tail of histone H3 in the context of a variety of developmental processes. This functional complexity involves the association with nuclear factors and non-coding RNAs. A number of exciting studies are uncovering the bases of these specific and diverse molecular interactions, which occur both at catalytic and non-catalytic regions of the enzyme. Alternative splicing and post-translation modifications represent further layers for modulating this complex molecular network. By combining structural methods with the usage of chemically modified histones, it is becoming possible to visualize how LSD1 and associated co-repressors recognize the nucleosome. The enzyme clamps the nucleosomal particle through multivalent interactions mediated by the non-catalytic domains, which represent prospective sites for drug design.

Address
Department of Biology and Biotechnology, University of Pavia, 27100 Pavia, Italy

Corresponding author: Mattevi, Andrea (andrea.mattevi@unipv.it)

Current Opinion in Structural Biology 2016, **41**:135–144

This review comes from a themed issue on **Catalysis and regulation**

Edited by **David W Christianson** and **Nigel S Scrutton**

For a complete overview see the Issue and the Editorial

Available online 26th July 2016

http://dx.doi.org/10.1016/j.sbi.2016.07.011

Introduction

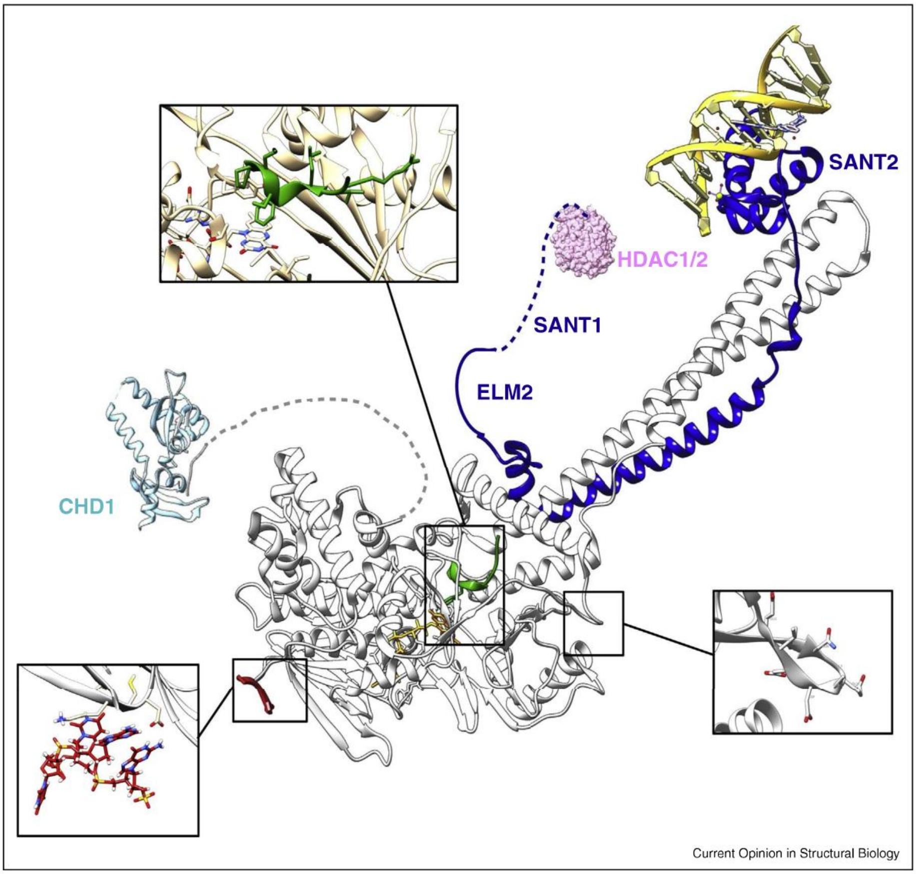
Among the variety of chromatin marks establishing the epigenome, histone methylation is one of the most versatile modifications. Mono-methyl, di-methyl, and tri-methyl groups can be found on lysine and arginine residues of nucleosomes at strategic chromatin positions. Specific patterns of histone methylation control enhancer commitment, promoter recognition by transcription factors, and co-translational gene-expression regulation. Until the discovery of the first histone lysine demethylase in 2004, this epigenetic mark had been considered to be irreversible [1]. However, presently a large number of enzymes with different specificities are known to remove methyl marks from a broad range of chromatin substrates [2•]. From a biochemical point of view, all these interesting epigenetic players can be clustered into two subclasses [3], the great majority of them rely on iron and oxoglutarate as co-factors, whereas there are only two enzymes known to exploit a flavin-dependent oxidative strategy. Of the latter, Lysine-Specific Demethylase 1 (also known as KDM1A) is the best characterized because it was the first histone demethylase discovered [1,4]. Since then, a wealth of research studies progressively unravelled many of its biological roles, highlighting a very complex, yet intriguing, scenario. Generally, LSD1 is emerging as a main executor of cell-fate determination from embryonic development to adult tissue regeneration. Evidence suggest that this demethylase and its associated proteins are implicated in the physiological processes that shape the identity of both stem and progenitor cells, along with their differentiation pathways to hematopoietic, neuronal, mesenchymal, sperm, and fat cells [5,6,7•,8•,9,10–12]. Pathological consequences of aberrant LSD1 activity range from cancer to neurological disorders and viral infection [13,14,15•,16].

It is beyond the scope of this review to discuss the diversity of biological processes that involve LSD1. The central notion is that such functional complexity is brought about by an equally intricate network of molecular interactions with various players: transcription factors, splicing factors, chromatin binders and remodelers, oncoproteins and tumour suppressors, DNA, non-coding RNAs including telomere-encoded RNAs, and obviously the nucleosomes. In particular, LSD1 is mostly found in complex with the REST co-repressor proteins (CoREST1–3) and the histone deacetylases 1 and 2 (HDAC1–2), forming a bi-functional system that integrates histone deacetylase and demethylase activities [17–19,20•]. Here, we review the state of the art about the structural and molecular properties of LSD1 with a focus on the binding properties at both catalytic and non-catalytic sites. Hence, we will summarize the proposed models for nucleosome recognition and the interesting differences with LSD2 (KDM1B), the other more recently discovered flavin-dependent histone demethylase.

A curiously elongated three-dimensional structure

LSD1 was initially discovered as a nuclear protein displaying sequence homology with FAD-dependent amine oxidases, a class of enzymes known to act on a broad range of substrates, either mono- or poly-amines [21]. Only later, it was found that LSD1 catalyses demethylation

Figure 1



Overall structure of LSD1-CoREST1, highlighting key functional sites. The H3 N-terminal tail (green) binds in the active-site cleft in the core of the amine oxidase domain. A surface loop (on the right) is subject to alternative splicing, forming a potential site for protein–protein interactions. The amine oxidase domain also harbours a binding-site for non-coding RNAs such as TERRA (orange). The N-terminal disordered residues (dashed line) can be methylated on specific Lys residues and are recognized by CHD1. The C-terminal SANT domain of CoREST binds with high affinity to DNA whereas the CoREST N-terminal residues are not present in the crystal structure and form the binding site for HDAC1/2.

of mono-methyl and di-methyl lysine 4 of histone H3 [1]. The chemistry of the reaction follows the classical amine oxidase scheme: two electrons are transferred from the methyl carbon of methylLys4 to the flavin in the form of a hydride anion [22•]. The reduced flavin then reacts with molecular oxygen, generating hydrogen peroxide. The oxidized imine-containing peptide finally hydrolyses, probably in a non-enzymatic process, to release formaldehyde and the Lys4-demethylated H3 product [4,23]. It is remarkable that a potentially toxic compound such as formaldehyde is produced in direct proximity to DNA. Intriguingly, folate has been shown to bind to the active site of the enzyme, and it remains to be seen whether this binding underlies a formaldehyde-scavenging role by this cofactor [24•,25].

Despite the similarity in sequence and catalytic mechanism, the amine-oxidase domain of LSD1 is embedded in an overall architecture that is substantially different from that of other mono-amine and poly-amine oxidases (Figure 1) [26]. The first characteristic feature is a long (about 150 amino acids) disordered segment at the N-terminus, which is involved in protein-protein interactions and is subject to post-translational modifications [27,28••]. This is followed by a SWIRM domain that rigidly packs against the FAD-binding core, thereby having a structural function. The third most striking feature of the enzyme architecture is the so-called tower domain, a long 90-residue insert, which divides the amine oxidase into two halves; it is formed by two antiparallel helices protruding from the globular catalytic core. This segment is essential for constitutive binding of CoREST, the presence of which is necessary to allow LSD1-dependent nucleosome demethylation [29,30••]. Indeed, the C-terminal residues of the co-repressor partner wrap around the tower domain in a very stable triple helical bundle, and fold into a DNA-binding SANT domain at the tip of the stalk. With inter-protein nanomolar affinities, LSD1-CoREST is such a tight complex that can be considered as a *bona fide* heterodimeric enzyme [31].

The catalytic centre lies at the interface between the two amine-oxidase domain lobes, buried within the enzyme main body (Figure 1). Differently from the closed cavities and narrow tunnels typically found in amine oxidases, LSD1 features a more open funnel-shaped active site, which accommodates a very long portion of the histone tail substrate [32]. This structural feature is accompanied by the outstanding property of specific interaction with each one of the first twenty N-terminal amino acids of H3. A set of negatively charged side chains at the rims of the active site establish an extensive network of specific hydrogen bonds and salt bridges with the histone tail substrate, which adopts a folded conformation within the enzyme active site. Lys4 side chain of H3 is found exactly in front of the LSD1-flavin N5 atom as required by the oxidative reaction to take place. Most importantly, this mode of binding has profound functional implications because it enables LSD1 to sense a broad range of epigenetic marks and their combinations on the histone tail and to select among them. This exquisite specificity may indicate a possible chromatin-reading, non-catalytic function of the enzyme [33•,34]. In this context, HDACs represent a very interesting example of interplay between LSD1 specificity and the activity of other chromatin modifiers. Acetylated histone tails are poorly recognized by LSD1, and a deacetylation step by HDAC1/2 is suggested to be necessary for demethylation to occur [35,36,23]. From a biological point of view, this is fascinating because it suggests that the sequential reactions of HDAC1/2 and LSD1 are intimately linked through their common partner CoREST (Figure 1).

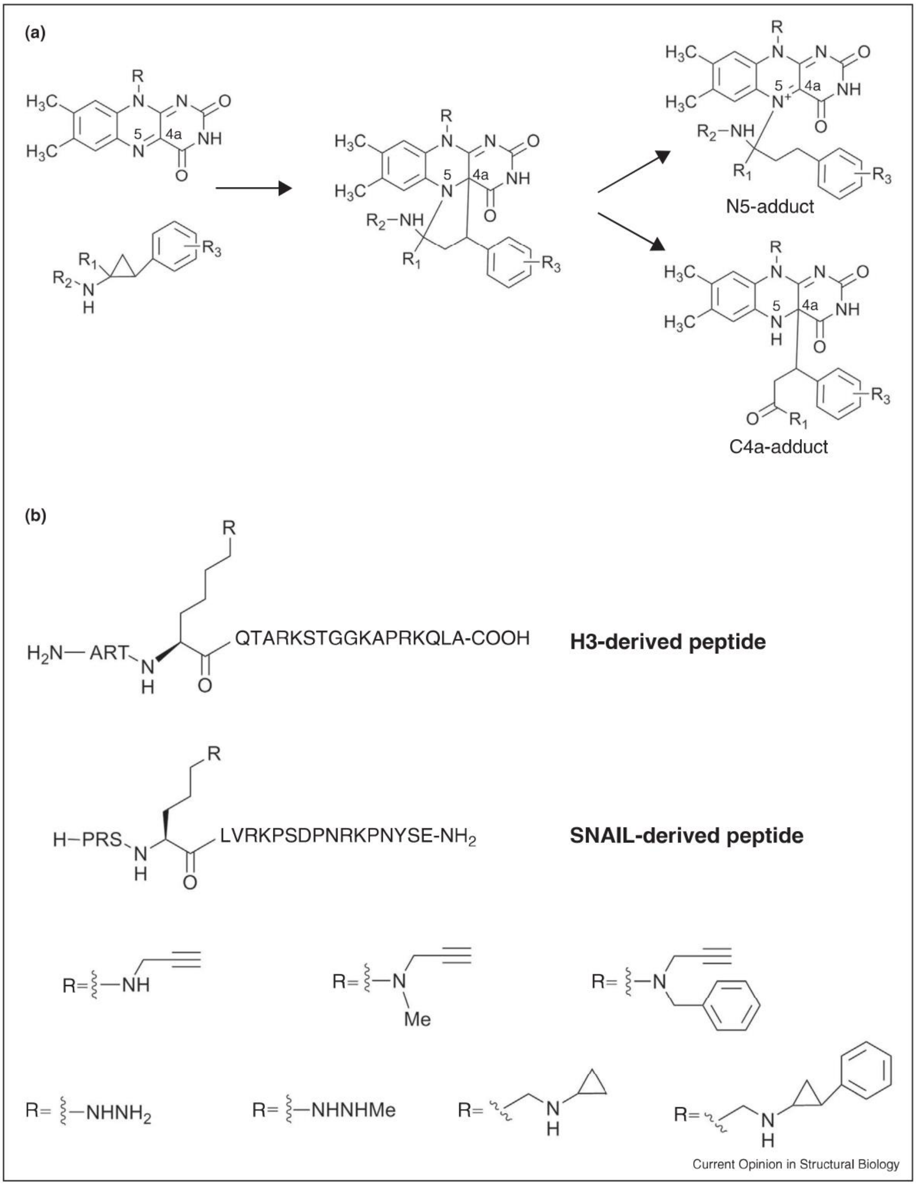
As revealed by recent studies, the functional importance of finely shaping the substrate-binding cavity bears pathological consequences. Three patients have been found to carry mutations in LSD1, which are associated to a newly described neurological disorder with intellectual disability. These mutations target active-site residues, causing mild-to-strong impairment in substrate binding and catalysis, likely altering the establishment and progress of the LSD1-dependent chromatin-modification programs that underlie cell differentiation in multiple tissues and organs [15•,30••].

Covalent strategies for LSD1 inhibition

The discovery that LSD1 is a histone demethylase functionally and structurally associated with HDAC1 and 2, both validated drug targets, immediately spurred intense efforts to generate LSD1 inhibitors and evaluate their pharmacological activity in relevant disease models. Many recent patents and articles describe novel small molecules, and support a general consensus that LSD1 is indeed a highly promising drug target in oncology [37,38•]. Three LSD1 inhibitors have entered Phase I clinical trials for the treatment of acute myeloid leukaemia, myelodysplastic syndrome, and small cell lung cancer (for a review see [39]). From a biochemical and structural standpoint, the similarity between LSD1 and monoamine oxidases A and B (MAOs A and B) had a ground-breaking effect as it inspired the investigation of well-known and widely used anti-MAO compounds as potential anti-demethylase molecules. Among them, tranylcypromine was found to be a weak inhibitor of LSD1 activity and it served as a useful lead for further medicinal chemistry optimization [35]. Crystallographic studies confirmed that tranylcypromine derivatives irreversibly react with FAD, forming a covalent adduct with the flavin ring as found in MAOs [40] (Figure 2a). This provided the foundation for the structure-based decoration of the tranylcypromine scaffold leading to compounds with greatly improved potency and specificity for LSD1 over MAOs. The landmark finding that LSD1 inhibition by tranylcypromine had synergistic activity when combined with all-*trans* retinoid acid further strengthened the rationale for LSD1 inhibition in haematopoietic myeloid malignancies [41]. The presumed mechanism is that the de-repression of transcription, following LSD1 inhibition, allows retinoic acid to drive terminal differentiation and tumour cell apoptosis [42,13]. These results suggest that LSD1 inhibition combined with other anti-cancer agents in the treatment of myeloid or other neoplasms maybe be highly effective [43]. In summary, this is a remarkable success story in which the wealth of knowledge accumulated on well-characterised homologous enzymes (i.e. MAOs) was effectively exploited to enhance drug development campaigns against the newly identified LSD1 target.

The search for additional LSD1 inhibitors is embracing new and ingenuous strategies. One widely explored idea focuses on employing novel warheads (Figure 2b). Many current inhibitors functioning as suicide substrates were designed as mechanism-based LSD1 inhibitors, such as propargyl amine, cyclopropylamine, and phenylcyclopropane peptides, which bind to the enzyme, react with the flavin, and form a covalent adduct [44–46]. The hydrazine chemotype was identified as a highly potent LSD-inhibiting functionality and H3-derived peptides with modified lysine residue, such as a highly cell-active

Figure 2



Strategies for inhibitor design against LSD1. (**a**) Tranylcypromine inhibits the enzyme by forming a covalent adduct with the flavin. N5, C4a, and cyclic adducts have been observed in different enzyme-inhibitor complexes. (**b**) Strategies for active-site targeting by means of suicide-peptide substrates, developed by introducing chemical groups (propargylamine, hydrazine, cyclopropylamine) that react with the flavin.

azaLys-containing peptide, have been recently reported [47]. Finally, an innovative approach has been the development of pan-histone demethylase inhibitors, obtained by coupling the chemical features of tranylcypromine with the scaffolds of inhibitors of the iron/oxoglutarate dependent de-methylase. This dual compounds exhibited potent anti-cancer activity in cells and represent a first example of hybrid molecules modulating different classes of histone-modifying enzymes [48•].

An on-going challenge is the targeting of LSD1-CoREST in the context of processes involving specific protein complexes, through the exploration of new druggable spaces [49]. Innovative inhibitors, as opposed to the more 'classic' covalent and non-covalent [50] active-site ligands, could pave the way to the 'next-generation' anti-LSD1 drug development. These avenues require an in-depth investigation of protein-protein interactions, involving the non-catalytic regions of LSD1, as described in the next section.

An array of interactions with proteins and nucleic acids

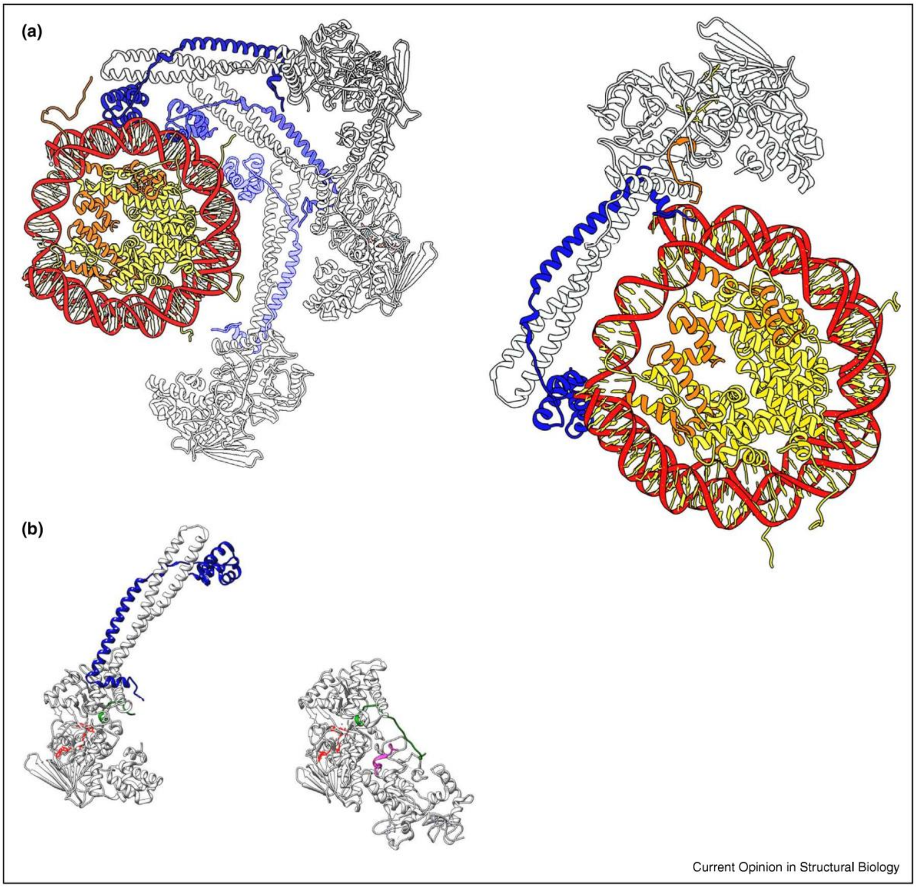
As often found for chromatin enzymes, a fundamental notion is that LSD1-CoREST is recruited by specific protein complexes in different cell types and differentiation programs. How can the same demethylase complex establish such specific interactions with so many different proteins and nucleic acids? What is the role of the non-catalytic domains? In the past few years, there have been several insightful reports addressing these issues. A first well characterized case-study elucidated the interaction with transcription factors of the so-called SNAIL family [51]. These factors function as master regulators of various differentiation programs from pituitary gland development to haematopoiesis [52]. They have been found to repress some of their target genes through recruitment of LSD1 and consequent H3-Lys4 demethylation at the target chromatin loci. This process is brought about by a fascinating histone mimicry strategy [53,54]. Namely, these transcription factors feature an N-terminal sequence that contains a pattern of positively charged residues that resembles the H3 tail (Figure 2b). In this way, their N-terminal residues bind tightly to the LSD1 active-site cleft, adopting a conformation very similar to that observed for the LSD1-bound H3 tail. Upon binding to the specific target genes, the transcription factors are released by LSD1, which becomes free to carry out its catalytic histone demethylase function. The important implication of this histone mimicry mechanism is that active-site inhibitors inherently disrupt interactions with these transcription factors, thereby functioning as protein–protein association inhibitors. In the case of SNAIL, this is a very relevant finding as this transcription factor recruits LSD1 to promote epithelial–mesenchymal transition through the repression of the E-cadherin gene, a hallmark of metastatic transformation.

Although SNAIL indicates that also the active site can be involved in protein–protein interactions, the non-catalytic regions remain the main players in these processes. This concept is effectively documented by the recent discovery that CHD1, a reader of the H3-trimethylLys4 mark, binds to LSD1 (Figure 1) [28••]. More specifically, it has been found that Lys114 on the N-terminal disordered region of LSD1 can be dimethylated by histone methyltransferase EHMT2 and, therefore, it is specifically recognized by CHD1. Interestingly, structural superpositions indicate that the binding of H3-trimethylLys4 and LSD1-dimethylLys114, though occurring in the same site of CHD1, is not fully overlapping. This interplay between EHMT2, LSD1Lys114me2, CHD1, and the H3 N-terminal tail is heavily involved in androgen-dependent gene-expression chromosomal-rearrangement, with far-reaching implications for prostate cancer treatment. This is a fascinating case of a chromatin-modifying enzyme that is modified by another methylase, thereby becoming target for specific recognition by a chromatin protein.

A splicing variant of LSD1 has been identified which is specifically expressed in neurons [55]. Structural studies have shown that this neuro-specific variant features a modified surface loop that contains four additional amino acids that are fully exposed at the protein surface and can be subject to phosphorylation (Figure 1). It is very likely that this newly introduced phosphorylation site represents a hook for specific interactions with factors present in neuronal cells and/or modifiers of 'core-corepressor association'; indeed, a phosphomimetic mutant loses the ability to retain association with CoREST and HDAC1/2 [56]. Recent papers indicate that the neuro-specific LSD1 variant might be endowed with altered substrate specific properties, acting on H3-Lys9 or H4-Lys20, although the possibility remains that these activities are due to another demethylase that might be part of neuro-specific LSD1-containing complexes [57•,58•]. Mice lacking the neuronal-specific LSD1 isoform display altered behaviour with regard to memory consolidation, emotional behaviour and perturbed response to psychological stress, beautifully highlighting the functional interplay between splicing, chromatin modification, and neuronal functions [57•,59•].

Also non-coding RNAs can be part of the LSD1 molecular biology. Initially, HOTAIR was found to associate to LSD1-containing protein complexes [60]. More recently, it has been discovered that LSD1 engages TERRA RNAs, which are encoded by telomeric sequences. This interaction elicits DNA-damage activity in uncapped telomeres [61•]. The crystal structure of RNA-bound LSD1 has now been published: the oligonucleotide binds in a surface cleft of the amine oxidase-SWIRM domain interface, which is distinct and remote from the active site

Figure 3



Catalytic and non-catalytic domains in LSD demethylases. (**a**) Schematic overview of the proposed mechanism of nucleosome binding by LSD1-CoREST. First, a DNA-scanning step (left panel) promotes the detachment of the histone tail from DNA to achieve the catalytically competent positioning of the enzyme over the nucleosomal disk (right panel). LSD1 is in light grey, CoREST in blue, and H3 in orange. H2a, H2b, and H4 are in yellow. (**b**) Comparison of LSD1 (left) and LSD2 (right) crystal structures, with their similar amine oxidase domains in the same orientation to outline the different non-catalytic domains. FAD and N-terminal H3 residues are in red and green, respectively. NPAC bound to LSD2 is in magenta and CoREST bound to LSD1 in blue.

[62••] (Figure 1). Interestingly, this surface was predicted by computational studies to be a possible target for intermolecular interactions [49]. Further experiments will clarify how specific this binding site is and whether it can be targeted by specific protein–RNA inhibitors, opening a new research avenue in the field.

On the mechanism of nucleosome recognition

A long-standing question in the field has been the structural basis of nucleosome binding and modification by LSD1-CoREST. The initial key finding was that CoREST is essential for nucleosome demethylation at H3-Lys4 [29]. Two recent papers have thoroughly addressed the issue by a combination of structural, biophysical and biochemical approaches [30••,63••]. A critical element has been the possibility to produce covalent enzyme-substrate complexes through nucleosome reconstitution with a semi-synthetic histone H3, carrying a propargyl moiety that covalently reacts with the flavin (Figure 2b). At the heart of the process is the C-terminal SANT domain of CoREST, which first binds with high-affinity but non-specifically to DNA (Figure 1). The fact that the length of extranucleosomal DNA strands proportionally enhances complex formation further supports the idea that the first step for the docking on LSD1-CoREST on the nucleosome is DNA binding [63••]. Next, LSD1-CoREST slides over the DNA, scanning different orientations over the nucleosome edge (Figure 3a, left panel). This process induces the detachment and displacement of the positively charged histone tail from the DNA. The histone tail can then be captured by the demethylase active site to attain the final catalytically productive docking of the enzyme heterodimer on the nucleosomal particle (Figure 3a, right panel) [30••]. Molecular dynamics simulations suggest that the elongated stalk of LSD1-CoREST likely explores different conformations during nucleosome binding and recognition, opening and closing akin to a flexible clamp, facilitating the process [64]. Of notice, it is a recent study indicating that two Lys residues on the midpoint of the tower domain surface are subject to acetylation by the MOF acetylase complex [65•]. Lys acetylation impairs the ability of LSD1-CoREST to demethylate the nucleosome. This is perfectly in line with the proposed model: LSD1-CoREST clamps the nucleosome with the helical stalk extending along and juxtaposed to the nucleosomal DNA (Figure 3a).

LSD1 versus LSD2: a matter of decoration

The roles of the non-catalytic domains are further highlighted by LSD2 (KDM1B), the second FAD-dependent histone lysine demethylase. Its reaction mechanism, substrate specificity (H3-Lys4), and catalytic domain structure are identical to those of LSD1. Nonetheless, it was clear since its discovery that LSD2 has distinct kinetic, structural, and functional properties [66–68]. Above all, LSD2 does not contain a tower domain, nor an interaction site for CoREST, but it rather features a zinc-finger domain [69•,70•] (Figure 3b). Differently from LSD1, this second flavoenzyme demethylase instead interacts with NPAC, a putative H3K36me3 reader with poorly-defined functions, which is suggested to augment the nucleosome demethylase activity of LSD2. Interestingly, structural studies have shown that a short peptide stretch of NPAC binds in proximity of the active site, enlarging the interaction surface with the H3 N-terminal residues for a tighter substrate binding [71••]. Thus, LSD1 and LSD2 perfectly outline the case of the same enzymatic module (an amine oxidase domain) decorated by different non-catalytic domains that impart distinct modes of nucleosome recognition and specific protein-protein interactions. It will be fascinating to uncover how these mechanistic and molecular distinctions lead to specific biological roles in epigenomic reprogramming. They will surely be exploited for the development of inhibitors that exert different pharmacological responses, despite targeting enzymes endowed with the same demethylase activities.

Open questions

- Is LSD1 able to demethylate other histone Lys residues and/or other non-histone substrates? Are these activities mediated by other demethylases directly or indirectly associated to LSD1?
- What is the fate of the formaldehyde product? Is folate involved in its scavenging?
- What are the non-catalytic functions of LSD1-CoREST? Does it function as a chromatin reader?
- What is the structure of the ternary HDAC-CoREST-LSD1 complex? Is there any channelling between the enzyme active sites?
- Will inhibitors targeting non-catalytic regions exert pharmacologically-relevant, distinct effects?

Conflict of interest

Nothing declared.

Acknowledgements

Research in the authors' laboratory is supported by AIRC (IG-15208) and the flagship MIUR-EPIGEN project.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y: **Histone demethylation mediated by the nuclear amine oxidase homolog LSD1**. *Cell* 2004, **119**:941-953.

2. • Walport LJ, Hopkinson RJ, Chowdhury R, Schiller R, Ge W, Kawamura A, Schofield CJ: **Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases.** *Nat Commun* 2016, **7**:11974.

Oxoglutarate-dependent demethylases are demonstrated to function also as Arg-demethylases, a ground-breaking observation in the field.

3. Walport LJ, Hopkinson RJ, Schofield CJ: **Mechanisms of human histone and nucleic acid demethylases**. *Curr Opin Chem Biol* 2012, **5–6**:525-534.

4. Forneris F, Binda C, Vanoni MA, Mattevi A, Battaglioli E: **Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process.** *FEBS Lett* 2005, **579**:2203-2207.

5. Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M *et al.*: **Disruption of histone methylation in developing sperm impairs offspring health transgenerationally**. *Science* 2015, **350** aab2006.

6. Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA: **Enhancer decommissioning by LSD1 during embryonic stem cell differentiation.** *Nature* 2012, **482**:221-225.

7. • Ancelin K, Syx L, Borensztein M, Ranisavljevic N, Vassilev I, Briseño-Roa L, Liu T, Metzger E, Servant N, Barillot E *et al.*: **Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation.** *Elife* 2016, **5** pii: e08851.

The back-to-back articles of Refs. [7•,8•] uncover the roles of maternally provided LSD1 in establishing, after fertilization, the epigenomic landscape that initiates development.

8. • Wasson JA, Simon AK, Myrick DA, Wolf G, Driscoll S, Pfaff SL, Macfarlan TS, Katz DJ: **Maternally provided LSD1/KDM1A enables the maternal-to-zygotic transition and prevents defects that manifest postnatally.** *Elife* 2016, **5** pii: e08848.

See annotation to Ref. [7•].

9. Thambyrajah R, Mazan M, Patel R, Moignard V, Stefanska M, Marinopoulou E, Li Y, Lancrin C, Clapes T, Möröy T *et al.*: **GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1.** *Nat Cell Biol* 2016, **18**:21-32.

10. Di Stefano B, Collombet S, Jakobsen JS, Wierer M, Sardina JL, Lackner A, Stadhouders R, Segura-Morales C, Francesconi M, Limone F *et al.*: **C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of LSD1 and BRD4.** *Nat Cell Biol.* 2016, **18**:371-381.

11. Hino S, Sakamoto A, Nagaoka K, Anan K, Wang Y, Mimasu S, Umehara T, Yokoyama S, Kosai K, Nakao M: **FAD-dependent lysine-specific demethylase-1 regulates cellular energy expenditure.** *Nat Commun* 2012, **3**:758.

12. Fuentes P, Cánovas J, Berndt FA, Noctor SC, Kukuljan M: **CoREST/LSD1 control the development of pyramidal cortical neurons.** *Cereb Cortex* 2012, **6**:1431-1441.

13. Harris WJ, Huang X, Lynch JT, Spencer GJ, Hitchin JR, Li Y, Ciceri F, Blaser JG, Greystoke BF, Jordan AM *et al.*: **The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells.** *Cancer Cell* 2012, **21**:473-487.

14. Hill JM, Quenelle DC, Cardin RD, Vogel JL, Clement C, Bravo FJ, Foster TP, Bosch-Marce M, Raja P, Lee JS *et al.*: **Inhibition of LSD1 reduces herpesvirus infection, shedding, and recurrence by promoting epigenetic suppression of viral genomes.** *Sci Transl Med* 2014, **6**:265ra169.

15. • Tunovic S, Barkovich J, Sherr EH, Slavotinek AM: **De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome.** *Am J Med Genet A* 2014, **164A**:1744-1749.

The discovery of a novel genetic disease caused by mutations in LSD1-coding gene.

16. Adamo A, Atashpaz S, Germain PL, Zanella M, D'Agostino G, Albertin V, Chenoweth J, Micale L, Fusco C, Unger C *et al.*: **7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages.** *Nat Genet* 2015, **47**:132-141.

17. Yang P, Wang Y, Chen J, Li H, Kang L, Zhang Y, Chen S, Zhu B, Gao S: **RCOR2 is a subunit of the LSD1 complex that regulates ESC property and substitutes for SOX2 in reprogramming somatic cells to pluripotency.** *Stem Cells* 2011, **29**:791-801.

18. Upadhyay G, Chowdhury AH, Vaidyanathan B, Kim D, Saleque S: **Antagonistic actions of Rcor3 proteins regulate LSD1 activity and cellular differentiation.** *Proc Natl Acad Sci U S A* 2014, **111**:8071-8076.

19. Barrios ÁP, Gómez AV, Sáez JE, Ciossani G, Toffolo E, Battaglioli E, Mattevi A, Andrés ME: **Differential properties of transcriptional complexes formed by the CoREST family.** *Mol Cell Biol* 2014, **34**:2760-2770.

20. • Wang Y, Wu Q, Yang P, Wang C, Liu J, Ding W, Liu W, Bai Y, Yang Y, Wang H, Gao S, Wang X: **LSD1 co-repressor Rcor2 orchestrates neurogenesis in the developing mouse brain.** *Nat Commun* 2016, **7**:10481.

The role of CoREST2 in neurogenesis outlines the notion that each CoREST isoform is selectively involved in developmental programs.

21. Binda C, Mattevi A, Edmondson DE: **Structure–function relationships in flavoenzyme-dependent amine oxidations: a comparison of polyamine oxidase and monoamine oxidase.** *J Biol Chem* 2002, **277**:23973-23976.

22. • Karasulu B, Patil M, Thiel W: **Amine oxidation mediated by lysine-specific demethylase 1: quantum mechanics/molecular mechanics insights into mechanism and role of lysine 661.** *J Am Chem Soc* 2013, **135**:13400-13413.

A thorough computational study of the mechanism of amine oxidation.

23. Forneris F, Binda C, Vanoni MA, Battaglioli E, Mattevi A: **Human histone demethylase LSD1 reads the histone code.** *J Biol Chem* 2005, **280**:41360-41365.

24. • Garcia BA, Luka Z, Loukachevitch LV, Bhanu NV, Wagner C: **Folate deficiency affects histone methylation.** *Med Hypotheses* 2016, **88**:63-67.

Is formaldehyde a simple reaction byproduct? How is it scavenged? These issues are discussed in this perspective article.

25. Luka Z, Pakhomova S, Loukachevitch LV, Calcutt MW, Newcomer ME, Wagner C: **Crystal structure of the histone lysine specific demethylase LSD1 complexed with tetrahydrofolate.** *Protein Sci* 2014, **23**:993-998.

26. Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, Otwinowski Z, Yu H: **Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase.** *Mol Cell* 2006, **23**:377-387.

27. Peng B, Wang J, Hu Y, Zhao H, Hou W, Zhao H, Wang H, Liao J, Xu X: **Modulation of LSD1 phosphorylation by CK2/WIP1 regulates RNF168-dependent 53BP1 recruitment in response to DNA damage.** *Nucleic Acids Res* 2015, **43**:5936-5947.

28. •• Metzger E, Willmann D, McMillan J, Forne I, Metzger P, Gerhardt S, Petroll K, von Maessenhausen A, Urban S, Schott AK *et al.*: **Assembly of methylated KDM1A and CHD1 drives androgen receptor-dependent transcription and translocation.** *Nat Struct Mol Biol* 2016, **23**:132-139.

A fascinating discovery: the N-terminal disordered residues of LSD1 are subject to methylation. Methylated LSD1 is 'read' by CHD1 as part of androgen-dependent transcriptional programs.

29. Lee MG, Wynder C, Cooch N, Shiekhattar R: **An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation.** *Nature* 2005, **437**:432-435.

30. •• Pilotto S, Speranzini V, Tortorici M, Durand D, Fish A, Valente S, Forneris F, Mai A, Sixma TK, Vachette P, Mattevi A: **Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation.** *Proc Natl Acad Sci U S A* 2015, **112**:2752-2757.

This article uncovers the mechanism for nucleosome recognition by LSD1, taking advantage of semisynthetic nucleosomes as highly effective experimental tool.

31. Hwang S, Schmitt AA, Luteran AE, Toone EJ, McCafferty DG: **Thermodynamic characterization of the binding interaction between the histone demethylaseLSD1/KDM1 and CoREST.** *Biochemistry* 2011, **50**:546-557.

32. Forneris F, Binda C, Adamo A, Battaglioli E, Mattevi A: **Structural basis of LSD1-CoREST selectivity in histone H3 recognition.** *J Biol Chem* 2007, **282**:20070-20074.

33. • Burg JM, Gonzalez JJ, Maksimchuk KR, McCafferty DG: **Lysine-specific demethylase 1A (KDM1A/LSD1): product recognition and kinetic analysis of full-length histones.** *Biochemistry* 2016, **55**:1652-1662.

An interesting extension of previous biochemical and kinetic studies. The enzymatic activity is probed using full-length histone H3, which exhibits tighter binding and activity compared to N-terminal H3 peptides normally used in enzymatic assays.

34. Yang Z, Jiang J, Stewart MD, Qi S, Yamane K, Li J, Zhang Y, Wong J: **AOF1 is a histone H3K4 demethylase possessing demethylase activity-independent repression function.** *Cell Res* 2010, **20**:276-287.

35. Lee MG, Wynder C, Schmidt DM, McCafferty DG, Shiekhattar R: **Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications.** *Chem Biol* 2006, **13**:563-567.

36. Shi YJ, Matson C, Lan F, Iwase S, Baba T, Shi Y: **Regulation of LSD1 histone demethylase activity by its associated factors.** *Mol Cell* 2005, **19**:857-864.

37. Neelamegam R, Ricq EL, Malvaez M, Patnaik D, Norton S, Carlin SM, Hill IT, Wood MA, Haggarty SJ, Hooker JM: **Brain-penetrant LSD1 inhibitors can block memory consolidation.** *ACS Chem Neurosci* 2012, **3**:120-128.

38. • Mohammad HP, Smitheman KN, Kamat CD, Soong D, Federowicz KE, Van Aller GS, Schneck JL, Carson JD, Liu Y, Butticello M *et al.*: **A DNA hypomethylation signature predicts antitumor activity of LSD1 inhibitors in SCLC.** *Cancer Cell* 2015, **28**:57-69.

LSD1 is shown to be a potential pharmacological target for treatment of lung cancer in presence of specific DNA methylation patterns.

39. Stazi G, Zwergel C, Valente S, Mai A: **LSD1 inhibitors: a patent review (2010–2015).** *Expert Opin Ther Pat* 2016, **26**:565-580.

40. Yang M, Culhane JC, Szewczuk LM, Jalili P, Ball HL, Machius M, Cole PA, Yu H: **Structural basis for the inhibition of the LSD1 histone demethylase by the antidepressant trans-2-phenylcyclopropylamine.** *Biochemistry* 2007, **46**:8058-8065.

41. Binda C, Valente S, Romanenghi M, Pilotto S, Cirilli R, Karytinos A, Ciossani G, Botrugno OA, Forneris F, Tardugno M *et al.*: **Biochemical, structural, and biological evaluation of tranylcypromine derivatives as inhibitors of histone demethylases LSD1 and LSD2.** *J Am Chem Soc* 2010, **132**:6827-6833.

42. Schenk T, Chen WC, Göllner S, Howell L, Jin L, Hebestreit K, Klein HU, Popescu AC, Burnett A, Mills K *et al.*: **Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia.** *Nat Med* 2012, **18**:605-611.

43. McGrath JP, Williamson KE, Balasubramanian S, Odate S, Arora S, Hatton C, Edwards TM, O'Brien T, Magnuson S, Stokoe D *et al.*: **Pharmacological inhibition of the histone lysine demethylase KDM1A suppresses the growth of multiple acute myeloid leukemia subtypes.** *Cancer Res* 2016, **76**:1975-1988.

44. Culhane JC, Wang D, Yen PM, Cole PA: **Comparative analysis of small molecules and histone substrate analogues as LSD1 lysine demethylase inhibitors.** *J Am Chem Soc* 2010, **132**:3164-3176.

45. Dancy BC, Ming SA, Papazyan R, Jelinek CA, Majumdar A, Sun Y, Dancy BM, Drury WJ 3rd, Cotter RJ, Taverna SD, Cole PA: **Azalysine analogues as probes for protein lysine deacetylation and demethylation.** *J Am Chem Soc* 2012, **134**:5138-5148.

46. Ogasawara D, Itoh Y, Tsumoto H, Kakizawa T, Mino K, Fukuhara K, Nakagawa H, Hasegawa M, Sasaki R, Mizukami T, Miyata N, Suzuki T: **Lysine-specific demethylase 1-selective inactivators: protein-targeted drug delivery mechanism.** *Angew Chem Int Ed Engl* 2013, **52**:8620-8624.

47. Itoh Y, Aihara K, Mellini P, Tojo T, Ota Y, Tsumoto H, Solomon VR, Zhan P, Suzuki M, Ogasawara D *et al.*: **Identification of SNAIL1 peptide-based irreversible lysine-specific demethylase 1-selective inactivators.** *J Med Chem* 2016, **59**:1531-1544.

48. • Rotili D, Tomassi S, Conte M, Benedetti R, Tortorici M, Ciossani G, Valente S, Marrocco B, Labella D, Novellino E *et al.*: **Pan-histone demethylase inhibitors simultaneously targeting Jumonji C and lysine-specific demethylases display high anticancer activities.** *J Med Chem* 2014, **57**:42-55.

The design, synthesis, and evaluation of innovative inhibitors that target both flavin-dependent and oxoglutarate-dependent histone demethylases.

49. Robertson JC, Hurley NC, Tortorici M, Ciossani G, Borrello MT, Vellore NA, Ganesan A, Mattevi A, Baron R: **Expanding the druggable space of the LSD1/CoREST epigenetic target: new potential binding regions for drug-like molecules, peptides, protein partners, and chromatin.** *PLoS Comput Biol* 2013, **9**:e1003158.

50. Ma LY, Zheng YC, Wang SQ, Wang B, Wang ZR, Pang LP, Zhang M, Wang JW, Ding L, Li J *et al.*: **Design, synthesis, and structure–activity relationship of novel LSD1 inhibitors based on pyrimidine–thiourea hybrids as potent, orally active antitumor agents.** *J Med Chem* 2015, **58**:1705-1716.

51. Lin Y, Wu Y, Li J, Dong C, Ye X, Chi YI, Evers BM, Zhou BP: **The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1.** *EMBO J* 2010, **29**:1803-1816.

52. Welcker JE, Hernandez-Miranda LR, Paul FE, Jia S, Ivanov A, Selbach M, Birchmeier C: **Insm1 controls development of pituitary endocrine cells and requires a SNAG domain for function and for recruitment of histone-modifying factors.** *Development* 2013, **140**:4947-4958.

53. Baron R, Binda C, Tortorici M, McCammon JA, Mattevi A: **Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1–CoREST complex.** *Structure* 2011, **19**:212-220.

54. Tortorici M, Borrello MT, Tardugno M, Chiarelli LR, Pilotto S, Ciossani G, Vellore NA, Bailey SG, Cowan J, O'Connell M *et al.*: **Protein recognition by short peptide reversible inhibitors of the chromatin-modifying LSD1/CoREST lysine demethylase.** *ACS Chem Biol* 2013, **8**:1677-1682.

55. Zibetti C, Adamo A, Binda C, Forneris F, Toffolo E, Verpelli C, Ginelli E, Mattevi A, Sala C, Battaglioli E: **Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system.** *J Neurosci* 2010, **30**:2521-2532.

56. Toffolo E, Rusconi F, Paganini L, Tortorici M, Pilotto S, Heise C, Verpelli C, Tedeschi G, Maffioli E, Sala C, Mattevi A, Battaglioli E: **Phosphorylation of neuronal lysine-specific demethylase 1 LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2.** *J Neurochem* 2014, **128**:603-616.

57. • Wang J, Telese F, Tan Y, Li W, Jin C, He X, Basnet H, Ma Q, Merkurjev D, Zhu X *et al.*: **LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control.** *Nat Neurosci* 2015, **18**:1256-1264.

The manuscripts [57•,58•,59•] collectively demonstrate the roles of the LSD1 neurospecific variant in neuronal differentiation and animal behavior.

58. • Laurent B, Ruitu L, Murn J, Hempel K, Ferrao R, Xiang Y, Liu S, Garcia BA, Wu H, Wu F, Steen H, Shi Y: **A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation.** *Mol Cell* 2015, **57**:957-970.

See annotation to Ref. [57•].

59. • Rusconi F, Grillo B, Ponzoni L, Bassani S, Toffolo E, Paganini L, Mallei A, Braida D, Passafaro M, Popoli M, Sala M, Battaglioli E: **LSD1 modulates stress-evoked transcription of immediate early genes and emotional behavior.** *Proc Natl Acad Sci U S A* 2016, **113**:3651-3656.

See annotation to Ref. [57•].

60. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY: **Long noncoding RNA as modular scaffold of histone modification complexes.** *Science* 2010, **329**:689-693.

61. • Porro A, Feuerhahn S, Lingner J: **TERRA-reinforced association of LSD1 with MRE11 promotes processing of uncapped telomeres.** *Cell Rep* 2014, **6**:765-776.

The discovery that LSD1 associates to non-coding RNAs, expanding the pattern of molecular interactions involving the enzyme.

62. •• Hirschi A, Martin WJ, Luka Z, Loukachevitch LV, Reiter NJ: **G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme.** *RNA* 2016. [Epub ahead of print].

The structure of LSD1/CoREST bound to a telomeric RNA fragment is described. The RNA binds in a non-catalytic site, raising several interesting questions about its roles.

63. •• Kim SA, Chatterjee N, Jennings MJ, Bartholomew B, Tan S: **Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex.** *Nucleic Acids Res* 2015, **43**:4868-4880.

An in-depth and methodologically advanced investigation of nucleosome recognition by LSD1/CoREST, demonstrating the role of linker DNA in the process.

64. Baron R, Vellore NA: **LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition.** *Proc Natl Acad Sci U S A* 2012, **109**:12509-12514.

65. • Luo H, Shenoy AK, Li X, Jin Y, Jin L, Cai Q, Tang M, Liu Y, Chen H, Reisman D *et al.*: **MOF acetylates the histone demethylase LSD1 to suppress epithelial-to-mesenchymal transition.** *Cell Rep* 2016, **15**:2665-2678.

Lys residues of the tower domain are subject to acetylation, which impairs binding and demethylation of the nucleosome.

66. Karytinos A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, Mattevi A: **A novel mammalian flavin-dependent histone demethylase.** *J Biol Chem* 2009, **284**:17775-17782.

67. Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E, Chen T: **KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints**. *Nature* 2009, **461**:415-418.

68. Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C *et al.*: **Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation**. *Mol Cell* 2010, **39**:222-233.

69. • Chen F, Yang H, Dong Z, Fang J, Wang P, Zhu T, Gong W, Fang R, Shi YG, Li Z, Xu Y: **Structural insight into substrate recognition by histone demethylase LSD2/KDM1b**. *Cell Res* 2013, **23**:306-309.
The structural basis of histone tail recognition and demethylation by LSD2 highlights similarities and differences compared to LSD1.

70. • Zhang Q, Qi S, Xu M, Yu L, Tao Y, Deng Z, Wu W, Li J, Chen Z, Wong J: **Structure–function analysis reveals a novel mechanism for regulation of histone demethylase LSD2/AOF1/KDM1b**. *Cell Res* 2013, **23**:225-241.
The description of the three-dimensional crystal structure of LSD2.

71. •• Fang R, Chen F, Dong Z, Hu D, Barbera AJ, Clark EA, Fang J, Yang Y, Mei P, Rutenberg M *et al.*: **LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation**. *Mol Cell* 2013, **49**:558-570.
A milestone study, which unravels how LSD2 associates to and is activated by NPAC, a nuclear protein with poorly characterized functions.

References

Alberts, B. *et al.* (2002) *Molecular Biology of the Cell . Fourth Edition., New York: Garland Science.* doi: 10.1086/377840.

Barski, A. *et al.* (2007) ‘High-Resolution Profiling of Histone Methylations in the Human Genome’, *Cell*, 129, pp. 823–837. doi: 10.1016/j.cell.2007.05.009.

Bednar, J. *et al.* (2017) ‘Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1.’, *Molecular cell*, 66(3), pp. 384-397.e8. doi: 10.1016/j.molcel.2017.04.012.

Bilokapic, S., Strauss, M. and Halic, M. (2018a) ‘Histone octamer rearranges to adapt to DNA unwrapping’, *Nature Structural and Molecular Biology*. Springer US, 25(1), pp. 101–108. doi: 10.1038/s41594-017-0005-5.

Bilokapic, S., Strauss, M. and Halic, M. (2018b) ‘Structural rearrangements of the histone octamer translocate DNA’, *Nature Communications*, pp. 1–11. doi: 10.1038/s41467-018-03677-z.

Bintu, B. *et al.* (2018) ‘Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells’, *Science*. doi: 10.1126/science.aau1783.

Chen, F. *et al.* (2013) ‘Structural insight into substrate recognition by histone demethylase LSD2/KDM1b’, *Cell Research*, 23(2), pp. 306–309. doi: 10.1038/cr.2013.17.

Cirillo, L. A. and Zaret, K. S. (2007) ‘Specific Interactions of the Wing Domains of FOXA1 Transcription Factor with DNA’, *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2006.11.087.

Clapier, C. R. *et al.* (2017) ‘Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes’, *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm.2017.26.

Clinton, B. (2000) *President clinton: Announcing the Completion of the First Survey of the Entire Human Genome Project*. Available at: https://clintonwhitehouse3.archives.gov/WH/Work/062600.html.

Collings, C. K., Waddell, P. J. and Anderson, J. N. (2013) ‘Effects of DNA methylation on nucleosome stability’, *Nucleic Acids Research*. doi: 10.1093/nar/gks893.

Cooper, G. M. (2000) *The cell : a molecular approach*. ASM Press.

Crick, F. H. D. and Watson, J. D. (1953) ‘A Structure for Deoxyribose Nucleic Acid’, *Nature*,. doi: 10.1038/171737a0.

Dechassa, M. L. *et al.* (2011) ‘Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes.’, *Nature communications*, 2, p. 313. doi: 10.1038/ncomms1320.

Dixon, J. R., Gorkin, D. U. and Ren, B. (2016) ‘Chromatin Domains: The Unit of Chromosome Organization’, *Molecular Cell*. doi: 10.1016/j.molcel.2016.05.018.

Eidahl, J. O. *et al.* (2013) ‘Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes’, *Nucleic Acids Research*. doi: 10.1093/nar/gkt074.

Fang, R. *et al.* (2013) ‘LSD2/KDM1B and Its Cofactor NPAC/GLYR1 Endow a Structural and Molecular Model for Regulation of H3K4 Demethylation’, *Molecular Cell*, 49(3), pp. 558–570. doi: 10.1016/j.molcel.2012.11.019.

Farnung, L., Vos, S. M. and Cramer, P. (2018) ‘Structure of transcribing RNA polymerase II-nucleosome complex’, *Nature Communications*. doi: 10.1038/s41467-018-07870-y.

Fei, J. *et al.* (2018) ‘NDF, a nucleosome-destabilizing factor that facilitates transcription through nucleosomes’, *Genes and Development*, 32, pp. 1–13. doi: 10.1101/gad.313973.118.

Feil, R. (2009) ‘Epigenetics: Ready for the marks’, *Nature*. doi: 10.1038/461359a.

Finch, J. T. *et al.* (1977) ‘Structure of nucleosome core particles of chromatin.’, *Nature*, 269(5623), pp. 29–36. doi: 10.1038/269029a0.

Fu, I. *et al.* (2017) ‘Nucleosome Histone Tail Conformation and Dynamics: Impacts of Lysine Acetylation and a Nearby Minor Groove Benzo[a]pyrene-Derived Lesion’, *Biochemistry*. doi: 10.1021/acs.biochem.6b01208.

Gamarra, N. *et al.* (2018) ‘The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h’. doi: 10.7554/eLife.35322.001.

Gurdon, J. B. (1962) ‘The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles’, *Development*, 10(4).

Hendrich, B. and Bird, A. (1998) ‘Identification and characterization of a family of mammalian methyl-CpG binding proteins.’, *Molecular and cellular biology*, 18(11), pp. 6538–47. doi: 10.1128/mcb.18.11.6538.

Hirschhorn, J. N. *et al.* (1992) ‘Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure.’, *Genes & development*, 6(12A), pp. 2288–98. doi: 10.1101/gad.6.12a.2288.

Højfeldt, J. W., Agger, K. and Helin, K. (2013) ‘Histone lysine demethylases as targets for anticancer therapy’, *Nature Reviews Drug Discovery*. doi: 10.1038/nrd4154.

Iwafuchi-Doi, M. *et al.* (2016) ‘The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation’, *Molecular Cell*. doi: 10.1016/j.molcel.2016.03.001.

Iwasaki, W. *et al.* (2013) ‘Contribution of histone N-terminal tails to the structure and stability of nucleosomes’, *FEBS Open Bio*. doi: 10.1016/j.fob.2013.08.007.

Jablonka, E. and Lamb, M. J. (1998) ‘Bridges between Development and Evolution’, *Biology and Philosophy*. doi: 10.1023/A:1006520407209.

Kalashnikova, A. A. *et al.* (2013) ‘The role of the nucleosome acidic patch in modulating higher order chromatin structure’, *Journal of the Royal Society Interface*. doi: 10.1098/rsif.2012.1022.

Kim, S. *et al.* (2015) ‘Extranucleosomal DNA enhances the activity of the LSD1 / CoREST histone demethylase complex’, *Nucleic Acids Research*, 43(10), pp. 4868–4880. doi: 10.1093/nar/gkv388.

Kornberg, R. D. (1974) ‘Chromatin structure: a repeating unit of histones and DNA.’, *Science (New York, N.Y.)*, 184(4139), pp. 868–71. doi: 10.1126/science.184.4139.868.

Kouzarides, T. (2007) 'Chromatin Modifications and Their Function', *Cell*. doi: 10.1016/j.cell.2007.02.005.

Kruger, W. *et al.* (1995) 'Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription.', *Genes & development*, 9(22), pp. 2770–9. doi: 10.1101/gad.9.22.2770.

Kujirai, T. *et al.* (2018) 'Structural basis of the nucleosome transition during RNA polymerase II passage.', *Science (New York, N.Y.)*, 362(6414), pp. 595–598. doi: 10.1126/science.aau9904.

Li, G. *et al.* (2005) 'Rapid spontaneous accessibility of nucleosomal DNA', *Nature Structural and Molecular Biology*, 12(1), pp. 46–53. doi: 10.1038/nsmb869.

Li, Meijing *et al.* (2019) 'Mechanism of DNA translocation underlying chromatin remodelling by Snf2', *Nature*, 567(7748), pp. 409–413. doi: 10.1038/s41586-019-1029-2.

Lind, M. I. and Spagopoulou, F. (2018) 'Evolutionary consequences of epigenetic inheritance', *Heredity*. doi: 10.1038/s41437-018-0113-y.

Lowary, P. T. and Widom, J. (1998) 'New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-directed Nucleosome Positioning', *Journal of Molecular Biology*, 276, pp. 19–42.

Luger, K. *et al.* (1997) 'Crystal structureof thenucleosome core particle at 2.8A° resolution', *Nature*·, 389(6648), pp. 251–260. doi: 10.1107/S0907444900011847.

Makde, R. D. *et al.* (2010) 'Structure of RCC1 chromatin factor bound to the nucleosome core particle', *Nature*. doi: 10.1038/nature09321.

Marabelli, C. *et al.* (2019) 'A Tail-Based Mechanism Drives Nucleosome Demethylation by the LSD2/NPAC Multimeric Complex', *Cell Reports*, 27, pp. 387–399. doi: 10.1016/j.celrep.2019.03.061.

Marabelli, C., Marrocco, B. and Mattevi, A. (2016) 'The growing structural and functional complexity of the LSD1/KDM1A histone demethylase', *Current Opinion in Structural Biology*. Elsevier Ltd, 41, pp. 135–144. doi: 10.1016/j.sbi.2016.07.011.

Marmorstein, R. and Zhou, M. M. (2014) ‘Writers and readers of histone acetylation: Structure, mechanism, and inhibition’, *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a018762.

Maskell, D. P. *et al.* (2015) ‘Structural basis for retroviral integration into nucleosomes.’, *Nature*, 523(7560), pp. 366–9. doi: 10.1038/nature14495.

McGinty, R. K., Henrici, R. C. and Tan, S. (2014) ‘Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome.’, *Nature*, 514(7524), pp. 591–6. doi: 10.1038/nature13890.

McGinty, R. K. and Tan, S. (2015) ‘Nucleosome structure and function’, *Chemical Reviews*, pp. 2255–2273. doi: 10.1021/cr500373h.

Milner, G. R. and Hayhoe, F. G. J. (1968) ‘Ultrastructural localization of nucleic acid synthesis in human blood cells [32]’, *Nature*. doi: 10.1038/218785a0.

Morrison, Emma A. *et al.* (2018) ‘The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome’, *eLife*. doi: 10.7554/eLife.31481.

Morrison, Emma A *et al.* (2018) ‘The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome’, *eLife*, 7, p. e31481. doi: 10.7554/eLife.31481.001.

Mosammaparast, N. and Shi, Y. (2010) ‘Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases.’, *Annual review of biochemistry*, 79, pp. 155–79. doi: 10.1146/annurev.biochem.78.070907.103946.

Narlikar, G. J., Sundaramoorthy, R. and Owen-Hughes, T. (2013) ‘Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes’, *Cell*. doi: 10.1016/j.cell.2013.07.011.

Noble, I. (2003) ‘Human genome finally complete’, *BBC News*. Available at: http://news.bbc.co.uk/2/hi/science/nature/2940601.stm.

North, J. A. *et al.* (2012) ‘Regulation of the nucleosome unwrapping rate controls DNA accessibility’, *Nucleic Acids Research*. doi: 10.1093/nar/gks747.

Van Nuland, R. *et al.* (2013) ‘Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain’, *Epigenetics and Chromatin*. Epigenetics & Chromatin, 6:12(1), p. 1. doi: 10.1186/1756-8935-6-12.

Oudet, P., Gross-Bellard, M. and Chambon, P. (1975) ‘Electron microscopic and biochemical evidence that chromatin structure is a repeating unit’, *Cell*. doi: 10.1016/0092-8674(75)90149-X.

Patel, D. J. (2016) ‘A structural perspective on readout of epigenetic histone and DNA methylation marks’, *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a018754.

Pilotto, S. *et al.* (2015) ‘Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation’, *Proceedings of the National Academy of Sciences*, 112(9), pp. 2752–2757. doi: 10.1155/2014/349065.

Poirier, M. G. *et al.* (2009) ‘Dynamics and function of compact nucleosome arrays’, *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.1650.

Richmond, T. J. *et al.* (1984) ‘Structure of the nucleosome core particle at 7 resolution’, *Nature*. doi: 10.1038/311532a0.

Ricketts, M. D. *et al.* (2019) ‘Molecular basis for chromatin assembly and modification by multi-protein complexes’, *Protein Science*, 2, pp. 329–343. doi: 10.1002/pro.3535.

Rondelet, G. *et al.* (2016) ‘Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B’, *Journal of Structural Biology*. doi: 10.1016/j.jsb.2016.03.013.

Savitsky, P. *et al.* (2016) ‘Multivalent Histone and DNA Engagement by a PHD/BRD/PWWP Triple Reader Cassette Recruits ZMYND8 to K14ac-Rich Chromatin’, *Cell Reports*. doi: 10.1016/j.celrep.2016.11.014.

Suzuki, M. M. and Bird, A. (2008) ‘DNA methylation landscapes: provocative insights from epigenomics’, *Nature Reviews Genetics*, 9, pp. 465–476. doi: 10.1038/nrg2341.

Tachiwana, H. *et al.* (2011) ‘Crystal structure of the human centromeric nucleosome containing CENP-A.’, *Nature*, 476(7359), pp. 232–5. doi: 10.1038/nature10258.

Taguchi, H. *et al.* (2014) ‘A method for evaluating nucleosome stability with a protein-binding fluorescent dye’, *Methods*. doi: 10.1016/j.ymeth.2014.08.019.

Wagner, E. J. and Carpenter, P. B. (2012) ‘Understanding the language of Lys36 methylation at histone H3’, *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm3274.

Weaver, T. M., Morrison, E. A. and Musselman, C. A. (2018) ‘Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains’, *Molecules,* 23(2614). doi: 10.3390/molecules23102614.

Wen, H. *et al.* (2014) ‘ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression’, *Nature,* 508, pp. 263–268. doi: 10.1038/nature13045.

Wilmut, I. *et al.* (1996) ‘Sheep cloned by nuclear transfer from a cultured cell line.’, *Nature*. doi: 10.1038/380064a0.

Wilson, M. D. *et al.* (2016) ‘The structural basis of modified nucleosome recognition by 53BP1’, *Nature*. doi: 10.1038/nature18951.

Woodcock, C. (2010) ‘Gallus gallus, nucleate erythrocyte.’, *Cell Image Library 709*. Available at: https://doi.org/doi:10.7295/W9CIL709.

Zaret, K. S., Lerner, J. and Iwafuchi-Doi, M. (2016) ‘Chromatin Scanning by Dynamic Binding of Pioneer Factors’, *Molecular Cell.* doi: 10.1016/j.molcel.2016.05.024.

Zhao, Y. and Garcia, B. A. (2015) ‘Comprehensive catalog of currently documented histone modifications’, *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a025064.

Zhou, K., Gaullier, G. and Luger, K. (2018) ‘Nucleosome structure and dynamics are coming of age’, *Nature Structural & Molecular Biology*. Springer US, 26, pp. 3–13. doi: 10.1038/s41594-018-0166-x.

Chapter II – Biological and pathological effects of LSD1

LSD1 is the first discovered of the only two flavin-dependent histone demethylases. Both LSD1 and LSD2 perform demethylation of H3k4me1/me2 nucleosomes. However, LSD1 is the most pharmacologically relevant of the two, as the many studies and researches performed till now have evidenced its role in more than thirty cancer-related processes. Various LSD1-targeting inhibitors are nowadays in phase I and II clinical trials all around the world for the treatment of hematopoietic malignancies in particular.

Structural and biochemical characterization of three pathological variants of LSD1 opens new perspectives on the biological role of the demethylase, whose function is not only restricted to the catalysis of H3K4me1/2 demethylation, but also to its recruitment of a wide network of chromatin regulators and transcription factors, such as p53 and SNAIL1.

Background

1. LSD1 substrate-binding pocket: a docking point for multiple interactors

The active site of LSD1 is very large and rich in invaginations. Residues exposed within the cavity are also extremely conserved among vertebrates. They establish highly specific H-bonds and salt bridges with the first 21 residues of the substrate histone tail, promoting its folding inside the active site. The effect of PTMs on H3 substrate recognition by LSD1 have already been extensively studied (Forneris, Binda, Vanoni, Battaglioli, *et al.*, 2005; Forneris *et al*., 2007). LSD1 preference for H3 tail stripped of all covalent marks, suggests that other histone modifying "eraser" enzymes, as phosphatases, demethylases and deacetylases, may precede LSD1 activity in order to make the tail suitable for H3K4me1/me2 demethylation (Forneris, Binda, Vanoni, Mattevi, *et al*., 2005). LSD1 can bind with the same affinity histone tails with different K4 methylation states, including the tri-methylated Lys 4 that is not substrate for LSD1. The intrinsic high affinity of LSD1 for a long stretch on the histone H3 tail suggests also a non-catalytic role for this enzyme, which in some contexts may act as simply an epigenetic "reader".

Moreover, LSD1 activity is tightly regulated: not only the epigenetic state of the substrate H3 tail, but also through the enzyme recruitment within different complexes. LSD1 has indeed many interactors and regulators able to bind it at various and separate surface motifs (Marabelli *et al*., 2016).. Interestingly, LSD1 interactors are known which bind to the enzyme catalytic pocket through a histone tail mimicry strategy (Figure 1). Zinc-finger transcription factors of the Snail1 family recruit the complex through binding of LSD1 substrate cavity. In fact, the N-terminal SNAG domain in Snail family mimics histone H3 tail. ((Tortorici *et al.*, 2013)). For example, Gfi-1 (Growth Factor independence 1) and Gfi-1b use this mechanism to drive LSD1, along with CoREST1 and HDAC, to target specific gene promoters in hematopoietic stem cells. Here, the histone demethylase represses hematopoietic stem and progenitor cell signatures during blood cell maturation (Lin *et al.*, 2010). During Epithelial-to-Mesenchymal Transition (EMT), SNAIL1 transcription factor carries the LSD1/CoREST1/HDAC repressor complex to particular gene targets thanks to this particular mechanism of mimicry (Lin y *et al.*, 2010).

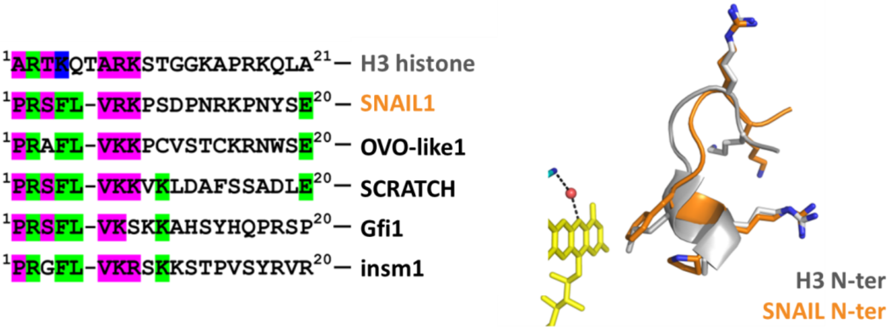


Figure 1. On the left, comparison of the amino acid sequences of the N-terminal segments of the SNAIL family of transcription factors. Residues mimicking histone H3 are evidenced. On the right, superposition of the crystallographic structures of the N-terminal peptides of H3 (grey) and SNAIL1 (orange), which assume identical conformations within LSD1 substrate-binding cavity. Image adapted from Tortorici *et al.*, 2013.

2. Biomedical relevance of LSD1

A variety of physiological and pathological processes have been ascribed to LSD1. Its importance is such that gene knockout causes early embryonic lethality in mice (Wang *et al.*, 2005). LSD1 is implicated in tumor formation and progression: it is overexpressed in acute myeloid leukemia and in solid tumors such as neuroblastoma, prostate, colon and breast cancer (Schenk *et al.*, 2012). Downregulation of LSD1 results in a repression of proliferation, migration and invasive potential in different cancer cells (Zheng *et al.*, 2015). The link between LSD1 and cancer is highlighted also by the fact that a variety of tumors present loss of H3K4 methylation and enrichment of H3K9 methylation (Wang *et al.*, 2009).

The recent development of fast and accurate gene sequencing methods allowed the identification of three dominant missense point mutations of LSD1: c.1207G>A, c.1739A>G and c.2353T>C, encoding for Glu379Lys, Asp556Gly and Tyr761His respectively (Tunovic *et al.*, 2014). In particular, the Tyr761His mutation affects a key tyrosine of the LSD1 substrate-binding cavity, which helps in positioning of H3K4me1/2 in front of the catalytic FAD ring; whereas Glu379 and Asp556 organize non-substrate residues of the H3 tail (Figure 2).

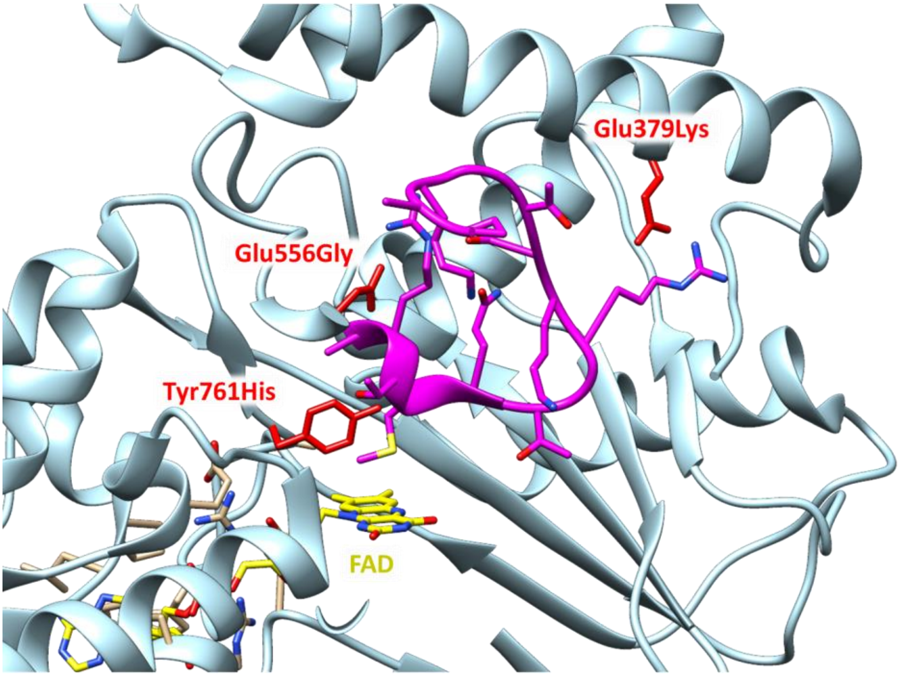


Figure 2. Zoomed view of the H3(1-21) (purple) substrate within LSD1 catalytic pocket (PDB 2V1D). LSD1 mutant residues are highlighted in red, whereas LSD1 ribbon structure is shown in light blue. The substrate residue at position 4 of the h3 tail sits exactly in front of the flavin cofactor (yellow).

Results and discussion

1. LSD1 pathological variants effect on substrate binding and catalysis

LSD1 mutants Glu379Lys, Asp556Gly and Tyr761His were expressed and co-purified with CoREST1 according to the standard protocol without significant changes in the yield with respect to LSD1 wild-type. Also, all three proteins were crystallized in the standard LSD1 wild-type conditions, with crystals diffracting at 2.6 to 3.3 Å resolution (PDB IDs are 5L3B: LSD1 D556G, 5L3C: LSD1 E379K, 5L3D: LSD1 Y761H). No overall structural rearrangements were observed in any case. On the other hand, the catalytic properties of the pathological mutants were at least 10/20 times less efficient compared with the wild-type enzyme. The most interesting case is that of the E379K mutation, which so strongly affects binding of the substrate that no catalytic activity was observed (Table 2 and Figure 2 in Pilotto, Speranzini *et al.*, 2016). Y761H affects a very well conserved residue among amine oxidases, because of its positive effect on the oxidation of the substrate amine group (Li *et al.*, 2006). Similarly to what is observed for this group of enzymes, mutation to this residue affects the catalytic efficiency of demethylation (Table 2 in Pilotto, Speranzini *et al.*, 2016). The D556G mutation replaces a strongly negative charged group with a hydrogen atom. Thus, the histone tail, despite being equally sequestered from the solvent, is no longer constrained into the catalytically competent conformation and other positions and orientations of the substrate tail can be adopted before demethylation of the lysine 4 can occur (Tables 2 and 3 in Pilotto, Speranzini *et al.*, 2016).

Last, nucleosome recognition by LSD1 pathological mutants was also affected, as demonstrated by analytical size-exclusion chromatography (SEC) experiments with semi-synthetic nucleosomes (Figure 4 in Pilotto, *et al.*, 2016). Briefly, the experiment consists in incubating the flavo-dependent demethylase with nucleosomes carrying a reactive propargyl group at position 4 of histone H3 (the substrate residue). Formation of a covalent complex between the FAD cofactor and the nucleosome is monitored by the absorbance profile (UV-vis spectrum), whereas relative quantification of the formed complex versus free nucleosome species is assessed by analytical SEC. Further details can be found paragraph 6.1, in the Materials and methods section of Chapter III).

2. LSD1 pathological variants effect on transcription factors recruitment

LSD1 pathological mutants were tested for their effects on transcription factor binding within the catalytic site. In particular, proteins were analysed for their ability to bind SNAIL1 N-terminus and p53 C-terminus peptides in comparison with the wild-type LSD1.

Thermal stability and fluorescence polarization experiments revealed different behaviours among LSD1 proteins. The wild type and the D556G mutant exhibited nano-molar affinity for the conserved N-terminal SNAIL1 peptide, whereas E379K and Y761H showed a >30 times reduced affinity (Table 3 in Pilotto, Speranzini *et al.*, 2016). Differently, studies on various p53 CTD (C-terminal domain) peptides revealed that D556G and Y761H retain the binding affinity of the wild-type protein, and that only the E379K mutant is 10times less able to bind it, with a similar fashion to what happens for the substrate H3 tail (Figure 4 and Table in Speranzini *et al.*, 2017). The retained binding affinity for the mutants acquires even more importance in light of the fact that LSD1 does not demethylate p53 CTD, hence the interaction between the two proteins has a wider and more complex significance than regulation of p53 recruitment by 53BP1 (Haupt *et al*., 1997; Lin *et al*., 2001).

It is evident that D556G mutation does not affect the affinity for any substrate or non-substrate peptide, as the role of the mutated residue is mostly to allow acquisition of the catalytically competent conformation of H3, which is of course not needed in non-catalytic interactions with transcription factors. On the other hand, the reversal of charge of the E379K mutation strongly affects binding to all peptides, because of the disruption of the ionic interaction with the Arginine residue conserved at position 2 of all peptides (Figure 1 in Pilotto, Speranzini *et al.*, 2016). The case of the Y761H mutation is very interesting because of the different effects on either H3 or SNAIL1 or p53. Indeed, interaction with SNAIL1 peptide is the only hampered one (about 70 times reduction in affinity), probably because SNAIL1 exposes a Phenylalanine residue at that location, instead of a Lysine residue as in the case of both H3 and p53.

Conclusions

LSD1 catalytic and non-catalytic biological and pathological effects

The only three known non-lethal mutations of LSD1 active site show different effects on either the catalysis rate, subtrate-binding affinity or other chromatin factors binding through histone-mimicry. However, all of these three mutants retain unaltered binding to CoREST1 and HDAC1/2, and this might explain the viable phenotypes of the heterozygote carriers.

It appears evident from both these articles, that the biological function of LSD1 is clearly not limited to its catalytic efficiency, which is completely abolilshed in the E379K mutant. The biological role of the LSD1/CoREST1 system seems more related to its stable non-catalytic association with various complexes, involving either SNAIL1 family of transcription factors, or p53, or the nucleosome itself.

Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A., *et al.* (2005) ‘Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process’, FEBS Letters, 579(10), pp. 2203–2207. doi: 10.1016/j.febslet.2005.03.015.

Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E., *et al.* (2005) ‘Human histone demethylase LSD1 reads the histone code’, Journal of Biological Chemistry. doi: 10.1074/jbc.M509549200.

Forneris, F. *et al.* (2007) ‘Structural basis of LSD1-CoREST selectivity in histone H3 recognition’, Journal of Biological Chemistry. doi: 10.1074/jbc.C700100200.

Haupt, Y. *et al.* (1997) ‘Mdm2 promotes the rapid degradation of p53’, Nature, 387(6630), pp. 296–299. doi: 10.1038/387296a0.

Li, M. *et al.* (2006) ‘Functional role of the “aromatic cage” in human monoamine oxidase B: structures and catalytic properties of Tyr435 mutant proteins.’, Biochemistry, 45(15), pp. 4775–84. doi: 10.1021/bi051847g.

Lin, J. *et al.* (2001) ‘Inhibition of p53 transcriptional activity by the S100B calcium-binding protein.’, The Journal of biological chemistry, 276(37), pp. 35037–41. doi: 10.1074/jbc.M104379200.

Lin, T. *et al.* (2010) ‘Requirement of the histone demethylase LSD1 in Snai1-mediated transcriptional repression during epithelial-mesenchymal transition’, Oncogene, 29(35), pp. 4896–4904. doi: 10.1038/onc.2010.234.

Schenk, T. *et al.* (2012) ‘Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia’, Nature Medicine, 18(4), pp. 605–611. doi: 10.1038/nm.2661.

Tortorici, M. *et al.* (2013) ‘Protein recognition by short peptide reversible inhibitors of the chromatin-modifying LSD1/CoREST lysine demethylase’, ACS Chemical Biology, 8(8), pp. 1677–1682. doi: 10.1021/cb4001926.

Tunovic, S. *et al*. (2014) ‘De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome’, American Journal of Medical Genetics, Part A. Wiley-Liss Inc., 164(7), pp. 1744–1749. doi: 10.1002/ajmg.a.36450.

Wang, L. *et al*. (2005) ‘Early Embryonic Lethality in Mice with Targeted Deletion of the CTP:Phosphocholine Cytidylyltransferase Gene (Pcyt1a)’, Molecular and Cellular Biology. American Society for Microbiology, 25(8), pp. 3357–3363. doi: 10.1128/mcb.25.8.3357-3363.2005.

Wang, Y. *et al*. (2009) ‘LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer’, Cell, 138(4), pp. 660–672. doi: 10.1016/j.cell.2009.05.050.

Zheng, Y. C. *et al*. (2015) ‘A Systematic Review of Histone Lysine-Specific Demethylase 1 and Its Inhibitors’, Medicinal Research Reviews. John Wiley and Sons Inc., 35(5), pp. 1032–1071. doi: 10.1002/med.21350.

Human Molecular Genetics, 2016, Vol. 25, No. 12 2578–2587

doi: 10.1093/hmg/ddw120
Advance Access Publication Date: 19 April 2016
Original Article

ORIGINAL ARTICLE

LSD1/KDM1A mutations associated to a newly described form of intellectual disability impair demethylase activity and binding to transcription factors

Simona Pilotto[1,†], Valentina Speranzini[1,†], Chiara Marabelli[1], Francesco Rusconi[2], Emanuela Toffolo[2], Barbara Grillo[2], Elena Battaglioli[2,3,*] and Andrea Mattevi[1,*]

[1]Department of Biology and Biotechnology, University of Pavia, 27100 Pavia, Italy, [2]Department of Medical Biotechnology and Translational Medicine, University of Milan, 20133 Milano, Italy and [3]CNR, Institute of Neuroscience, 20129 Milano, Italy

*To whom correspondence should be addressed at: Department of Medical Biotechnology and Translational Medicine, University of Milan, via Viotti 5, 20133 Milano, Italy. Tel: +390250315844; Fax: +390250335864; Email: elena.battaglioli@unimi.it (E.B); Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, via Ferrata 1, 27100 Pavia, Italy. Tel: +390382985525; Fax: +390382528496; Email: andrea.mattevi@unipv.it (A.M.)

Abstract

Genetic diseases often lead to rare and severe syndromes and the identification of the genetic and protein alterations responsible for the pathogenesis is essential to understand both the physiological and pathological role of the gene product. Recently, *de novo* variants have been mapped on the gene encoding for the lysine-specific histone demethylase 1 (LSD1)/ lysine(K)-specific histone demethylase 1A in three patients characterized by a new genetic disorder. We have analyzed the effects of these pathological mutations on the structure, stability and activity of LSD1 using both *in vitro* and cellular approaches. The three mutations (Glu403Lys, Asp580Gly and Tyr785His) affect active-site residues and lead to a partial impairment of catalytic activity. They also differentially perturb the ability of LSD1 to engage transcription factors that orchestrate key developmental programs. Moreover, cellular data indicate a decrease in the protein cellular half-life. Taken together, these results demonstrate the relevance of LSD1 in gene regulation and how even moderate alterations in its stability, catalytic activity and binding properties can strongly affect organism development. This depicts a perturbed interplay of catalytic and non-catalytic processes at the origin of the pathology.

Introduction

Histone lysine methylation is one of the most studied and characterized modifications, because it dynamically regulates multiple fundamental biological processes, including chromatin accessibility, transcription, DNA repair, cell cycle and development (1–4). Histone methylation is essential for the establishment of cell identity and maturation, as underlined by numerous studies on disease and tumorigenesis associated with mutations in either Lys methyltransferases or demethylases, as

---

[†]These authors equally contributed to this work.

Received: February 19, 2016. **Revised:** March 30, 2016. **Accepted:** April 8, 2016

well as in histone Lys residues (5). Likewise, novel sequencing studies are detecting mutations in different chromatin modifiers associated to neurodevelopmental disorders, showing how variations in chromatin regulation can be linked also to both intellectual and physical disabilities (6–8). In this context, functional alterations of lysine-specific histone demethylase 1 (LSD1/lysine(K)-specific histone demethylase 1A (KDM1A), hereafter referred to as LSD1), a H3-mono- (1) or di- (2) methylated lysine 4 demethylase, have been implied not only in a variety of physiological processes but also in pathological conditions, ranging from haematopoiesis impairment, neurological disorders and cancer (9–11). More recently, dominant missense point mutations in the *LSD1* gene (NM_001009999.2) have been identified, and correlated to a new genetic disorder that phenotypically resembles the Kabuki syndrome (OMIM 147920) but with distinctive facial features, skeletal anomalies and, above all, cognitive impairment (12,13).

In the reported case studies, three missense point mutations have been mapped on the *LSD1* gene: c.1207G > A predicting p.Glu403Lys (14), c.1739A > G encoding for Asp580Gly (14) and c.2353T > C, encoding for Tyr785His (13). The key issue is that these mutations represent *de novo* variants expected to be deleterious, as they affect residues in the catalytic core of the enzyme (Fig. 1). Indeed, the mutations are all in heterozygosis, indicating that they represent a dominant condition. This is especially interesting in light of the fact that the *LSD1* gene is essential, as demonstrated by mice homozygous mutants failing to develop properly after implantation (15,16).

With the aim of understanding how these newly identified pathological mutations affect the binding, catalytic, structural and repressive properties of this chromatin-associated enzyme, we produced and purified the three LSD1 disease-associated variants. Our combined analyses on the molecular features of the mutants both *in vitro* and *in vivo* provide new insights on how alterations of this epigenetic enzyme can translate into pathological conditions.

Results

LSD1 mutants are structurally stable yet catalytically impaired

For consistency with previously published works and to facilitate structural comparisons, the residue numbering used in the genetic studies for LSD1 protein (12–14) has been adapted to reference UniProt sequence O60341, which has been used so-far for all three-dimensional structures present in the Protein Data Bank. Therefore, the mutations Glu403Lys, Asp580Gly, Tyr785His will be hereafter referred to as Glu379Lys (E379K), Asp556Gly (D556G) and Tyr761His (Y761H). Domain and structural overview of recombinant LSD1-CoREST (corepressor of re1-silencing transcription factor) shows that these clinically relevant mutations all affect the active site. Glu379 and Asp556 are located towards the entry of the histone-tail binding site whereas Tyr761 occupies a more buried position to form the so-called aromatic cage, which is proximal to the flavin cofactor in the core of the catalytic centre (Fig. 1A).

We have recombinantly expressed, purified and characterized the three *de novo* LSD1 variants associated to pathogenic conditions. For all proteins, we performed co-purification with the corepressor CoREST, and in all cases this resulted in the same total recombinant protein yield in the range 1–3 mg/l of *Escherichia coli* culture, similar to LSD1 wild-type. All proteins showed no sign of folding instability (neither aggregation nor precipitation, as judged by gel filtration), therefore exhibiting the same tight association to CoREST. Similarly, thermal stability assays on the LSD1-CoREST complexes showed no significant changes in the unfolding melting temperatures (Table 1). Taken together, these results indicated that the pathological mutations do not affect the stability of the enzyme and its capability to bind to the co-repressor partner.

We next assessed the catalytic properties of the pathological LSD1 mutants. To test their ability to demethylate monomethylated-Lys4 of H3, we used a methylated peptide in a coupled assay (17). We found that D556G and Y761H are 10/20-fold less efficient compared with wild-type (essentially because of a decrease in the K_{cat}), whereas the activity of E379K is barely detectable, as listed in Table 2. To better understand if the decreased catalytic activities could be ascribed to a reduced binding affinity for histone H3 tail, we set up fluorescence polarization binding assays (Fig. 2). Furthermore, we assessed the thermal stability of LSD1 wild-type, E379K, D556G and Y761H LSD1 proteins in complex with CoREST when incubated with the same histone-derived peptide used for the enzymatic assays (Table 1). Both experimental approaches consistently indicated that LSD1 mutation E379K strongly impairs binding to the H3 N-terminal peptide, whereas mutations D556G and Y761H

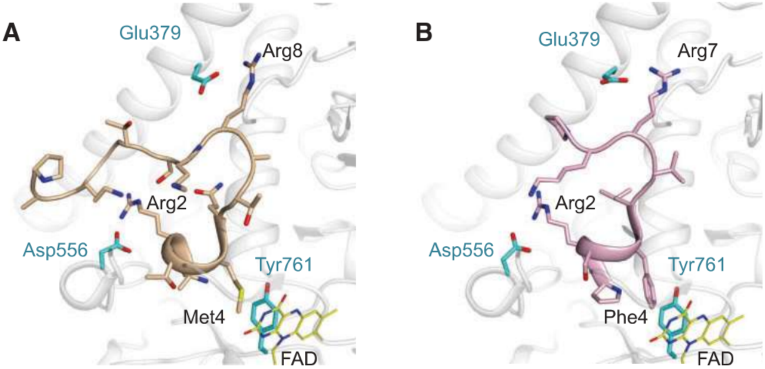


Figure 1. Structural analysis of Kabuki-like LSD1 mutations. Distributions in LSD1 (white) of the three residues affected by pathological mutations (cyan) and interactions with H3 tail substrate and interactor peptides. **(A)** Histone H3, depicted in wheat. The structure (PDB 2V1D) presents the histone peptide where Lys4 is substituted by a Met, mimicking demethylated product. **(B)** SNAIL1 (PDB 2Y48) in pink. Both substrate and the transcription factor peptides interact tightly with Glu379, Asp556 and Tyr761.

show binding similar to wild-type enzyme (Table 3). These biochemical assays were complemented by the elucidation of the three-dimensional structures of the LSD1 mutants in complex with CoREST (Table 4). The structures showed nearly identical quaternary arrangements, without overall or local conformational changes compared to the wild-type LSD1-CoREST native structure (Fig. 3). The conformation of both the catalytic and the FAD-binding sites is virtually identical in all structures, confirming the predicted positions of the pathogenic mutations in proximity to the H3 substrate binding site. Likewise, CoREST binding is indistinguishable from that observed in the wild-type enzyme, as expected from the biochemical analysis. Collectively, these investigations demonstrated that the mutations primarily affect enzyme catalysis and, to a minor extent, binding to the H3 N-terminal peptide without altering the protein conformation.

However, for the purpose of our analysis, it should be pointed out that each of the three alterations likely exerts its effect in a mutant-specific way. The strongest perturbation affecting substrate binding and catalysis is caused by E379K. Glu379 is directly involved in the recognition of Arg8 of histone H3 (Fig. 1A). It is conceivable that charge reversal at this locus essentially disrupts binding to the H3 N-terminal tail, which results in the mutant being enzymatically inactive. More subtle seem to be the effects of the other two mutations. Y761H affects a residue that is generally conserved among amine oxidases. Previous studies on human monoamine oxidases A and B have shown that a Tyr at this position creates a sterically and electrostatically favourable environment for the substrate amine group, promoting its oxidation (18). Indeed, the Tyr-to-His mutation in monoamine oxidases decreases the catalytic efficiency and the same effect is now found in LSD1 (Tables 2). D556G is the mutation with the milder consequences on the enzyme biochemical properties as it features only a 10-fold reduced K_{cat}. Asp556 is part of a cluster of negatively charged residues that interact with Arg2 of the H3 N-terminal tail (Fig. 1A). The D556G mutations still allows the active site to engage histone tail (Table 3). Likely, removal of the negatively charged Asp side chain leads to a non-productive histone binding conformation, which does not allow the correct positioning of methylated Lys4 to face the flavin for efficient oxidative demethylation.

These data are supported by experiments performed with nucleosomal particles bearing a propargyl methyl-Lys4 analogue. These semisynthetic nucleosomes effectively function as a suicide substrate whose modified H3-Lys4 forms a covalent adduct with the flavin cofactor upon binding to the enzyme (19). Consistently with the non-productive binding of the H3 tail, all three pathogenic mutations hamper the covalent reaction between the modified Lys4 and the LSD1 flavin cofactor, as none

Table 1. Thermal stability assays with LSD1-CoREST wild-type and mutants[a]

	No ligand[b]	H3 (1–21)	SNAIL1 (1–9)
LSD1-CoREST	0 (T_m=50.5)	+ 2.5	+ 7.5
$LSD1_{E379K}$-CoREST	+1.5	+0.5	+4.5
$LSD1_{D556G}$-CoREST	+1	+1.5	+10.5
$LSD1_{Y761H}$-CoREST	+2	+1.5	+2.5

[a] All temperature shifts are reported in °C and refer to the T_m value measured in the absence of ligands. H3 (1–21) is the Lys4-methylated H3 peptide corresponding to residues 1–21 of H3. SNAIL1 (1–9) is a peptide with sequence corresponding to residues 1–9 of human SNAIL1.

[b] The melting temperature (T_m) of the wild-type protein is in brackets. Temperature shifts refer to the T_m value measured for the wild-type in the absence of ligands.

Table 2. Kinetic parameters for activity of LSD1-CoREST wild-type and mutants[a]

	K_m (μM)	K_{cat} (min^{-1})	K_{cat}/K_m ($min^{-1}/\mu M$)
LSD1-CoREST1	3.89 ± 0.43	4.72 ± 0.16	1.21 ± 0.15
$LSD1_{E379K}$-CoREST	n.d.[b]	<0.01	n.d.[b]
$LSD1_{D556G}$-CoREST	2.55 ± 0.45	0.30 ± 0.01	0.12 ± 0.21
$LSD1_{Y761H}$-CoREST	1.28 ± 0.26	0.18 ± 0.01	0.14 ± 0.25

[a] All experiments were performed in quadruplicate.

[b] n.d., not detectable because of too low activity.

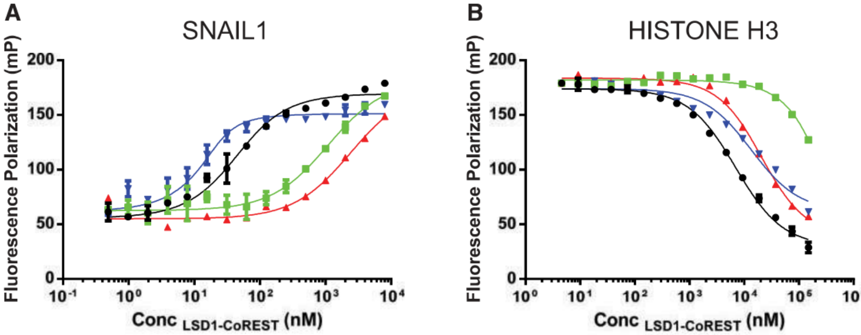


Figure 2. LSD1-CoREST mutants bind differently to histone substrate and transcription factor SNAIL1. Binding properties of LSD1 wild-type (black), Glu379Lys (green), Asp556Gly (blue) and Tyr761His (red; in complex with CoREST) to histone H3- and SNAIL-derived peptides. (A) Increasing concentrations of purified LSD1-CoREST wild-type and mutants were incubated with 5(6)-carboxytetramethylrhodamine-conjugated SNAIL1 N-terminal peptide (residues 1–9). Changes in polarization were measured in millipolarization (mP) units and plotted against the concentration of LSD1-CoREST. The corresponding K_d values are reported in Table 3. (B) Binding to me_2Lys4 1-21 H3 N-terminal peptide was measured in competition against SNAIL1. LSD1-CoREST wild-type and mutants were incubated with increasing (0–150 μM) concentrations of competing H3 peptide. Comparison with the curve and K_d values obtained by direct binding (A, black) allowed affinity measurements and K_d calculation, as reported in Table 3. Both in A and in B, error bars correspond to standard deviations for all measurements ($n \geq 3$). The choice of using SNAIL1 (instead of H3) for direct binding is based on the proven higher affinity for this peptide (19,28), which allowed a larger dynamic range of affinity.

Table 3. Affinity of LSD1-CoREST wild-type and mutants to histone and SNAIL1 peptides[a]

	H3 (1–21) K_d (μM)	SNAIL1 (1–9) K_d (μM)
LSD1-CoREST	3 ± 0.3	$3.43{\cdot}10^{-2} \pm 5.8{\cdot}10^{-3}$
LSD1$_{E379K}$-CoREST	> 150	1.1 ± 0.3
LSD1$_{D556G}$-CoREST	3.8 ± 1.3	$4{\cdot}10^{-3} \pm 1.8{\cdot}10^{-3}$
LSD1$_{Y761H}$-CoREST	9.8 ± 4.1	2.4 ± 0.9

[a] All experiments were performed in quadruplicate.

Table 4. Crystallographic data collection and refinement statistics for LSD1-CoREST mutants

	E379K	D556G	Y761H
PDB code	5L3C	5L3B	5L3D
Unit cell (Å)[a]	120.4, 180.4, 235.4	119.1, 179.1, 234.4	120.3, 177.7, 235.2
Resolution (Å)	3.3	3.3	2.6
R_{sym} (%)[a,b]	13.0 (81.0)	13.1 (61.3)	7.6 (97.0)
$CC_{1/2}$[b,c]	0.99 (0.47)	0.99 (0.71)	1.00 (0.52)
Completeness (%)	99.0 (99.0)	99.6 (99.9)	99.2 (100.0)
Unique reflections[b]	38276 (4604)	37869 (4593)	76680 (4596)
Redundancy[b]	3.8 (3.9)	3.8 (3.8)	4.4 (4.5)
Average Intensity/σI[b]	8.5 (1.2)	7.2 (1.7)	10.5 (1.0)
Number of protein atoms	6356	6342	6344
Number of solvent atoms	0	0	64
Ramachandran favoured (%)	96.3	96.0	96.9
Ramachandran allowed (%)	3.5	3.4	2.9
Ramachandran outliers (%)	0.2	0.6	0.2
Rwork (%)[d]	19.9	19.5	20.8
Rfree (%)[d]	22.3	20.8	21.9
Rmsd bond length (Å)	0.009	0.020	0.009
Rmsd bond angle (°)	1.40	1.30	1.20

[a] For all structures, space group is I222. $R_{sym}=\sum|I_i-|/\sum I_i$, where I_i is the intensity of ith observation and is the mean intensity of the reflection.

[b] Values in parentheses are for reflections in the highest resolution shell.

[c] The observed resolution limits of the diffraction data were selected by evaluating the mean(I) correlation between half datasets, as defined by Karplus and Diederichs (44).

[d] $R_{work}=\sum|F_{obs}-F_{calc}|/\sum|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. R_{work} and R_{free} were calculated using the working and test sets, respectively.

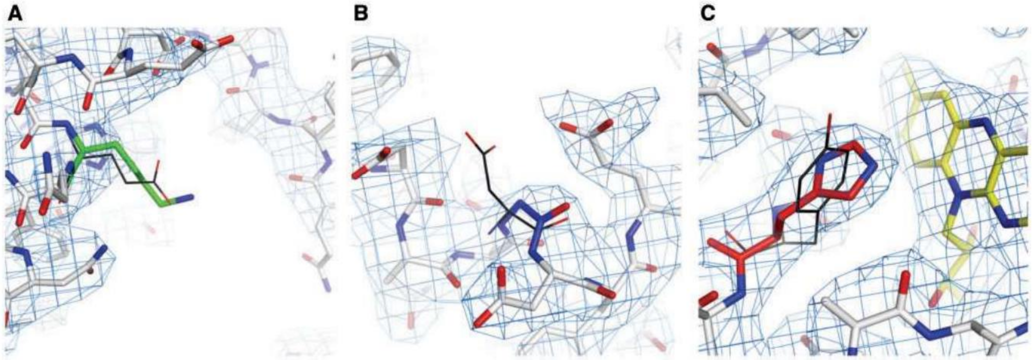


Figure 3. Three-dimensional structure of LSD1 pathological mutants. Distributions in LSD1 of each of the three residues affected by pathological mutations. Structures are compared to wild-type (black, thin line). (A) Glu at position 379 is substituted with a Lys (green), whose side chain shows to be disordered due to the lack of defined electron density. (B) Asp at position 556 is mutated to a Gly (blue). (C) Tyr at position 761 is replaced by a His (red). Electron density 2Fo-Fc map of each mutant LSD1 is represented as blue grid at contour level of 1.2σ.

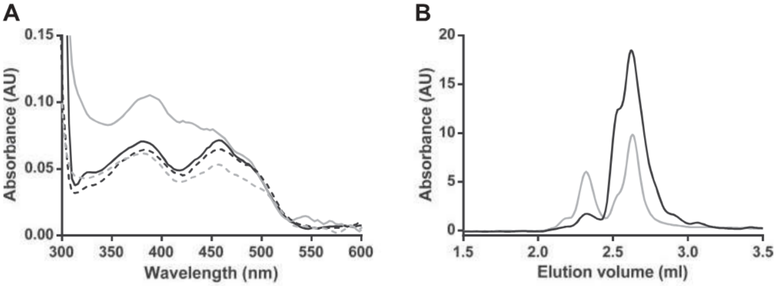


Figure 4. Pathological mutations in LSD1 hamper the binding to nucleosomal particles. (**A**) Wild-type (dashed grey) and mutant (dashed black, for simplicity only D556G mutant is shown) LSD1-CoREST (25 μM) show the typical absorbance peaks at 380 and 458 nm of the oxidized FAD. Binding of propargyl-modified H3 nucleosomes to wild-type LSD1 causes a spectral change with the formation of a single peak close to 400 nm (solid grey). Upon interaction with modified nucleosomes, the mutant spectra (solid black with reference to D556G) are unaltered, indicating no covalent binding to the FAD. (**B**) This observation was confirmed when purification of the covalent LSD1$_{D556G}$-CoREST-nucleosome was attempted. The analytical gel filtration chromatogram shows that LSD1 wild-type (grey) forms a 1:1 covalent complex with the nucleosome (elution volume 2.3 ml), whereas mutants (black, for simplicity only D556G mutant is shown) only show the elution profiles of free nucleosomes (2.6 ml) and free LSD1-CoREST (2.7 ml).

of the mutants are able to form such a stable complex with the modified nucleosomes (Fig. 4).

LSD1 inactive mutants retain the ability to associate to corepressor partners and gene repression capability in cell lines

LSD1 transcriptional repressive capability is tightly linked to and dependent on its association with corepressors histone deacetylase 1/2 (HDAC1/2) and CoREST inside the nucleus (20–23). We initially confirmed nuclear localization of hemagglutinin (HA)-tagged LSD1 mutants in non-neuronal and neuronal cells (Fig. 5A). Next, we assessed the ability of LSD1 mutants to recruit a proper corepressor complex through HDAC1/2 and CoREST binding in human cells. We transfected each HA-tagged LSD1 mutant into Henrietta Lack (HeLa) cells and then probed HA-immunoprecipitates for LSD1 association with endogenous HDAC1, HDAC2 and CoREST. All three mutants showed deacetylase and corepressor binding comparable to wild-type HA-LSD1 (Fig. 5B), in agreement with the biochemical and structural studies.

We then functionally assessed the effect of the mutations in a cell-based transcriptional assay by generating Gal4-LSD1 mutants that were transfected together with 5xUAS–TK–LUC reporter gene in HeLa and neuroblastoma SH-SY5Y cell lines (Fig 5C–E). For each mutant, the repressive activity was measured as *Luciferase* expression normalized over a cotransfected *Renilla* reporter. The repressive behaviour was ascertained as residual luciferase activity compared to the empty Gal4-vector, which was arbitrarily set to 100%. In parallel with the pathological mutations, we used the same assay to analyse the effect on transcription of a previously characterized enzymatically inactive LSD1 mutant, K661A, specifically designed to interfere with demethylase reaction (21,24). As shown in Figure 5C, Gal4-LSD1 K661A displayed a modest although significant reduction in repressive activity in HeLa cells compared with Gal4-LSD1, while in the neuronal cell context the effect of the inactivating mutation was stronger. Repressive activities of Gal4-LSD1 and Gal4-LSD1 K661A defined the dynamic range in which to evaluate demethylase-dependent transcriptional modulation induced by the three pathological mutations. Gal4-LSD1 D556G, Y761H and E379K did not substantially differ when compared to wild-type Gal4-LSD1. Only Gal4-LSD1 E379K showed a slight, yet significant, impairment of the repressive activity in HeLa cells, similarly to LSD1 K661A (Fig. 5C). The ability of LSD1 mutants to retain repressive capacity could be ascribed to the unaltered hallmark of all LSD1 mutants to efficiently interact with HDAC1/2 and CoREST in corepressor complexes (Fig. 5B). Collectively, these data suggest that, at least in this system, substantial impairment of the LSD1 demethylase activity does not affect repression of the target reporter gene. Consistently, previously reported experiments on the LSD1 K661A mutant showed that also this catalytically inactive enzyme is able to retain full association with the breast cancer type 2 susceptibility protein (BRCA2)-associated factor-HDAC complex, responsible for the repressive activity towards target genes (25).

LSD1 mutations reduce protein stability in vivo

Having observed that, despite being catalytically impaired, the mutants retain substantial repressive activity, we next investigated whether LSD1 stability is affected by the pathological mutations in human cell lines. This question was addressed by probing protein levels in HeLa cells after blocking protein synthesis via cycloheximide administration to the culture media (26), followed by time-course comparisons with the wild-type protein (Fig. 6). In our experimental conditions, HA-LSD1 displays a half-life of ~33 h. Interestingly, all mutants showed substantially shortened half-life, ranging from 15 h for Y761H down to 9 and 5.5 h for D556G and E379K, respectively. Such a significant reduction in protein levels will potentially add to the effects on catalytic and binding properties, contributing to the pathological phenotype associated to these mutations.

Pathological mutations in LSD1 have different effects on binding to transcription factors

To further examine the properties of LSD1 E379K, D556G and Y761H, we assessed their capability to bind the transcription factor zinc finger protein SNAI1 (SNAIL1), which is implicated in the differentiation of epithelial cells into mesenchymal cells during embryonic development (27). SNAIL1 and similar

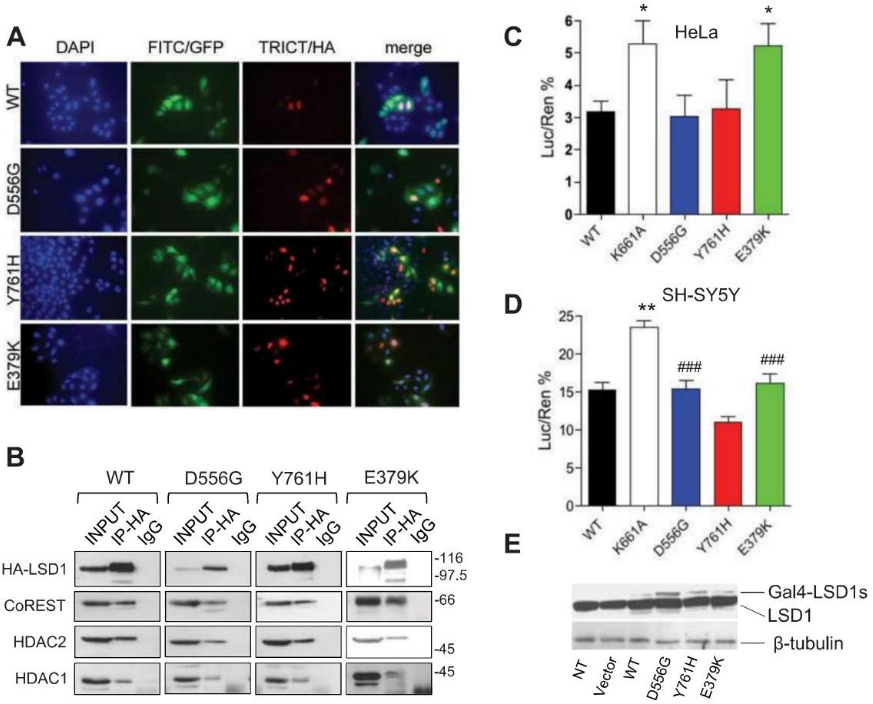


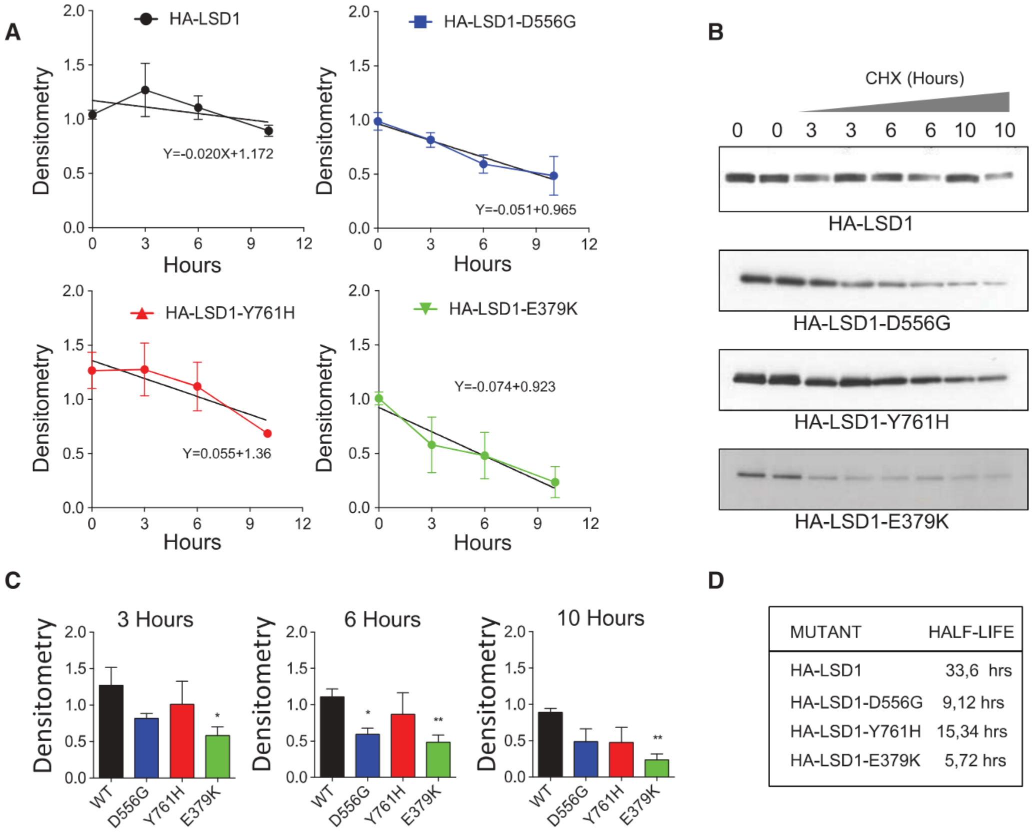
Figure 5. Effect of patient mutations on LSD1 repressive activity. (**A**) LSD1 D556G, Y761H and E379K do not alter LSD1 nuclear localization in HeLa cells. Immunofluorescence analysis of transiently expressed HA-LSD1 and annotated HA-tagged mutants co-transfected with EGFP, stained with anti-HA antibody. Nuclei are evidenced with DAPI staining (magnification 40×). (**B**) In human cell lines, LSD1 mutants retain the ability to bind to corepressors HDAC1/2 and CoREST. Protein extracts from HeLa cells transfected with HA-tagged LSD1 and related mutants (HA-LSD1-D556G, HA-LSD1-Y761H, HA-LSD1-E379K) immunoprecipitated with anti-HA antibody or pre-immune IgG, separated on SDS–PAGE and immunodecorated with HA, CoREST, HDAC1 and HDAC2 antibodies. (C-D) LSD1 and LSD1 mutants D556G, Y761H, E379K together with the demethylase-dead mutant LSD1 K661A, fused to Gal4, were assayed for their ability to repress the 5xUAS-TK-LUC reporter gene in (**C**) HeLa cells at 1:0,125 reporter:repressor molar ratio and in (**D**) neuroblastoma SH-SY5Y cells at 1:2 reporter:repressor molar ratio. The *Luciferase* activity normalized on the co-transfected *Renilla* reporter was expressed as a percentage of the Gal4 empty vector put to 100% (not shown). Mean values ± SE are shown. Student's t test was applied to percentage values (*refers to LSD1-WILD-TYPE; # refers to LSD1-K661A). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{\#\#\#}P < 0.001$. (E) Protein expression of LSD1 and mutant proteins, normalized on β-tubulin level.

transcription factors are known to recruit the repressive complex LSD1-CoREST-HDAC to their target gene promoters through binding to LSD1 active site, mimicking the histone H3 N-terminal tail (28,29). We performed thermal stability and binding assays using a peptide corresponding to the N-terminal tail of SNAIL1, whose sequence is strictly conserved in all members of SNAIL1-related family (30) (Tables 1 and 3). From these experiments, it emerged that the three pathological mutations differentially affect binding to SNAIL1: two of them (E379K and Y761H) drastically reduce affinity (> 30-fold; Table 3), whereas D556G features little variations compared to the wild-type enzyme.

These findings can be rationalized in light of the three-dimensional structure of LSD1-CoREST bound to SNAIL1 (Fig. 1B) (28). Removal of Asp556 negatively charged side chain (D556G) has no effect on transcription factor binding. A plausible explanation to this observation is that a glycine at this position may allow solvent molecules to interact with SNAIL1 Arg2 side chain, with little effect on the overall binding affinity for the transcription factor. A similar effect is observed on the affinity for the H3 peptide, although the precise histone binding mode and efficiency is likely affected, as shown by reduced K_{cat} rates. In contrast, E379K mutation profoundly affects SNAIL1 binding by reversing the charge of Glu379 which normally interacts with Arg7 of SNAIL1. Again, such a strong perturbation matches the drastic effect by this mutation on H3 binding. More specific is the impact on SNAIL1 binding induced by Y761H. In this case, a polar histidine side chain affects the characteristic hydrophobic interaction between the aromatic rings of Phe4 of SNAIL1 and Tyr761 of LSD1 (Fig. 1B). Hence, the Y761H mutation has a much stronger effect on SNAIL1 binding (70-fold decrease in affinity) compared with H3, which features a Lys (Lys4) rather than a Phe in direct contact with Tyr761. These data demonstrate how active-site mutations can differentially perturb binding of histone tail and histone-mimicking transcription factors.

Discussion

Exon sequencing data allowed the identification in three young patients of three different point mutations affecting residues in the active site of the histone demethylase LSD1 (12–14). In this work, we show that all of these mutations negatively influence both the demethylase catalytic efficiency and cellular protein stability, though to different extents. In contrast, these LSD1



MUTANT	HALF-LIFE
HA-LSD1	33,6 hrs
HA-LSD1-D556G	9,12 hrs
HA-LSD1-Y761H	15,34 hrs
HA-LSD1-E379K	5,72 hrs

Figure 6. Pathological mutations destabilize LSD1 protein. Cycloheximide (CHX, 100 µg/ml) time-courses stability assay on HeLa cells transfected with HA-LSD1, HA-LSD1-D556G, HA-LSD1-Y761H or HA-LSD1-E379K show that mutant LSD1 isoforms are less stable than the wild type. (**A**) Linear regression analysis of time course densitometry experiments relative to wild-type LSD1 and mutants. (**B**) Representative western blot films for wild type LSD1 and mutant isoforms. A loading control for each transfected cell line was performed (not shown). (**C**) Time-courses densitometry analysis relative to the untreated cells of each transfected HA-LSD1 and mutated isoforms. (**D**) Inferred protein half-lives. Mean values ± SE are shown. Student's t test was applied. $^{*}P < 0.05$; $^{**}P < 0.01$ referred to WT HA-LSD1.

mutations retain repressive activity on a reporter gene, most likely due to unaltered binding to the co-repressors HDAC1/2 and CoREST. Therefore, as H3-Lys4 methylation is a widespread and key epigenetic mark, alteration in its homeostasis and reprogramming is seemingly the main reason for the extensive developmental effects observed in the patients affected by the newly discovered pathologic condition caused by LSD1 mutations (12). Thus, the enzymatic histone demethylase activity is far from being dispensable for proper development, as already reported in literature for either embryonic stem cells and haematopoietic differentiation or for the establishment of proper neuronal identity (22,31–34).

Notwithstanding its catalytic function, LSD1 can also recognize and bind a variety of non-histone ligands, ranging from other chromatin modifiers such as transcription factors and the tumour suppressor p53 (10,21,29,35). In this work, we found that the pathological mutations do affect such interactions, with a focus on the binding to SNAIL1, known as a master regulator of the epithelial–mesenchymal transition during development and tumour progression. Indeed, this binding is deeply affected by two (Y761H and E379K) of the analysed mutations, which can further contribute to the pathogenesis. In this context, it cannot be excluded that pathological effect of D556G is exacerbated by adverse alterations on the binding to interactor(s) other than SNAIL1-related factors. For instance, it has been recently shown that the neuro-specific splicing variant of LSD1 (22) might be endowed with the ability to demethylate other Lys residues such as H3-Lys9 or H4-Lys20, depending on the interaction with specific co-repressor or co-activators different from CoREST. We cannot exclude that also these regulatory interactions could be affected by the mutations, impairing neuro-specific roles of LSD1 (36,37). Hence, it is clear that the function of the active site of LSD1 is not limited to the recognition and modification of the methylated histone H3 tail. Rather, it is a hub for competitive interactions with the nucleosomal tails as well as transcription factors, implying that even catalytically and biochemically mild mutations targeting this protein region can have pleiotropic (both catalytic and non-catalytic) effects. Along this line, it has been very recently demonstrated that simply a partial loss of maternal LSD1 demethylase is sufficient to block maternal oocytes epigenetic reprogramming, leading to deregulated gene expression during embryonic maturation and consequentially developmental aberrations at birth (38). In conclusion, LSD1 and associated factors are such key players in establishing and maintaining patterns of chromatin modifications that a moderate reduction in enzymatic activity and protein levels lead to severe pathological conditions.

Materials and Methods

Protein expression and purification

LSD1 mutants were prepared using standard mutagenesis procedures (QuickChange Mutagenesis Kit, Agilent Technologies Milano, Italy), and purified as the wild-type proteins. Protein expression and purification of all constructs were performed using previously published protocols (39).

Thermal stability assays

Protein thermostability was tested using ThermoFAD as previously described (40). All proteins were incubated at 4 μM final concentration together with peptides (H3 1–21 and SNAIL1 1–9; Table 1) at 200 μM final concentration in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)/NaOH pH 7.5 and stability was assessed over a temperature range 20–90°C.

LSD1 activity assays

Enzymatic activities were evaluated by a peroxidase-coupled assay monitoring hydrogen peroxide formation (39). The reaction mixture contained 50 mM HEPES/NaOH pH 7.5, 0.1 mM Amplex Red, 0.3 mM horseradish peroxidase, 0.1 μM LSD1-CoREST and varying concentrations of monomethylated H3 – K4 peptide substrate (0.3–40 μM). Fluorescence changes following the conversion of Amplex Red (Invitrogen ThermoFisher Scientific) to Resorufin were monitored at 535 and 590 nm for excitation and emission, respectively. Assays were performed at 25 °C in 96-plate format using a CLARIOstar plate reader (BMG Labtech, Germany). Initial velocity values were fitted to Michaelis–Menten equations.

Fluorescence polarization binding assays

Binding assays to LSD1-CoREST wild-type and mutants were carried out monitoring the change in polarization properties of SNAIL1 peptide (amino acids 1-9) fluorescently labelled with 5(6)-carboxytetramethylrhodamine (TAMRA) as previously described (19). Briefly, all experiments were carried out in 15 mM KH_2PO_4 pH 7.2, 1 mg/ml BSA, 5% glycerol at 25 °C. CLARIOstar (BMG Labtech, Germany) plate reader was used with 540 nm excitation and 590 nm emission filters. Experiments were performed in triplicates using 384-well microplates (CORNING, UK). Direct binding of SNAIL1-TAMRA at a final concentration of 10 nM was performed to calculate K_d values. LSD1-CoREST was kept constant at 30 nM for wild-type, 1 μM for E379K, 3 nM for D556G, 2.5 μM for Y761H. For competitive experiments, each well contained LSD1-CoREST at the listed concentrations and labelled peptide (10 nM final). Next, decreasing concentrations (typically in the 0–150 μM range) of the unlabelled competing H3 peptide (ARTme$_2$KQTARKSTGGKAPRKQLA) was added to the wells mixture and competitive curves recorded with the same settings.

Binding to semi-synthetic nucleosomes

Nucleosome binding by LSD1-CoREST wild-type, E379K, D556G and Y761H was performed by incubating each protein with semi-synthetic recombinant nucleosomes, which were prepared as previously described (19). LSD1-CoREST proteins (in 25 mM KH_2PO_4 pH 7.2, 5% glycerol) were mixed with semi-synthetic nucleosomes (in 20 mM Tris(hydroxymethyl)aminomethane (Tris)/HCl pH 7.5, 1 mM (ethylenedinitrilo)tetraacetic acid (EDTA), 1 mM 1,4-dithiothreitol (DTT)) at 1.5:1 molar ratio and incubated on ice for 1 h. Absorbance spectra were recorded in the UV-visible range of wavelengths using the NanoDrop spectrophotometer (ThermoScientific). Analytical size-exclusion chromatography was performed for each mix on silica gel column WTC-030N5 (4.6 × 300 mm) (Wyatt Technology, CA) equilibrated in buffer 200 mM KCl and 10 mM Tris/HCl pH 7.5 (4°C). Detection wavelength were set at 214 nm (peptide bond), 260 nm (DNA) and 280 nm (aromatic protein side chains) and elution profiles were recorded using an AKTAmicro purification system (GE Healthcare). Figure 4B shows only the 214 nm profiles for graphical simplicity.

X-ray crystallography and structural analysis

Crystals of mutant LSD1-CoREST complexes were obtained at 20 °C as previously described (39). X-ray diffraction data were collected at 100 K at beamline X06DA at the Swiss Light Source (SLS, Villigen, Switzerland) and at beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Data processing and crystallographic refinement were carried out with standard procedures using XDS (41) and programs of the CCP4 suite (42). Atomic co-ordinates for the E379K, D556G and Y761H mutants have been deposited in the Protein Data Bank with accession codes 5L3B (D556G), 5L3C (E379K), 5L3D (Y761H) (Table 4). Structural figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

Cell cultures

HeLa and SH-SY5Y cells were, respectively, cultured in Dulbecco's modified Eagle's medium (DMEM) and Roswell Park Memorial Institute (RPMI) medium supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin and 1% glutamax. Cycloheximide was solubilized in DMSO and used at 100 μg/ml to block HeLa cells protein synthesis. For stability assays, transfected cells were seeded 24 h upon transfection (day 1) in four different wells, one for each condition (control, 3–6 and 10 h treatment with cycloheximide). On day 2, cells were collected at different time points and analysed by western blot (26).

Plasmids and transfection

Gal4-LSD1 fusion constructs containing full-length LSD1 as well as pCGN vectors encoding HA-LSD1 have been described elsewhere (22,24). LSD1-K661A, D556G, Y761H and E379K were obtained by site-specific mutagenesis using as template Gal4-LSD1 and pCGN-HA-LSD1 plasmids with QuikChange II Site-Directed Mutagenesis Kits (Stratagene, La Jolla, CA, USA). All plasmids were sequenced. HeLa and SH-SY5Y cells were transiently transfected using Lipofectamine 2000 (Invitrogen, Waltham, MA, USA) and processed for transcription repression assay after 48 h. For stability and immunoprecipitation assays, transfection was performed with Lipofectamine LTX (Invitrogen) according to manufacturer's instructions.

Immunoprecipitation assays

Experiments were performed as reported (24). Briefly, HeLa cells protein extraction was performed in low-stringency buffer (10% glycerol, NaCl 150 mM, imidazole 10 mM, 0.5 mM EDTA 0.5% Triton-X100, DTT 0.5 mM) supplemented with 1 mM phenylmethanesulfonyl fluoride and 1× Protease Inhibitors Cocktail (Roche) and 1× Phosphatase Inhibitor Cocktail (Roche). 0.5 mg of cell extract were used and collected with HA-conjugated Agarose beads (Santa Cruz, Santa Cruz, CA). After incubation, beads were washed with IP buffer, immunoprecipitates eluted with 1X sodium dodecyl sulphate (SDS) Sample buffer and processed for western blot. Quantification was performed using ImageJ software. Western blotting experiments were performed as previously described (43). Lysates and immunoprecipitates were resolved on 8% SDS–PAGE (polyacrylamide gel electrophoresis) gels whereas for LSD1 stability assays, HA-LSD1 and HA-LSD1 mutants were resolved on 7% SDS–PAGE gels. Antibodies were the following: Anti LSD1 (C69G12 CST, Denver, MA, USA); anti CoREST (07-455 Merck Millipore, Billerica, MA, USA); anti-HDAC2 (ab7029 Abcam, Cambridge, UK); anti-HDAC1 (ab7028 Abcam, Cambridge, UK) anti-α/β Tubulin (2148 CST, Danvers, MA, USA), anti-HA (sc-7392 and sc-80 Santa Cruz Biotechnology, CA, USA).

Repression assays

5xUAS-TK-LUC reporter plasmid (24) was used at the indicated molar ratio relative to the expression plasmids pGal4-LSD1 and pGal4-LSD1 mutants. Control experiments were carried out by using equivalent molar amounts of pGal4 empty vectors. DNA was maintained constant through pBSIIKS buffering (Stratagene, La Jolla, CA, USA). pRL-TK-reporter vector (Promega, Madison, MI, USA) was used to normalize for transfection efficiency. The luciferase reporter activity was determined with the Dual-Luciferase reporter assay system (Promega) according to the manufacturer's instructions. Values of *Firefly* luciferase were normalized over *Renilla* luciferase (both expressed as relative luminescent units). The activity of each construct was expressed as a percentage of the pGal4 empty vector.

Nuclear localization in cells

HeLa cells were transfected using Lipofectamine 2000 (Invitrogen). After 2 days cells were fixed with a PBS solution containing 4% paraformaldehyde for 10 min. Cells were incubated with anti-HA (sc-80) for 3 h in GDB buffer (30 mM phosphate buffer, pH 7.4, containing 0.2% gelatin, 0.5% Triton X-100 and 0.8 M NaCl), followed by 1 h incubation with cyanine 3-conjugated secondary antibody (The Jackson Laboratory, USA) and mounted in VectaDAPI medium (Vector Laboratories, USA). Images were acquired using Nikon Eclipse E600 microscope.

Acknowledgements

We acknowledge SLS and ESRF for provision of synchrotron radiation facilities and their staff during data collection. We thank Dr. Peter Lorentzen and Dr. Karen M. Park for helpful information on the case studies.

Conflict of Interest statement. None declared.

Funding

This work was supported by AIRC (IG-15208 to A.M.), MIUR (Progetto Bandiera Epigenomica—EPIGEN), Telethon (GGP12007 to A.M. and GGP14074 to E.B.) and the ONLUS Insieme per la Ricerca (PCDH19 to E.B.).

References

1. Rothbart, S.B. and Strahl, B.D. (2014) Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta*, **1839**, 627–643.
2. Mosammaparast, N., Kim, H., Laurent, B., Zhao, Y., Lim, H.J., Majid, M.C., Dango, S., Luo, Y., Hempel, K., Sowa, M.E. *et al.* (2013) The histone demethylase LSD1/KDM1A promotes the DNA damage response. *J. Cell Biol.*, **203**, 457–470.
3. Laugesen, A. and Helin, K. (2014) Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell*, **14**, 735–751.
4. Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R. and Schule, R. (2005) LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature*, **437**, 436–439.
5. Herz, H.M., Morgan, M., Gao, X., Jackson, J., Rickels, R., Swanson, S.K., Florens, L., Washburn, M.P., Eissenberg, J.C. and Shilatifard, A. (2014) Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. *Science*, **345**, 1065–1070.
6. Shen, E., Shulha, H., Weng, Z. and Akbarian, S. (2014) Regulation of histone H3K4 methylation in brain development and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **369**, 20130514.
7. Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.P. *et al.* (2005) Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.*, **76**, 227–236.
8. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Genevieve, D., Cormier-Daire, V., van Esch, H., Fryns, J.P. *et al.* (2006) Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am. J. Hum. Genet.*, **79**, 370–377.
9. Rusconi, F., Paganini, L., Braida, D., Ponzoni, L., Toffolo, E., Maroli, A., Landsberger, N., Bedogni, F., Turco, E., Pattini, L. *et al.* (2015) LSD1 neurospecific alternative splicing controls neuronal excitability in mouse models of epilepsy. *Cereb Cortex*, **25**, 2729–2740.
10. Saleque, S., Kim, J., Rooke, H.M. and Orkin, S.H. (2007) Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. *Mol. Cell*, **27**, 562–572.
11. Hojfeldt, J.W., Agger, K. and Helin, K. (2013) Histone lysine demethylases as targets for anticancer therapy. *Nat. Rev. Drug Discov.*, **12**, 917–930.
12. Chong, J.X., Yu, J.H., Lorentzen, P., Park, K.M., Jamal, S.M., Tabor, H.K., Rauch, A., Saenz, M.S., Boltshauser, E., Patterson, K.E. *et al.* (2015) Gene discovery for Mendelian conditions via social networking: de novo variants in KDM1A cause developmental delay and distinctive facial features. *Genet. Med.*, 10.1038/gim.2015.161.
13. Tunovic, S., Barkovich, J., Sherr, E.H. and Slavotinek, A.M. (2014) De novo ANKRD11 and KDM1A gene mutations in a male with features of KBG syndrome and Kabuki syndrome. *Am. J. Med. Genet. A*, **164**, 1744–1749.

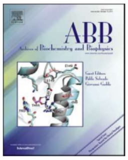
14. Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N. *et al.* (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*, **380**, 1674–1682.
15. Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G. *et al.* (2009) The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.*, **41**, 125–129.
16. Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I. *et al.* (2007) Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature*, **446**, 882–887.
17. Forneris, F., Binda, C., Vanoni, M.A., Mattevi, A. and Battaglioli, E. (2005) Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett.*, **579**, 2203–2207.
18. Li, M., Binda, C., Mattevi, A. and Edmondson, D.E. (2006) Functional role of the “aromatic cage” in human monoamine oxidase B: structures and catalytic properties of Tyr435 mutant proteins. *Biochemistry*, **45**, 4775–4784.
19. Pilotto, S., Speranzini, V., Tortorici, M., Durand, D., Fish, A., Valente, S., Forneris, F., Mai, A., Sixma, T.K., Vachette, P. *et al.* (2015) Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 2752–2757.
20. Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T. and Shi, Y. (2005) Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell*, **19**, 857–864.
21. Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.A., Cooch, N. and Shiekhattar, R. (2006) Functional interplay between histone demethylase and deacetylase enzymes. *Mol. Cell. Biol.*, **26**, 6395–6402.
22. Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., Ginelli, E., Mattevi, A., Sala, C. and Battaglioli, E. (2010) Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system. *J. Neurosci.*, **30**, 2521–2532.
23. Forneris, F., Binda, C., Vanoni, M.A., Battaglioli, E. and Mattevi, A. (2005) Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem.*, **280**, 41360–41365.
24. Toffolo, E., Rusconi, F., Paganini, L., Tortorici, M., Pilotto, S., Heise, C., Verpelli, C., Tedeschi, G., Maffioli, E., Sala, C. *et al.* (2014) Phosphorylation of neuronal Lysine-Specific Demethylase 1LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2. *J. Neurochem.*, **128**, 603–616.
25. Lee, M.G., Wynder, C., Cooch, N. and Shiekhattar, R. (2005) An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature*, **437**, 432–435.
26. Shen, C., Wang, D., Liu, X., Gu, B., Du, Y., Wei, F.Z., Cao, L.L., Song, B., Lu, X., Yang, Q. *et al.* (2015) SET7/9 regulates cancer cell proliferation by influencing beta-catenin stability. *FASEB J.*, **29**, 4313–4323.
27. Cano, A., Perez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F. and Nieto, M.A. (2000) The transcription factor Snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol.*, **2**, 76–83.
28. Baron, R., Binda, C., Tortorici, M., McCammon, J.A. and Mattevi, A. (2011) Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1-CoREST complex. *Structure*, **19**, 212–220.
29. Lin, Y., Wu, Y., Li, J., Dong, C., Ye, X., Chi, Y.I., Evers, B.M. and Zhou, B.P. (2010) The SNAG domain of Snail1 functions as a molecular hook for recruiting lysine-specific demethylase 1. *EMBO J.*, **29**, 1803–1816.
30. Barrallo-Gimeno, A. and Nieto, M.A. (2009) Evolutionary history of the snail/scratch superfamily. *Trends Genet.*, **25**, 248–252.
31. Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M. and Young, R.A. (2012) Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*, **482**, 221–225.
32. Foster, C.T., Dovey, O.M., Lezina, L., Luo, J.L., Gant, T.W., Barlev, N., Bradley, A. and Cowley, S.M. (2010) Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol. Cell. Biol.*, **30**, 4851–4863.
33. Nottke, A., Colaiacovo, M.P. and Shi, Y. (2009) Developmental roles of the histone lysine demethylases. *Development*, **136**, 879–889.
34. Thambyrajah, R., Mazan, M., Patel, R., Moignard, V., Stefanska, M., Marinopoulou, E., Li, Y., Lancrin, C., Clapes, T., Moroy, T. *et al.* (2016) GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat. Cell Biol.*, **18**, 21–32.
35. Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhattar, R., Bedford, M.T., Jenuwein, T. *et al.* (2007) p53 is regulated by the lysine demethylase LSD1. *Nature*, **449**, 105–108.
36. Laurent, B., Ruitu, L., Murn, J., Hempel, K., Ferrao, R., Xiang, Y., Liu, S., Garcia, B.A., Wu, H., Wu, F. *et al.* (2015) A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol. Cell*, **57**, 957–970.
37. Wang, J., Telese, F., Tan, Y., Li, W., Jin, C., He, X., Basnet, H., Ma, Q., Merkurjev, D., Zhu, X. *et al.* (2015) LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control. *Nat. Neurosci.*, **18**, 1256–1264.
38. Wasson, J.A., Simon, A.K., Myrick, D.A., Wolf, G., Driscoll, S., Pfaff, S.L., Macfarlan, T.S. and Katz, D.J. (2016) Maternally provided LSD1/KDM1A enables the maternal-to-zygotic transition and prevents defects that manifest postnatally. *Elife*, **5**, 10.7554/eLife.08848.
39. Forneris, F., Binda, C., Adamo, A., Battaglioli, E. and Mattevi, A. (2007) Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J. Biol. Chem.*, **282**, 20070–20074.
40. Forneris, F., Orru, R., Bonivento, D., Chiarelli, L.R. and Mattevi, A. (2009) ThermoFAD, a Thermofluor-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J.*, **276**, 2833–2840.
41. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
42. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Struct. Biol.*, **67**, 235–242.
43. Rusconi, F., Mancinelli, E., Colombo, G., Cardani, R., Da Riva, L., Bongarzone, I., Meola, G. and Zippel, R. (2010) Proteome profile in myotonic dystrophy type 2 myotubes reveals dysfunction in protein processing and mitochondrial pathways. *Neurobiol. Dis.*, **38**, 273–280.
44. Karplus, P.A. and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science*, **336**, 1030–1033.

Archives of Biochemistry and Biophysics 632 (2017) 202–208

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi



Probing the interaction of the p53 C-terminal domain to the histone demethylase LSD1

Valentina Speranzini [1], Giuseppe Ciossani [2], Chiara Marabelli, Andrea Mattevi*

Department of Biology and Biotechnology, University of Pavia, 27100 Pavia, Italy

ARTICLE INFO

Article history:
Received 21 June 2017
Received in revised form
10 July 2017
Accepted 13 July 2017
Available online 4 August 2017

Keywords:
Histone demethylase
p53
Protein-protein interaction
Post-translational modification
Substrate specificity
Electrostatic interactions

ABSTRACT

The p53 transcription factor plays a central role in the regulation of the expression of several genes, and itself is post-translationally regulated through its different domains. Of particular relevance for p53 function is its intrinsically disordered C-terminal domain (CTD), representing a hotspot for post-translational modifications and a docking site for transcriptional regulators. For example, the histone H3 lysine demethylase 1 (LSD1) interacts with p53 via the p53-CTD for mutual regulation. To biochemically and functionally characterize this complex, we evaluated the *in vitro* interactions of LSD1 with several p53-CTD peptides differing in length and modifications. Binding was demonstrated through thermal shift, enzymatic and fluorescence polarization assays, but no enzymatic activity could be detected on methylated p53-CTD peptides *in vitro*. These experiments were performed using the wild-type enzyme and LSD1 variants that are mutated on three active-site residues. We found that LSD1 demethylase activity is inhibited by p53-CTD. We also noted that the association between the two proteins is mediated by mostly non-specific electrostatic interactions involving conserved active-site residues of LSD1 and a highly charged segment of the p53-CTD. We conclude that p53-CTD inhibits LSD1 activity and that the direct association between the two proteins can contribute to their functional cross-talk.

1. Introduction

Histone modifications play a crucial role in tuning many biological processes, and methylation is one of the most studied and characterized thus far [1]. This modification regulates in a dynamic way different important pathways including chromatin accessibility, transcription, DNA repair, cell cycle and development [2–4]. In the past decade, new researches indicate that the enzymes responsible for the modulation of histone methylation can function on non-histone substrates in different environments other than chromatin [5–7]. This concept is exemplified by LSD1/KDM1A, the first lysine demethylase discovered and, to date, one of the most studied [8,9]. It has been shown that the non-catalytic domains within LSD1 and specific interactions with different partners determine the site-specific targeting and modulate demethylation activity, suggesting that this enzyme can also modify non-histone proteins [10,11]. For example, the introduction of an extra four-amino acid loop following alternative splicing is involved in the switch of substrate specificity of LSD1-containing complexes from H3Lys4 to H4Lys20 [12,13]. Similarly, it has been reported that LSD1-interacting proteins can change substrate specificity, as in the case of binding to androgen receptor and estrogen-related receptor α shown to form a demethylation complex [14,15]. Along this line, it has also been found that transcription factors can interact with LSD1 by mimicking the H3 tail as illustrated by the SNAIL1 family [16]. These proteins bind the active site of LSD1, which is recruited to specific loci, where the demethylase becomes engaged in specific chromatin complexes to selectively modify target gene(s). It is therefore clear that the binding and catalytic properties of LSD1 active site are versatile and can be finely tuned by a number of factors, enabling the demethylase to take part in very diverse processes of cell function, differentiation, and disease.

Many studies have reported that the tumour suppressor p53 is

Abbreviations: CTD, C-terminal domain; p53-CTD, C-terminal domain of p53; TAMRA, carboxytetramethylrhodamine; DAB, diaminobutyric acid.

* Corresponding author. Department of Biology and Biotechnology, University of Pavia, via Ferrata 9/A, 27100 Pavia, Italy.

E-mail address: andrea.mattevi@unipv.it (A. Mattevi).

[1] Present address: European Molecular Biology Laboratory, Grenoble Outstation, Grenoble Cedex 9, France.

[2] Present address: Department of Mechanistic Cell Biology, Max-Planck Institute of Molecular Physiology, Dortmund, Germany.

http://dx.doi.org/10.1016/j.abb.2017.07.021
0003-9861/

functionally associated to LSD1 for mutual regulation in the context of DNA damage response and cell death [17–20]. A wealth of cellular, biochemical and structural data have dissected the functions of the different domains within p53, generating a comprehensive and detailed picture of their functions [21–24]. Of relevance for this work, it has been largely reported that the role of the p53 C-terminal domain (CTD; residues 353–393 shown in Fig. 1A) is to stabilize the protein, to contribute to non-specific DNA-binding, and to recruit co-factor proteins, many being chromatin-associated such as the oncoprotein SET, p300/CBP histone acetyltransferase, and Mdm2 E3 ligase [25–27]. However, the data regarding the CTD have not yet clarified the precise function of this flexible domain, and fully reached a consensus model on how it may regulate p53 function. In this context, LSD1 was proposed to remove mono- and di-methylation on Lys370 on the CTD of p53, therefore blocking its interaction with the co-activator 53BP1 and the subsequent triggering of apoptosis in damaged cells [28]. More generally, LSD1 was also shown to be targeted to chromatin at gene-specific sites thanks to direct binding to p53, leading for example to the repression of transcription of alpha fetoprotein in hepatocytes [17].

A biochemical characterisation of this functionally crucial LSD1-p53 interaction is the focus of this study. Using several complementary assays [29–32], we find that a well-defined segment of p53-CTD is capable of binding to LSD1 active site, thereby inhibiting the enzymatic activity of the demethylase. This association is mostly electrostatic in nature and likely involves an active-site region that has recently been demonstrated to be targeted by a new class of LSD1 inhibitors.

2. Materials and methods

2.1. Protein expression and purification

Protein expression and co-purification of LSD1-CoREST1 (wild-type and mutants) heterodimer were performed following established protocols as previously described [29,31]. Protein complex was stored in 25 mM KH_2PO_4 pH 7.2, 5% glycerol.

2.2. LSD1 activity and inhibition assays

Custom p53 peptides were synthesized and purchased from Thermo Scientific and ChinaPeptides. Enzyme inhibition was evaluated using a horseradish peroxidase-coupled enzymatic assay to monitor hydrogen peroxide formation during demethylation [29]. The reaction mixture contained 50 mM HEPES/NaOH pH 7.5, 0.1 μM Amplex Red, 2.8 μM horseradish peroxidase, 0.1 μM LSD1-CoREST1, varying p53-derived peptide concentrations (in the range 0–1000 μM) and monomethylated H3K4 peptide substrate (0–15 μM concentration range). Fluorescence changes associated to the conversion of Amplex Red to Resorufin were monitored at 535 nm and 590 nm for excitation and emission, respectively. A formaldehyde dehydrogenase-coupled assay that monitors formation of formaldehyde was also employed to double-check for activity on p53-methylated peptides [8]. All experiments were performed at 25 °C using a CLARIOstar plate reader (BMG Labtech, Germany). Initial velocity values were fitted to the equations below:

$$v_0 = \frac{V_{maxapp} * [S]}{Km_{app} + [S]} \tag{1.1}$$

$$V_{maxapp} = \frac{V_{max}}{1 + \frac{[I]}{\alpha Ki}} \tag{1.2}$$

$$Km_{app} = \frac{Km * \left(1 + \frac{[I]}{Ki}\right)}{1 + \frac{I}{\alpha Ki}} \tag{1.3}$$

A

p53$_{(353)}$ AQAGKEPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD $_{(393)}$

H3$_{(1)}$ ARTKQTARKSTGGKAPRKQLA $_{(21)}$

SNAIL1$_{(1)}$ PRSFLVRKP $_{(9)}$

B

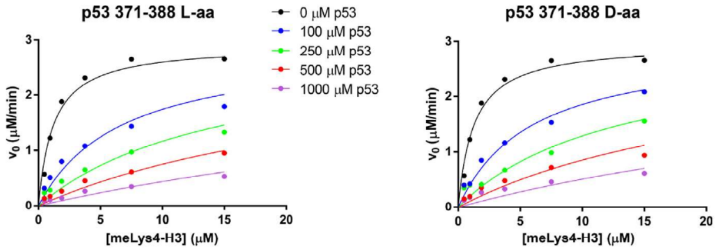


Fig. 1. Binding of p53-CTD to LSD1. (A) Sequences of the peptides used in this study. The C-terminal domain of human p53 (residues 353–393; p53-CTD) is subject to extensive post-translational modifications on Lys and Ser residues, which are highlighted in red and blue, respectively [21–24]. The monomethylLys4 peptide comprising residues 1–21 of H3 (monomethylLys4-H3) is employed as reference substrate for all enzymatic assays [16]. SNAIL comprises residues 1–9 of the transcription factor SNAIL1, whose N-terminal tail tightly associates to the LSD1 active site without undergoing demethylation (Phe rather than a Lys in position 4). SNAIL peptide was employed for competitive binding assays. (B) Michaelis–Menten plots for the inhibition of LSD1-CoREST1 activity on monomethylLys4-H3 by p53-derived peptides at variable peptide (0–1000 μM) and substrate (0–15 μM) concentrations. Full substitution of L-amino acids (left panel) with D-amino acids (right panel) does not alter significantly binding of the peptide 371–388 to LSD1 active site (Table 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In this model, α is the constant that will determine the type of inhibition (α would be $= 1$ for pure noncompetitive inhibition, >1 for mixed noncompetitive inhibition, $>>1$ for competitive inhibition, $0<\alpha<<1$ for uncompetitive inhibition). Values for the kinetic parameters along with their associated errors were obtained with GraphPad Prism inbuilt analysis tools (see Supplementary Information).

2.3. Thermal shift assay

Qualitative binding of p53-derived peptides to LSD1-CoREST1 was assessed using ThermoFAD [30]. The assay was performed in LSD1 storage buffer at final protein concentration of 4 μM and peptide concentration of 200 μM.

2.4. Fluorescence polarization binding assays

Binding of p53 peptides to LSD1-CoREST1 were evaluated monitoring the change in polarization properties of SNAIL1 (1–9) peptide fluorescently labeled with 5(6)-carboxytetramethylrhodamine (TAMRA) using previously reported procedures [32]. Briefly, experiments were typically carried out in 15 mM KH_2PO_4 pH 7.2, 1 mg/ml BSA, 5% glycerol at 25 °C. CLARIOstar (BMG Labtech, Germany) plate reader was used with 540 nm excitation and 590 nm emission filters. Experiments were performed in triplicates using 384-well microplates (CORNING, UK). Direct binding of SNAIL-TAMRA at a final concentration of 10 nM was performed to calculate K_d and proceed with competitive binding assays. For competitive experiments, each well contained LSD1-CoREST1 (50 nM) and labeled peptide (10 nM). Next, decreasing concentrations (typically in the 0–500 μM concentration range) of the unlabeled competing peptides were added. For direct and competitive experiments assessing effect of ionic strength on ligand binding, the reaction buffer was supplemented with 50-100-200 mM KCl. In all experiments, total fluorescence intensity was also analyzed to ensure that the final polarization signal would not be affected by formation of colloidal aggregates in the mix.

3. Results

Our studies started from the reported observations that histone demethylase LSD1 and tumour suppressor p53 interact for reciprocal regulation [17,28]. Upon binding of LSD1 to p53-CTD, lysine 370 of p53 would become demethylated, resulting in a downregulation of p53 activity through a weakened interaction with 53BP1. LSD1 is in return redirected to specific chromatin loci, triggering gene expression modulation via H3Lys4 demethylation. Lysine 370 is part of p53-CTD, which is characterized by an enrichment of positively charged residues and represents one of the two disordered regions within the tumour suppressor protein (Fig. 1A) [33]. We first wanted to explore the catalytic activity of LSD1 in complex with its co-repressor CoREST1 on this p53-Lys370 residue *in vitro*. To this aim, we employed established enzymatic assays with three CTD-derived peptides dimethylated at Lys370 (residues 353–388; 363–388; 353–378). Although these same peptides were used by other authors for pull-down and co-immunoprecipitation experiments [28], in our hands no evident catalytic Lys370-demethylation activity was observed. We employed two spectroscopic assays that measure hydrogen peroxide and formaldehyde production, respectively. In either case no detectable change in absorbance was observed over a time of 20 min with substrate concentrations up to 400 μM (i.e. estimated lower limit of detection is k_{cat} ~0.001 min^{-1}) [34]. This result was replicated using no less than five different LSD1-CoREST1 enzyme preparations which were fully active in H3Lys4 demethylation (see Fig. 1B, black curves). Nonetheless, we found that the incubation with both dimethylLys370 353–388 and 363–388 peptides as well as unmodified 363–388 and 363–393 peptides lead to a 2 °C increase in the unfolding temperature as measured by a thermal shift assay [30], suggesting binding to the enzyme. Consistently, we observed that these peptides inhibit the demethylase activity on methylLys4-H3 peptide (data shown in Fig. 1B and Fig. S1 and summarized in Table 1 and Supplementary Tables 1 and 2). These findings supported the idea that, though not a direct LSD1 substrate, p53 can physically interact with LSD1-CoREST1 through the CTD residues, hampering the histone demethylation catalytic activity.

We then aimed at characterizing a minimal unit within p53-CTD for interaction with LSD1-CoREST1 and the factors that may modulate this interaction. To do so, we expanded the set of peptides to a total of twelve CTD-derived sequences and probed their ability to compete with histone H3 tail to bind to the LSD1 active site, therefore inhibiting its enzymatic activity (Fig. 1B and Table 1). These peptides varied in sequence length, covering differing segments within residues 358–393 of p53-CTD, and bore different mutations or amino acid modifications that possibly modulate binding properties [35–37]. With respect to modifications, we analyzed Ser phosphorylation, Lys methylations, and Lys acetylations which are known to occur in the post-translationally modified p53 as detailed in Table 1 (Fig. 1A) [21,22]. We found that the strongest binding was obtained with the peptide covering the amino acids 363–388 (K_i of 6.4 μM) and that the overall lengths of the peptides do not greatly influence binding affinity as long as the sequence 379–388 is retained (Fig. 1A and Table 1). Indeed, peptide 353–378 does not show detectable inhibition of LSD1-CoREST1 whereas the 379–388 peptide retains an affinity similar to that comprising residues 371–388 (K_i of 22 μM). Consistent with the fact that the 379–388 segment is the key player for the association of p53-CTD to LSD1-COREST1, Lys370 methylation does not strongly affect binding causing only a two-fold increase in K_i from 6.4 μM to 11.4 Conversely, significant (though not drastic) effects were observed for side chain modifications targeting the core segment. Binding is decreased by Lys382 and Lys386 acetylation, Ser378 phosphorylation, and Lys381Met mutation.

We also evaluated the effect of varying ionic strength. These experiments included previously characterized H3 (residues 1–21) and SNAIL (residues 1–9) peptides [8,34] to allow a comparative analysis with different ligands (Fig. 1A). As shown in Fig. 2 and listed in Table 2, it is evident that a high sensitivity to salt concentration is a feature shared by all these peptides in their association to LSD1. Specifically, binding of SNAIL1 is the least perturbed (15–20 times Kd decrease at 200 mM salt), while p53 seems most affected (75–80 times Kd decrease at 200 mM salt). Collectively, these experiments gave a clear indication on the features characterizing p53-CTD/LSD1-CoREST1 interaction: binding is mostly electrostatic, mediated by residues 379–388, and moderately affected by site-specific variations on these amino acids (Fig. 2 and Table 2).

To further probe whether the binding was due to non-specific interactions, a p53 peptide (residues 371–388) consisting of D- (rather than L-) amino acid was used in our studies. Surprisingly, it turned out that this D-peptide does inhibit LSD1-CoREST1 equally well as the L-amino acid peptide (Table 1, Fig. 1B). This intriguing result was validated by fluorescence polarization using a SNAIL1 peptide (residues 1–9) conjugated with a specific fluorophore to assess competitive binding properties of the D- and L-p53-CTD peptides. We were able to confirm the binding constants obtained with the enzyme inhibition assays and we also verified that there is no significant difference between L-amino acid and D-amino acid ligands (Fig. 1B and Table 1).

Table 1
K_i values for p53-derived peptides on LSD1 activity.

p53 Peptide	Sequence	K_i (μM)[a,b]
Unmodified		
363-388	RAHSSHLKSKKGQSTSRHKKLMFKTE	6.4 ± 1.4
363-393	RAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD	13.1 ± 2.3
379-388	RHKKLMFKTE	35.1 ± 7.6
Methylated		
353–388 - DimethylK370	AQAGKEPGGSRAHSSHL$\mathbf{K}_{\mathbf{me2}}$SKKGQSTSRHKKLMFKTE	10.6 ± 2.5
363–388 - DimethylK370	RAHSSHL$\mathbf{K}_{\mathbf{me2}}$SKKGQSTSRHKKLMFKTE	11.4 ± 2.9
353–378 - DimethylK370	AQAGKEPGGSRAHSSHL$\mathbf{K}_{\mathbf{me2}}$SKKGQSTS	>150 (>120)[b]
Acetylated		
363–388 - Acetyl-K373, K382	RAHSSHLKSK$\mathbf{K}_{\mathbf{Ac}}$GQSTSRHK$\mathbf{K}_{\mathbf{Ac}}$LMFKTE	29.3 ± 3.9
371–388 - Acetyl-K386	SKKGQSTSRHKKLMF$\mathbf{K}_{\mathbf{Ac}}$TE	40.5 ± 10.4
Phosphorylated		
371–388 - Phospho-S378	SKKGQST$\mathbf{S}_{\mathbf{Ph}}$RHKKLMFKTE	44.0 ± 6.2
Mutants		
371–388 - R379A	SKKGQSTS**A**HKKLMFKTE	12.6 ± 4.1
371–388 - K381M	SKKGQSTSRH**M**KLMFKTE	45.5 ± 8.9
Chirality		
371-388 L-aa	SKKGQSTSRHKKLMFKTE	22.0 ± 2.8 (24.4 ± 5.4)[b]
D-aa	SKKGQSTSRHKKLMFKTE	27.0 ± 3.7 (30.0 ± 3.8)[b]

[a] Inhibition of catalytic activity by fluorescent assay on histone-derived peptide monomethylLys4-H3 peptide (residues 1–21; Fig. 1A). Data fitting yielded large α values (>9) indicating a competitive type of inhibition (see Methods, Tables S1–S2, and Fig. S1).

[b] Competitive fluorescence polarization assay on SNAIL-derived peptide 1–9 (Fig. 1A).

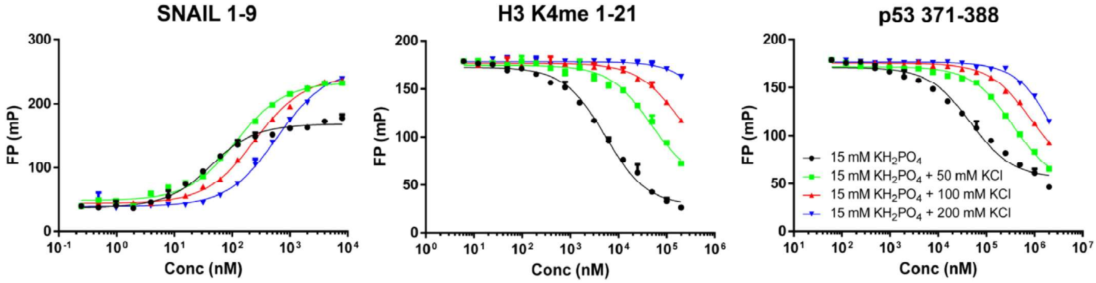


Fig. 2. Effect of ionic strength on peptide binding to LSD1-COREST1. Binding properties of LSD1 to SNAIL (left), histone H3-derived peptide (mid) and p53-CTD 371–388 (right) at different ionic strength (0–200 mM KCl in 15 mM KH_2PO_4, pH 7.2). In the case of SNAIL, increasing concentrations of purified LSD1-CoREST1 (0.25 nM-8 μM) were incubated directly with TAMRA-conjugated SNAIL1 peptide, whereas binding to histone H3 N-terminal tail and p53-CTD were measured in competition against SNAIL. Changes in polarization were measured in millipolarization units (mP) and plotted against peptide concentration for K_d determination (Table 2).

Table 2
Effect of ionic strength on peptides binding to LSD1-CoREST1.

Buffer	K_d (μM)		
	SNAIL 1–9[a]	monomethylLys4-H3[b]	p53-CTD 371–388[b]
15 mM KH_2PO_4	0.03 ± 0.01	2.6 ± 0.3	24.4 ± 5.4
15 mM KH_2PO_4 + 50 mM KCl	0.11 ± 0.01	23.5 ± 3.9	186.0 ± 22.0
15 mM KH_2PO_4 + 100 mM KCl	0.24 ± 0.01	83.7 ± 25.0	462.0 ± 54.0
15 mM KH_2PO_4 + 200 mM KCl	0.58 ± 0.03	>200	>2000

[a] Fluorescence polarization direct binding (Fig. 2, left panel).

[b] Fluorescence polarization competitive binding (Fig. 2, middle and right panels).

We further analyzed the binding of the p53 peptides to LSD1 by including in our studies three mutants, which were recently discovered in patients suffering by a newly described genetic disease [31]. Of interest, all three disease-associated mutations (E379K, D556G, and Y761H) affect charged or H-bonding side chains in the active site (Fig. 3). Although they do not perturb the protein conformation, they substantially impair catalysis, primarily by negatively affecting H3 binding. The affinity to p53-CTD turned out to be unmodified by Y761H and D556G mutations whereas charge-reverted E379K caused a significant 10-fold decrease in CTD binding (data shown in Fig. 4 and summarized in Table 3). Thus, it appears that the binding of p53-CTD to LSD1-CoREST1 is strongly affected by a localized mutation (E379K) that targets an active-site residue.

This pattern of observations resembles the case of polymyxins [38]. These polycationic cyclic peptides comprise a cyclic heptapeptide loop with a DAB-DAB-Leu(or Phe)-Leu-DAB-DAB-Thr sequence that resembles the ^{381}Lys-Lys-Leu-Met-Phe-Lys-Thr^{387} segment of p53 [38,39]. Polymyxins bind to LSD1 with much higher nanomolar affinity than p53-CTD. The tighter binding probably

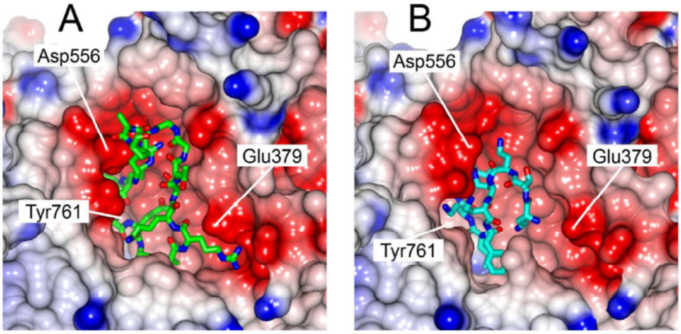


Fig. 3. LSD1-CoREST1 ligands bind to the active site cleft through extensive interactions with negatively charged surface. (A) Histone H3 (PDB: 2V1D) [29] and (B) Polymyxin B bound to LSD1-CoREST1 (PDB: 5L3F) [38]. LSD1-CoREST1 is represented as surface, which is colored according to electrostatic potential from red (negative) to blue (positive). Ligands bound to the active site are shown as sticks. Carbons are in green for H3 (panel A) and cyan for Polymyxin B (panel B); the other atoms are colored by atom type using standard color codes. Labels indicate the location of LSD1 pathological mutants used in this study. Both images were generated using built-in functions in CCP4MG [45]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

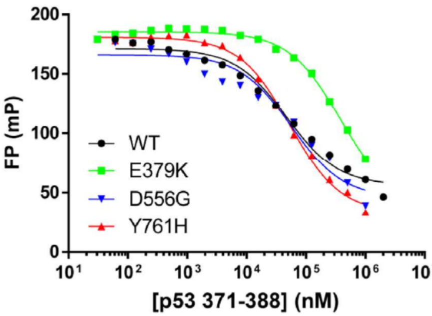


Fig. 4. Effect of LSD1 mutations on p53-CTD binding. Competitive binding of p53-CTD peptide were measured against SNAIL using fluorescence polarization for LSD1-CoREST1 wild-type and mutants. Changes in polarization were measure in milli-polarization units (mP) and plotted against peptide concentration for K_d determination (Table 3).

Table 3
Effect of LSD1 mutations on p53 371–388 binding.[a]

LSD1 mutant	K_d (μM)	
	p53-CTD 371-388	monomethylLys4-H3[b]
WT	24.4 ± 5.4	3.0 ± 0.3
E379K	195.2 ± 59.3	>150
D556G	16.1 ± 6.2	3.8 ± 1.3
Y761H	26.3 ± 9.4	9.8 ± 4.1

[a] Measured by fluorescence polarization competitive assay as shown in Fig. 4.
[b] Data for the H3 peptide are shown as a reference and taken from Ref. [31].

reflects the more elaborated and conformationally restrained cyclic structure of polymyxins compared to the highly flexible p53-CTD sequence [35–37]. Bearing in mind these differences, it is notable that, as now found for p53-CTD, the affinity of polymyxins to LSD1 is greatly diminished by E379K [38]. Polymyxins were found to bind at the entrance of the enzyme catalytic cleft where Glu379 is located (Fig. 3B). This region likely represents the site that electrostatically steers the H3 tail to favour substrate admission and ordered binding to the depth of the catalytic site where flavin-mediated demethylation takes place [38]. In light of the observation that E379K specifically perturbs the binding of both the highly charged polymyxin and p53-CTD peptides, it seems that also p53-CTD exploits this negatively charged area on the active-site outer surface for its binding to LSD1 (Fig. 3).

4. Discussion

The intrinsically disordered C-terminal domain of p53 is characterized by the presence of charge-rich sub-regions that are an ideal target for a large number of post-translational modifications by diverse enzymes, such as p300/CBP acetyltransferases, Mdm2 ubiquitin ligase and protein kinases [25–27] (Fig. 1A). Consistently, alteration of the degree and combination of these modifications influences the binding of p53 to its regulators [40,41]. Furthermore, the positively charged CTD contributes to DNA binding, supposedly though electrostatic interactions between Lys, Arg and the phosphate backbone of DNA [42]. Interestingly, a recently published work on the interaction between p53-CTD and the oncoprotein SET indicates this charge-dependency as "*widespread regulatory mode*" [43]. In this context, our experiments were motivated by several published reports demonstrating a functional cross-talk between p53 and LSD1, whose molecular bases were suggested to involve a direct interaction between p53-CTD with the demethylase enzyme [17–20,28,44]. To tackle the issue of the biochemical nature of this association we have performed complementary assays which probe binding to LSD1 by enzyme activity, thermal shift, enzyme inhibition and binding assays using a 21-amino acid H3 tail substrate, and fluorescence polarization using a 9-amino acid binder. Our experiments demonstrate that LSD1-CoREST does not show a catalytic demethylase activity on p53-CTD (at least on Lys370) *in vitro*. On the other hand, they also coherently indicate that there is a micromolar-affinity association between the LSD1-CoREST1 and p53-CTD and that binding can be mostly ascribed to residues 379–388 of p53-CTD (Fig. 1A). Modifications of the Lys side chains on this CTD region reduce binding affinity, which is consistent with the idea that the interaction is mainly enforced by electrostatic attraction. The clear dependency of the affinity on the ionic strength further confirms this notion (Table 2).

The finding that a peptide consisting of D-amino acids binds to LSD1-CoREST1 with equal affinity as the L-amino acid ligand indicates that the assays detect a generally non-specific association between two highly charged molecules. Nonetheless, p53-CTD is an effective inhibitor of LSD1 and its binding to the enzyme is markedly and specifically reduced by a charge-reversal mutation

(E379K), which is located on the outer rim of the LSD1 active site (Fig. 3). This protein region is rich in negatively-charged residues and was previously shown to represent the binding site for the potent polymyxin inhibitors [38]. Of notice, these peptide-like molecules feature a sequence that closely resembles the LSD1-binding sequence of p53-CTD. Collectively, these findings indicate that p53-CTD and the active site of LSD1 can be envisioned to form an encounter complex whose association is mostly mediated by non-specific electrostatic interactions. Such an encounter complex can potentially further promote more specific interactions between p53 and nucleosomal DNA and/or other proteins associated to LSD1-COREST1 core complex. This complex would also lead to the recruitment of LSD1 and associated partners such as histone deacetylases onto p53 target genes, inducing demethylation and/or other chromatin modifications in agreement with previous literature [17–19,44]. At the same time, engagement by LSD1 can mask the positive charges on the CTD, thereby hampering the interaction of p53 with 53BP1 in agreement with published data [28]. Consistently with the idea of “association initiated by electrostatics”, the changes of post-translational modifications (i.e. change of surface charges) within the CTD represent an ideal tool to modulate binding affinities among several possible partners that can associate to the CTD. The essence of these processes is that a plethora of low-affinity binding events can occur, which are nonetheless finely tuned and all together contribute to the correct targeting.

Acknowledgments

This work was supported by AIRC (IG-15208) and MIUR (Progetto Epigen).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.abb.2017.07.021.

References

[1] S.B. Rothbart, B.D. Strahl, Interpreting the language of histone and DNA modifications, Biochim. Biophys. Acta 1839 (2014) 627–643.
[2] H. Alam, B. Gu, M.G. Lee, Histone methylation modifiers in cellular signaling pathways, Cell. Mol. Life Sci. 72 (2015) 4577–4592.
[3] A.J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications, Cell Res. 21 (2011) 381–395.
[4] R. DesJarlais, P.J. Tummino, Role of histone-modifying enzymes and their complexes in regulation of chromatin biology, Biochemistry 55 (2016) 1584–1599.
[5] R. Hamamoto, V. Saloura, Y. Nakamura, Critical roles of non-histone protein lysine methylation in human tumorigenesis, Nat. Rev. Cancer 15 (2015) 110–124.
[6] P. Martinez-Redondo, A. Vaquero, The diversity of histone versus nonhistone sirtuin substrates, Genes & cancer 4 (2013) 148–163.
[7] L. Peng, E. Seto, Deacetylation of nonhistone proteins by HDACs and the implications in cancer, Handb. Exp. Pharmacol. 206 (2011) 39–56.
[8] F. Forneris, C. Binda, M.A. Vanoni, A. Mattevi, E. Battaglioli, Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process, FEBS Lett. 579 (2005) 2203–2207.
[9] Y.J. Shi, F. Lan, C. Matson, P. Mulligan, J.R. Whetstine, P.A. Cole, R.A. Casero, Y. Shi, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1, Cell 119 (2004) 941–953.
[10] T.B. Nicholson, T. Chen, LSD1 demethylates histone and non-histone proteins, Epigenetics 4 (2009) 129–132.
[11] X. Zhang, Y. Huang, X. Shi, Emerging roles of lysine methylation on non-histone proteins, Cell. Mol. Life Sci. 72 (2015) 4257–4272.
[12] J. Wang, F. Telese, Y. Tan, W. Li, C. Jin, X. He, H. Basnet, Q. Ma, D. Merkurjev, X. Zhu, Z. Liu, J. Zhang, K. Ohgi, H. Taylor, R.R. White, C. Tazearslan, Y. Suh, T.S. Macfarlan, S.L. Pfaff, M.G. Rosenfeld, LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control, Nat. Neurosci. 18 (2015) 1256–1264.
[13] C. Zibetti, A. Adamo, C. Binda, F. Forneris, E. Toffolo, C. Verpelli, E. Ginelli, A. Mattevi, C. Sala, E. Battaglioli, Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system, J. Neurosci. 30 (2010) 2521–2532.
[14] E. Metzger, M. Wissmann, N. Yin, J.M. Muller, R. Schneider, A.H. Peters, T. Gunther, R. Buettner, R. Schule, LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription, Nature 437 (2005) 436–439.
[15] J. Carnesecchi, C. Forcet, L. Zhang, V. Tribollet, B. Barenton, R. Boudra, C. Cerutti, I.M. Billas, A.A. Serandour, J.S. Carroll, C. Beaudoin, J.M. Vanacker, ERRalpha induces H3K9 demethylation by LSD1 to promote cell invasion, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 3909–3914.
[16] R. Baron, C. Binda, M. Tortorici, J.A. McCammon, A. Mattevi, Molecular mimicry and ligand recognition in binding and catalysis by the histone demethylase LSD1-CoREST complex, Structure 19 (2011) 212–220.
[17] W.W. Tsai, T.T. Nguyen, Y. Shi, M.C. Barton, p53-targeted LSD1 functions in repression of chromatin structure and transcription in vivo, Mol. Cell. Biol. 28 (2008) 5139–5146.
[18] A. Scoumanne, X. Chen, The lysine-specific demethylase 1 is required for cell proliferation in both p53-dependent and -independent manners, J. Biol. Chem. 282 (2007) 15471–15475.
[19] N. Mosammaparast, H. Kim, B. Laurent, Y. Zhao, H.J. Lim, M.C. Majid, S. Dango, Y. Luo, K. Hempel, M.E. Sowa, S.P. Gygi, H. Steen, J.W. Harper, B. Yankner, Y. Shi, The histone demethylase LSD1/KDM1A promotes the DNA damage response, J. Cell Biol. 203 (2013) 457–470.
[20] X. Li, T. Li, D. Chen, P. Zhang, Y. Song, H. Zhu, Y. Xiao, Y. Xing, Overexpression of lysine-specific demethylase 1 promotes androgen-independent transition of human prostate cancer LNCaP cells through activation of the AR signaling pathway and suppression of the p53 signaling pathway, Oncol. Rep. 35 (2016) 584–592.
[21] O. Laptenko, D.R. Tong, J. Manfredi, C. Prives, The tail that wags the dog: how the disordered C-Terminal domain controls the transcriptional activities of the p53 tumor-suppressor protein, Trends Biochem. Sci. 41 (2016) 1022–1034.
[22] N. Raj, L.D. Attardi, The transactivation domains of the p53 protein, Cold Spring Harb. Perspect. Med. 7 (2017).
[23] R. Kamada, Y. Toguchi, T. Nomura, T. Imagawa, K. Sakaguchi, Tetramer formation of tumor suppressor protein p53: structure, function, and applications, Biopolymers 106 (2016) 598–612.
[24] T. Saha, R.K. Kar, G. Sa, Structural and sequential context of p53: a review of experimental and theoretical evidence, Prog. Biophys. Mol. Biol. 117 (2015) 250–263.
[25] Y. Haupt, R. Maya, A. Kazaz, M. Oren, Mdm2 promotes the rapid degradation of p53, Nature 387 (1997) 296–299.
[26] J.Y. Kim, K.S. Lee, J.E. Seol, K. Yu, D. Chakravarti, S.B. Seo, Inhibition of p53 acetylation by INHAT subunit SET/TAF-Ibeta represses p53 activity, Nucleic Acids Res. 40 (2012) 75–87.
[27] J. Lin, M. Blake, C. Tang, D. Zimmer, R.R. Rustandi, D.J. Weber, F. Carrier, Inhibition of p53 transcriptional activity by the S100B calcium-binding protein, J. Biol. Chem. 276 (2001) 35037–35041.
[28] J. Huang, R. Sengupta, A.B. Espejo, M.G. Lee, J.A. Dorsey, M. Richter, S. Opravil, R. Shiekhattar, M.T. Bedford, T. Jenuwein, S.L. Berger, p53 is regulated by the lysine demethylase LSD1, Nature 449 (2007) 105–108.
[29] F. Forneris, C. Binda, A. Adamo, E. Battaglioli, A. Mattevi, Structural basis of LSD1-CoREST selectivity in histone H3 recognition, J. Biol. Chem. 282 (2007) 20070–20074.
[30] F. Forneris, R. Orru, D. Bonivento, L.R. Chiarelli, A. Mattevi, ThermoFAD, a Thermofluor-adapted flavin ad hoc detection system for protein folding and ligand binding, FEBS J. 276 (2009) 2833–2840.
[31] S. Pilotto, V. Speranzini, C. Marabelli, F. Rusconi, E. Toffolo, B. Grillo, E. Battaglioli, A. Mattevi, LSD1/KDM1A mutations associated to a newly described form of intellectual disability impair demethylase activity and binding to transcription factors, Hum. Mol. Gen. 25 (2016) 2578–2587.
[32] S. Pilotto, V. Speranzini, M. Tortorici, D. Durand, A. Fish, S. Valente, F. Forneris, A. Mai, T.K. Sixma, P. Vachette, A. Mattevi, Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 2752–2757.
[33] S. Bell, C. Klein, L. Muller, S. Hansen, J. Buchner, p53 contains large unstructured regions in its native state, J. Mol. Biol. 322 (2002) 917–927.
[34] M. Tortorici, M.T. Borrello, M. Tardugno, L.R. Chiarelli, S. Pilotto, G. Ciossani, N.A. Vellore, S.G. Bailey, J. Cowan, M. O'Connell, S.J. Crabb, G. Packham, A. Mai, R. Baron, A. Ganesan, A. Mattevi, Protein recognition by short peptide reversible inhibitors of the chromatin-modifying LSD1/CoREST lysine demethylase, ACs Chem. Biol. 8 (2013) 1677–1682.
[35] J.L. Avalos, K.M. Bever, C. Wolberger, Mechanism of sirtuin inhibition by nicotinamide: altering the NAD(+) cosubstrate specificity of a Sir2 enzyme, Mol. Cell 17 (2005) 855–868.
[36] M.S. Cosgrove, K. Bever, J.L. Avalos, S. Muhammad, X. Zhang, C. Wolberger, The structural basis of sirtuin substrate affinity, Biochemistry 45 (2006) 7511–7521.
[37] R.R. Rustandi, D.M. Baldisseri, D.J. Weber, Structure of the negative regulatory domain of p53 bound to S100B(betabeta), Nat. Struct. Biol. 7 (2000) 570–574.
[38] V. Speranzini, D. Rotili, G. Ciossani, S. Pilotto, B. Marrocco, M. Forgione, A. Lucidi, F. Forneris, P. Mehdipour, S. Velankar, A. Mai, A. Mattevi, Polymyxins and quinazolines are LSD1/KDM1A inhibitors with unusual structural features, Sci. Adv. 2 (2016) e1601017.
[39] T. Velkov, P.E. Thompson, R.L. Nation, J. Li, Structure–activity relationships of polymyxin antibiotics, J. Med. Chem. 53 (2010) 1898–1916.

[40] C.J. DeHart, J.S. Chahal, S.J. Flint, D.H. Perlman, Extensive post-translational modification of active and inactivated forms of endogenous p53, Mol. Cel. Proteomics 13 (2014) 1–17.

[41] M.V. Poyurovsky, C. Katz, O. Laptenko, R. Beckerman, M. Lokshin, J. Ahn, I.J. Byeon, R. Gabizon, M. Mattia, A. Zupnick, L.M. Brown, A. Friedler, C. Prives, The C terminus of p53 binds the N-terminal domain of MDM2, Nat. Struct. Biol. 17 (2010) 982–989.

[42] A. Friedler, D.B. Veprintsev, S.M. Freund, K.I. von Glos, A.R. Fersht, Modulation of binding of DNA to the C-terminal domain of p53 by acetylation, Structure 13 (2005) 629–636.

[43] D. Wang, N. Kon, G. Lasso, L. Jiang, W. Leng, W.G. Zhu, J. Qin, B. Honig, W. Gu, Acetylation-regulated interaction between p53 and SET reveals a widespread regulatory mode, Nature 538 (2016) 118–122.

[44] L. Yi, Y. Cui, Q. Xu, Y. Jiang, Stabilization of LSD1 by deubiquitinating enzyme USP7 promotes glioblastoma cell tumorigenesis and metastasis through suppression of the p53 signaling pathway, Oncol. Rep. 36 (2016) 2935–2945.

[45] S. McNicholas, E. Potterton, K.S. Wilson, M.E. Noble, Presenting your structures: the CCP4mg molecular-graphics software, Acta Crystallogr. D. Biol. Crystallogr. 67 (2011) 386–394.

Chapter III – Nucleosome recognition by LSD2/NPAC

Despite sharing the same substrate and the same catalytic mechanism, the homologous flavin-dependent histone lysine demethylases LSD1 and LSD2 are recruited by different partners and play opposite biological roles. Differently from its counterpart LSD1/CoREST complex, very little is known about LSD2 and its partner NPAC.

To gain a comprehensive view on the physiological role of the LSD2/NPAC system within the transcription elongation machinery, multiple and diverse biochemical and structural approaches have been undertaken, for whom a detailed description is given in the “Materials and Methods” section. For simplicity, results have been grouped into two parts. In the first part, the effect of the NPAC-linker on LSD2 activity will firstly be addressed, whereas in part 2, the characterization of the other domains of NPAC will complete the scenario.

These studies revealed a novel mechanism for chromatin remodeling, in which the NPAC-linker sustains the H3 nucleosomal tail recognition by LSD2. NPAC tetramer holds multiple copies of the enzyme, and along with the nucleosome binding NPAC-PWWP, increases the processivity and avidity of the demethylase system.

Background

The LSD2/NPAC system

The physiological partner of LSD2 is a short linker peptide from the unknown protein NPAC. LSD2 and its partner interact in the very low micromolar range (0.92±0.08 µM; Fang *et al.*, 2013). Despite its small size compared to that of both LSD2 and its nucleosomal substrate, the NPAC-linker is sufficient to enhance 5-fold the demethylase activity on nucleosomes (Fang *et al.*, 2013). In the crystal structure of LSD2 in complex with the NPAC linker peptide and the N-terminal 26-residues H3 peptide determined by Chen and colleagues (Chen *et al.*, 2013), the NPAC-linker occupies a deep hydrophobic groove very close to the H3-interacting residues of LSD2 (Figure 1). NPAC is then expected to stabilize LSD2-H3 secondary docking site, through hydrogen bonds and hydrophobic interactions (Chen *et al.*, 2013). Yet, the only data available showed that LSD2 affinity for H3 1-21 peptide (0.99±0.06 µM) increases only by a factor of 1.5 in the presence of NPAC (0.68±0.07 µM; Fang *et al.*, 2013). Unravelling the mechanism of nucleosome recognition by LSD2/NPAC-linker system could be the only way to explain the molecular basis of the NPAC-linker effect on LSD2 activity.

Moreover, another aspect that was not deeply investigated is the physiological role of the LSD2 partner, namely the full-length 553-aa long NPAC protein. Comparison of the denominations of NPAC is also significative to get how poorly characterized this protein is. The alias GLYR1 (glyoxylate reductase homolog 1) is due to the sequence and structural homology of NPAC C-terminus with *A. thaliana* GLYR1 and GLYR2 enzymes. These two are known succinic semi-aldehyde/glyoxilate reductases playing a role in stress response within the plastids (GLYR2) or the cytoplasm (GLYR1). However, human NPAC localizes in the cell nucleus, from which the name NP60 (nuclear protein 60). From bioinformatic analysis it can be easily assessed there are at least three nuclear domains, plus two nuclear localization sequences (NLSs). The short LSD2-binding linker (residues 214-225) and typical DNA-binding AT-hook sequence (Reeves, 2001; Fonfría-Subirós *et al.*, 2012) localize within the central disordered region of the protein. The AT-hook motif is required for the interaction and the activation of p38α (Fu *et al.*, 2006). Nevertheless, in all experiments with p38α, it has been always used a longer NPAC sequence, hence, the site binding and activating this MAP kinase might be close to the AT-hook and not necessarily the AT-hook itself.

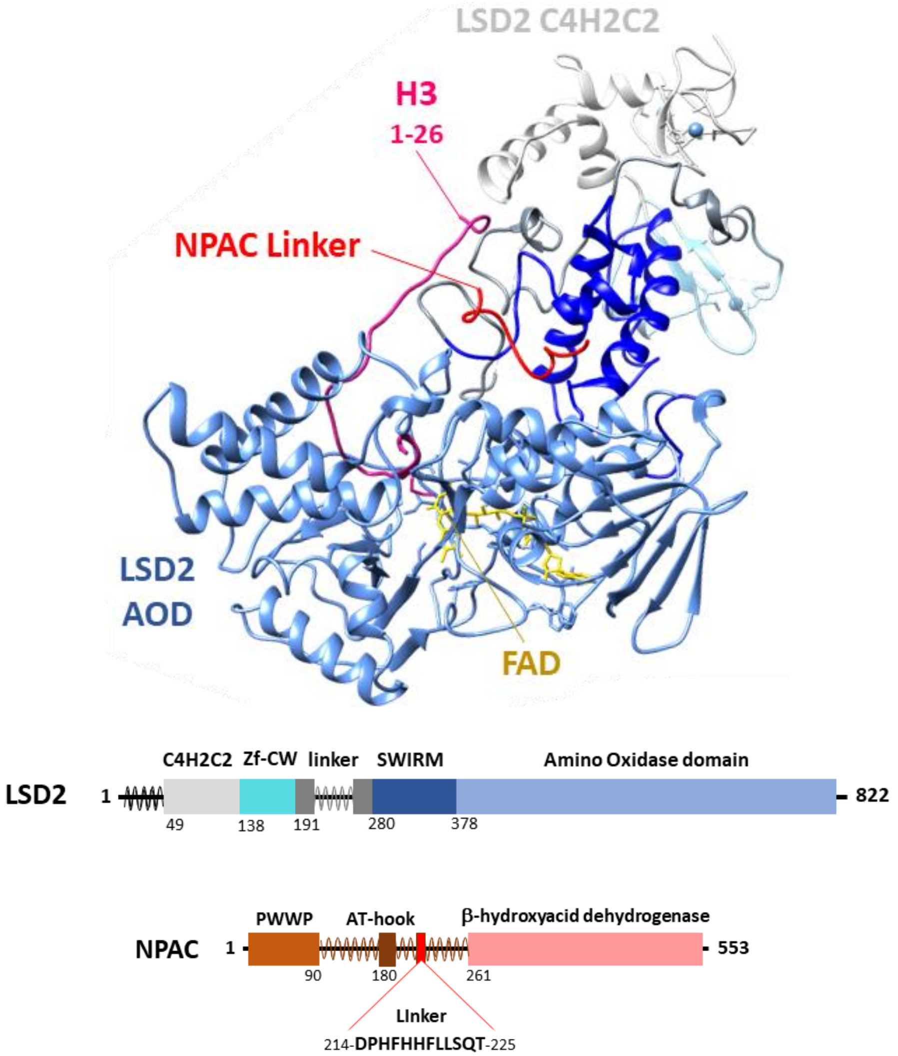


Figure 1. Crystal structure of LSD2 in complex with the NPAC-linker and the substrate H3 (1-26) tail. LSD2 is coloured accordingly with the domain organization legend below. NPAC-linker is a very short segment (residues 214-225) of the multidomain protein. Image adapted from Marabelli *et al.*, 2019.

The N-terminal sequence of NPAC well-aligns with the chromatin reader family of PWWP domains, whose name derives from a conserved Pro-Trp-Trp-Pro motif. Through cooperation between a positively charged surface patch and the nearby aromatic cage, each PWWP recognizes a specific histone methyl-lysine residue in the context of the nucleosome: DNMT3A binds H3K36me3, the HDGF2 (Hepatoma-derived growth factor-related protein 2) binds H3K79me3 whereas H4K20me3 and the BRPF1 (bromo and plant homeodomain (PHD) finger–containing protein 1) bind H3K36me3 (Vezzoli *et al.*, 2010; Qin and Min, 2014; Weaver, Morrison and Musselman, 2018). The PWWP site on NPAC sequence has been shown to localize at H3K36me3-rich chromatin loci (Vermeulen *et al.*, 2010), similarly to LSD2 (Fang *et al.*, 2010). Given that H3K36me3 mostly correspond to transcriptionally active gene bodies, and considering that NPAC-linker complexes LSD2, it becomes even more interesting to observe that the two proteins have been independently demonstrated to stimulate RNA Pol-II during transcription elongation (Fang *et al.*, 2010; Fei *et al.*, 2018).

Materials and methods

Investigation of the LSD2/NPAC system required cloning, expression and purification of multiple proteins, constructs and mutants. In the case of LSD2 production in *E. coli*, a robust protocol had to be implemented, whereas histone purification was optimized thanks to a novel strategy found in literature. Alkylated semi-synthetic nucleosomes proved to be a valid strategy for the analysis of the parameters involved in LSD2/nucleosome complex formation, and few caveats about this method are also here explained. Moreover, collaborations with Panagis Filippakopoulos (SGC, Oxford) and Sriram Subramaniam (Columbia Uinversity, Vancouver) allowed us to tackle the characterization of NPAC-PWWP "chromatin reader" properties, and to solve the cryo-EM structures of the LSD2/NPAC-linker/nucleosome complex.

1. Chemicals and instruments

Chemicals for the preparation of buffers and crystallization solutions were bought from Sigma-Aldrich. Affinity and size exclusion chromatography columns were purchased from GE Healthcare, as well as the ÄKTA purifiers. Concentrators were from Merck. Bio-rad instruments were used to run PCR and for SDS-PAGE and acrylamide gel imaging. Master mix components for PCR were from ThermoFisher Scientific and New England Biolabs. Microscopes for crystals imaging were from Olympus, while the crystallization robot was a Douglas Instruments Oryx-8. Crystallization plates were from Hampton Research. Loops for crystals collection came from Hampton Research and Molecular Dimensions. Screening kits for crystallization conditions were purchased from Quiagen, Jena Bioscience and Molecular Dimensions.

2. Human LSD2 cloning, expression and purification

LSD2 protein construct 31-822, named LSD2Δ30,was cloned and expressed both in *E.coli* and in *P.pastoris* cells. After first test of expression and purification in *E.coli* were successful, as described in detail in paragraphs 2.1-2.4, we moved for further cloning and mutagenesis of LSD2 in *E.coli.* Indeed,, the bacterial expression system was optimal for the production of the many mutant LSD2 proteins in a suitable time. I decided to separate the methods for the production of LSD2 according to the expression system used. Paragraph 2.1 focuses on the protocols used for preparation of the LSD2Δ30 protein from *Pichia pastoris*, whereas paragraphs 2.2, 2.3 and 2.4 introduce respectively: LSD2 cloning and mutagenesis in *E.coli* vectors, their expression and the purification of the obtained proteins from *E.coli.*

2.1 LSD2Δ30 cloning, expression and purification

The gene coding for LSD2 (LSD2Δ30, residues 31-822) was already available in our lab within the pJexpress902 vector modified with a Prescission cleavable site in between the protein C-terminus and the GFP tag-10xHis. The plasmid was maintained in *E. coli* TOP10 (Invitrogen) cells, grown in 15 ml Agar plates (5 gr/L yeast extract, 10 gr/L tryptone, 5 gr/L NaCl, 15 gr/L agar) supplemented with 25 µg/ml Zeocin. *Pichia pastoris* strain KM71-H (Invitrogen) cells were subjected to electroporation for plasmid acquisition. Cells were grown onto MD agar plates (13.4 gr/L YNB, 40 µgr/L biotin, 10 gr/L glycerol, 15 gr/L agar) supplemented with 2% w/v dextrose, 100 mM KH_2PO_4 pH 7.0 and Zeocyn 200 µgr/ml. After two days at 30°c the colonies were collected with a sterile plastic tip and plunged into a 50 ml capacity Falcon containing 15 ml of YPD medium (10 gr/L yeast extract, 20 gr/L bacto-peptone, 20gr/L dextrose). After 3-4 hours of growth at 30°C, 2.8 ml of the pre-inoculum were inoculated into each plastic baffled flask, containing 220 ml of BMGY medium each (100 mM KH_2PO_4 pH 6.0, 13,4 gr/L YNB, 200 µgr/L biotin, 1% v/v glycerol). Cells were grown at 30°c for 60 hours in a shaking incubator at 250 rpm. Cells were then centrifuged at 4000 rpm for 10 minutes and the surnatant buffer was discarded. Cells were then resuspended, in sterile atmosphere, in 110 ml of BMM medium (100 mM KH_2PO_4 pH 6.0, 13,4 gr/L YNB, 200 µgr/L biotin, 0.5 % v/v methanol) and placed back in a shaking incubator at 30°C, 250 rpm. After 24 hours , 0.5% v/V methanol were supplemented. After further 24 hours, cells were collected by centrifugation at 4000 rpm for 10 minutes, flash-frozen in liquid nitrogen and stored at -20°C.

All subsequent steps were performed at 4°C. *P. pastoris* cells were re-suspended in Nip-A buffer (50 mM NaH_2 PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 25 mM imidazole) supplemented with protease inhibitors (1µM leupeptin, 1 µM pepstatin, 1 mM PMSF) and 2 µg/ml DNAse. Cells were lysed in a abead-beater with Zirconia beads (BioSpec products). Then beads were removed with a Miracloth filter paper (Merck) and the cell extract was centrifuged at 70,000 g for 30 minutes. The supernatant was loaded onto a His-Trap column (GE Healthcare) pre-equilibrated in Nip-A. After a wash with 40 mM imidazole in Nip-A buffer, Nip-B buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 250 mM imidazole) was passed through the column, and the sample collected. The elution was supplemented with Prescission protease and dialyzed overnight in Nip-A buffer without imidazole. The day after a second passage through a His-Trap column pre-equilibrated in Nip-A was performed to purify the tag-free protein, eluting in Nip-A 40 mM imidazole. The ability to bind histidine resin is a characteristic of those proteins exposing a hydrophobic patch. Indeed, LSD2 exposes hydrophobic residues on the surface area just ouside the substrate-binding cleft. The tag-free protein sample was then gel filtered through a Superdex 200 10/300 (GE Healthcare) equilibrated in 20 mM Tris pH 8.5 (4 °C), 100 mM NaCl, 1 mM TCEP (Figure 2). The peak eluting at ml was concentrated through an Amicon 30 KDa till a final concentration of 70-100 µM, aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.

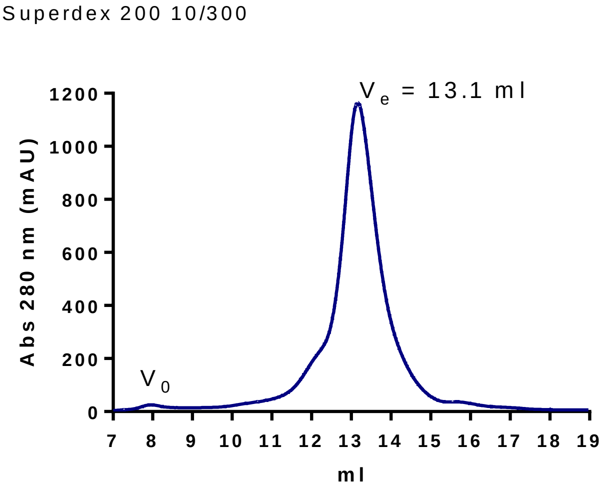


Figure 2. Elution profile of LSD2Δ29 on a preparative Superdex 200 10/300 column (GE healthcare). Running buffer is 20 mM Tris pH 8.5 (4 °C), 100 mM NaCl, 1 mM TCEP.

2.2 LSD2 cloning

The sequence coding for human LSD2Δ30 (residues 31-822) was cloned from the full-length human gene (purchased from GeneArt) into a pJ902Express vector (DNA 2.0) with a C-terminal eGFP-8xHis tag fused to a cleavable Prescission-protease cleavage site. The vector was linearized with PmeI and inserted in the P. pastoris KM71-H strain (Invitrogen) genome through electroporation. In order to perform mutagenesis studies on LSD2, I decided to switch the expression of the protein from *P. pastoris* to *E.coli*. Indeed, expression and purification from bacterial expression systems are quicker and cheaper than from yeast. My first attempts to clone the LSD2Δ30 gene from the pJexpress vector for yeast, to the pET28a(+) bacterial expression vector were not successful. Hence I tried with the full-length construct of human LSD2 (KDM1B), which was already present in house, into a pGEX-6P-1 vector with a N-terminal tag fused to a cleavable Prescission-protease cleavage site. The template for LSD2 site-specific mutagenesis was thus hLSD2fl in pGERX6p1 plasmid (Amp^R), and the primers used were ordered from GeneArt (ThermoScientific) (Table 1).

Table 1. Nucleotide sequences of the DNA primers used for site-directed mutagenesis of LSD2 wild-type gene.

LSD2 mutation	Primer sequence
H103D FW	CGCAGCGATAAAGATG
H103D RV	CATCTTTATCGCTGCG
K104E FW	GCCATGAAGATGGCTAT
K104E RV	ATAGCCATCTTCATGGC
K109E FW	GCTATGACGAATATACCACC
K109E RV	GGTGGTATATTCGTCATAGC
K114E FW	ACCTGGGAAAAAATTTGGACCAGCAAT
K114E RV	AATTTTTTCCCAGGTGGTATATTTGTCATAGCC
K115E FW	TGGAAAGAAATTTGGACCAGCAATGGTAAA
K115E RV	CCAAATTTCTTTCCAGGTGGTATATTTGTCATAGC
N120D FW	ACCAGCGATGGTAAAACCGAACC

N120D RV	TTTACCATCGCTGGTCCAAATTTTTTTCCA
K122E FW	AATGGTGAAACCGAACCGAGCC
K122E RV	TTCGGTTTCACCATTGCTGGTCC
R302D FW	TATAGCGATGATCCGACCATGTATCTGG
R302D RV	CGGATCATCGCTATATTCCGGAAACTCATA
K114E-K115E FW	CTGGGAAGAAATTTGGACCAGCAATGGTAAAACCG
K114E-K115E RV	CAAATTTCTTCCCAGGTGGTATATTTGTCATAGCCATCTTTATG
N120D-K122E-R302D FW	CCAGCGATGGTGAAACCGAACCGAGCCCGAAAGC
N120D-K122E-R302D RV	TTTCACCATCGCTGGTCCAAATTTTTTTCC
H103D-K104E-K109E FW	TATCGCAGCGATGAAGATGGCTATGACgAATATAC
H103D-K104E-K109E RV	TTCATCGCTGCGATAATAGTCATCGAAGCATTC
Δ[241-258] FW	GTACCAGCACCAATCGTGCAAGCGTTCATGTTCCG
Δ[241-258] RV	CATACCCGGAACATGAACGCTTGCACGATTGGTGCTG
K481E-R482D FW	GATGAAGACATGGATTTTCATTTTAACGCC
K481E-R482D RV	CATGTCTTCATCAATGGTCGGATCG

Lyophylized primers were diluted in sterile water with a final concentration of 100µM. Aliquots at 10 µM final concentration of each primer were then prepared in sterile water. Each PCR tube was prepared with 25 µl of MasterMix 2x Phusion Polymerase (ThermoFisher), and 2,5 µl of both forward and reverse 10 µM primers were added, along with 30-50 ng of template DNA. All the reaction tubes were then filled with sterile water till a final volume of 50 ul and the reactiona were performed as follows in a thermocycler (BioRad):

98°C x 30''

98°C x 20''
} x 30
72°C x 1' 45''

72°C x 5'

The final step was set at 4°C for "infinite" time. The PCR tube was then digested with DpnI O/N at 37°C to remove excess template DNA. The next morning, an analytical DNA gel was run to check the right size and to perform gel extraction of the selected band using WizardSV Gel extraction Kit. The DNA was eluted in 30 µl of DNase free water. Then 4 µl of the gel-extracted DNA solution were mixed with 1 µl of InFusion Mix 5x (Takara Clontech). The recombinase reaction was let proceed for 15 minutes at 50 °C, and then cooled on ice.

DH5α new competent cells were used as recipients of the LSD2 gene, ligated on pGEX-6P-1 vector. For each tube, 5 ul were added to a single sterile tube containing 25 ul of competent cells. Tubes were left on ice for 30 minutes, then shocked for 45 seconds at 42°C. After 3 minutes on ice, 250 ul of sterile LB medium were added, at and the cells were incubated in the shaker at 37°C for 1 hour. Then the tubes were centrifuged for 10 minutes at 10'000 g, 150 ul of supernatant were discarded and the remaining cell-containing medium was plated and grown O/N 37°C on an ampicillin-resistance selective medium. To test whether the transforming plasmid was the one predicted, one colony for each plate was inoculated in 5 ml of sterile LB medium and grown O/N at 37°C. The mini-prep Kit (Quiagen) was used to purify plasmid DNA from the bacterial suspension. Cells were lysed under alkaline conditions and the cleared solution was placed on a silica-gel membrane, where only DNA could absorb. After washing and elution of the mini-prep column, the purified plasmid was eluted in 50 µl of sterile EB Buffer, and the samples were sent to for sequencing at the GATC Biotech service.

2.3 LSD2 full-length proteins expression protocol setup

Once confirmed by sequencing the integrity of the expression vector and of the inserted gene, the following *E.coli* strains were transformed: BL21(DE3), BL21(DE3) pLysS, BL21(DE3) RPplus and ROSETTA pLysS, in order to do small-scale expression trials. BL21(DE3) pLysS and BL21(DE3) RPplus cells show very slowed growth rates, since colony appearance in the plate to the pre-inoculum growth. Fout different induction protocols were then tested on the two best-expressing strains: BL21(DE3) and ROSETTA pLysS. Inductionw as performed with IPTG 0.25 mM or 0.50 mM for 5 or 16 hours, for a total of eight samples tested. The cells were collected at 5000 rpm for 10 minutes, then resuspended in 2 ml each of 20 mM Tris pH 8.0, NaCl 500 MM. The cell suspension was then sonicated at 70 % Amplitude, for a total of 8 seconds and hence centrifuged at 70,000g for 10 minutes. Each surnatant was loaded onto 150 µl of Glutathione resin beads (pre-equilibrated in buffer) for a small-scale purification. After 1 hour of loading at 4 °C, the resin was washed in the same buffer used to resuspend cells, and finally elution was carried out with 200 µl of 50 mM GSH. SDS-PAGE showed ROSETTA pLysS cells gave the best yields in LSD2 expression, in particular both O/N inductions (0.25 and 0.5 mM) showed a clearly over-expressed band around 116 KDa, that is the expected molecular weight of the GST-LSD2 protein (LSD2 full length 91 KDa + 25 KDa for the GST tag). After evaluating the grams of cell growths, the protein expression by SDS-PAGE, and the quality of the purified protein after the first small-scale purification trials, I set up the expression protocol for LSD2 full-length wild-type and mutant proteins as follows. A single BL21(DE3) colony was grown in about 100 ml LB media, O/N in a shaking incubator at 37°C, 200 rpm. Then 20 ml of the pre-inioculum were added to each flask containing 1 L of sterile LB, supplemented with Ampicillin and 3-5 drops of Anti-foam. Flasks were incubated at 37 °C, 200 rpm, till O.D.$_{600nm}$ =0.8. The temperature of the incubator was set to 17 °C, and each flask was supplemented with IPTG 0.25 mM and ZnSO4 100 µM. induction was performed for 16-18 hours. Then cells were collected by centrifugation at 10’000 g for 10 minutes. The usual yield was about 4.9 grams of cells for each liter of growth.

2.4 LSD2 full length purification protocol set up

To optimize the purification procedure, I tried different buffers and chromatographic steps. All purification trials were performed at 4°C. About 12gr of cells were resuspended in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol supplemented with PMSF, Pepstatin, Leupeptin and DNase, sonicated at 30%Amplitude, 5 sec ON + 9.5 sec OFF for a total ON of 1 minute, and then centrifuged at 56,000 g for 45 minutes. The extract surnatant was loaded onto a HiTrap 5 ml GST column (GE Healthcare) at 0.2-0.3 ml/min. As first trial, overnight cleavage by GST-Prescission Protease was tried in the GST-column. The protein was mostly found in the GSH elution, indicating that in-column cleavage is not efficient (Figure 3A). I then tested cleavage in dialysis against resuspension buffer with either the GST-tagged or the His-tagged Prescission Protease. After an overnight incubation, both proteases were able to completely remove the GST tag from LSD2 (Figure 3B). When I tried to remove the His-tagged prescission protease, I unexpectedly found that LSD2 also binds to the nickel resin, probably because of hydrophobic patches exposed on its surface (Figure 3B). Better results were obtained the GST-tagged protease as after the cleavage step the use of a GST Hi-Trap resin allowed to remove the free GST tag along with the protease itself (Figure 3B). Free GST is present in the imidazole elution along with the protein, meaning they interact. In order to get rid of the GST tag remaining bound to LSD2 after the nickel column step, I tried an ionic exchange step. Briefly, the uncleaved sample was dialyzed O/N with His-Prescission protease in a 6-8 kDa cut-Off dialysis membrane. The cleaved sample was then collected and loaded onto a 15Q anionic exchange column in 50 mM

NaH2PO4 pH 8.0, 100 mM NaCl, 5% (v/v) glycerol buffer. The elution was carried in gradient of 50 mM NaH2PO4 pH 8.0, 1 M NaCl, 5% (v/v) glycerol buffer. LSD2 all eluted in the FT, which was concentrated and checked by Superdex 200 10/300 in running buffer 20 mM Tris-HCl pH 8.5 (4°C), 200 mM NaCl,1 mM DTT. Many degradation bands and/or contaminants, along with free GST, were present still (Figure 3).

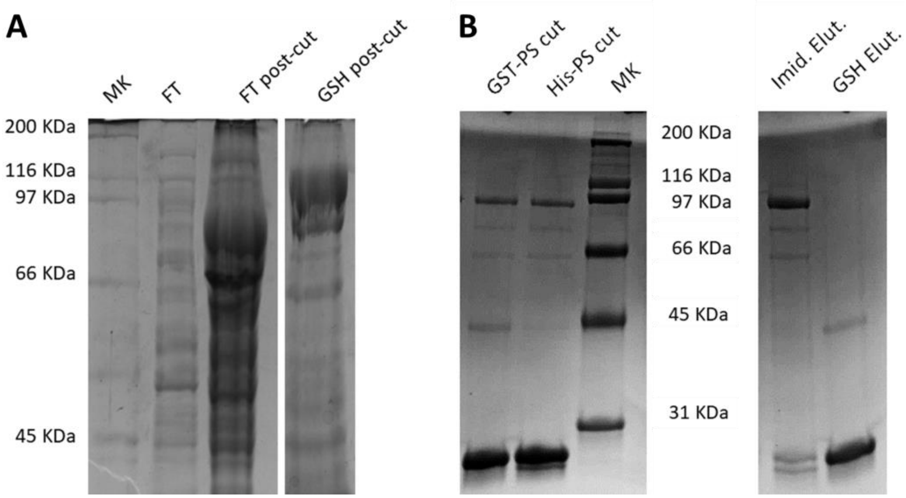


Figure 3. A) GST-prescission protease does not cleave efficiently the GST-LSD2 protein in column. After the O/N incubation in column, only part of the protein of interest has been cleaved (expected MW= 91 KDa), whereas the uncleaved GST-LSD2 remained bound to the column (expected MW=116 KDa). From left to right, each lane refers to: the molecular weight ruler (MK), the flow-through , the Flow-through after O/N incubation with the protease in column (FT post-cut), and the eluted sample with 50 mM GSH after the O/n incubation with the protease in column (GSH post-cut). B) Both His-tagged and GST-tagged prescission proteases efficiently cleave the GST-LSD2 protein in dialysis (GST-PS cut and his-PS cut refer to the samples cleaved O/N with the respective protease). Two identical samples were dialysed O/n against 1 L each of the same resuspension buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol). The cleaved protein binds to the Nichel resin and elutes with 250 mM imidazole (Imid Elut. lane on the right gel). Part of the free GST tag (26 KDa) contaminates the LSD2-containing sample and can be removed through passage in a GST-trap column (GSH Elut. Refers to the GST-tag bound to the GST column after the LSD2-containing sample has been flown through).

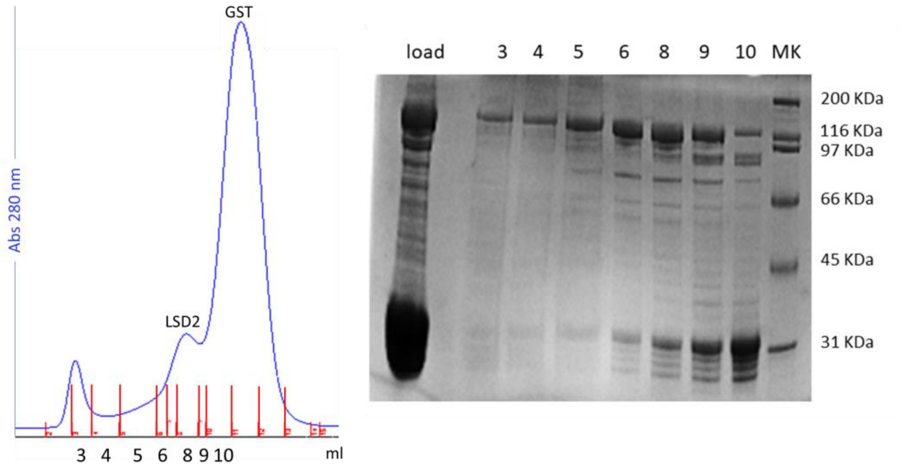


Figure 4. After anionic exchange free GST still remains as contaminant in the LSD2 sample. On the left the elution chromatogram from a Superdex 200 10/300 is shown. The sample loaded is the concentrated flow-through of a 15Q anionic exchange column. SDS-PAGE analysis of the SEC fractions confirms that LSD2 is still contaminated by various species, among which free GST tag is the predominant one.

I tried to get rid of the many contaminants being purified along with LSD2fl, by switching the expression from ROSETTA pLysS to the BL21(DE3) *E.coli* cells, which also gave good results in the small-scale expression trials. I followed a protocol very similar to that of the ROSETTA expression. Given that BL21(DE3) cells were growing slower, I simply increased the volume of the O/N pre-inoculum (up to 1:30 with respect to the growth media volume) growth to reach the desired O.D.$_{600nm}$ for induction within similar intervals. The new expression system proved to be effective in increasing the stability and purity of LSD2. After sonication, centrifugation and GST Hi-Trap column loading according to the standard protocol, I eluted with 50 mM GSH buffer. The protein was purer than in the previous trials. The only contaminant remaining the free GST-tag. Again, neither a nickel Hi-Trap nor an ionic exchange step were sufficient to separate the pure LSD2 from the free GST tag. Finally, passing the cleaved sample through a GST and a nickel Hi-Trap 5 ml columns in tandem, significantly reduced the amount of contaminating GST.

The final purification protocol is summarized as follows. LSD2 expressing E. coli BL21(DE3) cell pellets were re-suspended in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 25 mM imidazole, 2% (w/v) lysozyme, and 1 mM PMSF. Cell suspension was sonicated and centrifuged at 70,000 g for 30 minutes. The supernatant was loaded onto a GST-Trap column (GE Healthcare) and the tagged-protein eluted with 50 mM GSH. His- Prescission protease was added (1 mg/15 ml of sample). The sample was dialyzed overnight in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl and 5% glycerol. After dialysis, the sample was loaded onto a second GST-Trap to remove the GST-tag. A second His-Trap column was mounted in succession onto the ÄKTA system. to remove the His-tagged protease. A final size-exclusion chromatography step was performed on a Superdex 200 10/300 equilibrated in 20 mM Tris pH 8.0 (25 °C), 200 mM NaCl, 1 mM DTT.

GST contamination remained persistent through all purifications, but the amount was deemed to be insignificant. Also, the purification of all tested LSD2 mutants was consistent among them. Here below, the final SEC step of one of the mutants is shown as an example (Figure).

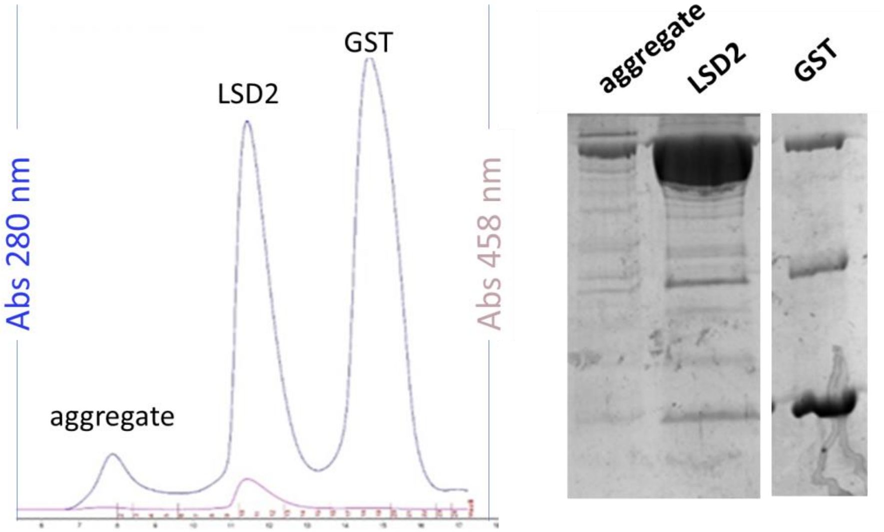


Figure 5. Final purification step of the K104E LSD2 mutant. Left: Elution profile of the purification containing three distinct peaks: aggregate, LSD2 K104E and free GST tag. On the right, SDS-PAGE of the peak fractions.

3. Recombinant nucleosomes preparation

3.1 Recombinant histones cloning and expression.

The four histone proteins were expressed and purified separately. The pET3 vector (Ampicillin resistance) containing the sequences of the histone proteins from *Xenopus laevis* are a kind gift from Doctor Toshiya Senda (Biomedicinal Information research Center, National Institute of Industrial Science and Technology, Tokyo). Histone H3 carrying the two mutations K4C-C110A also was already present in house, on a pET3 vector. The same vector was the template to obtain the K4C-K23M-K27M-C110A H3 protein through site-directed mutagenesis. This H3 mutant is expressed and purified in order to produce H3 histone specifically alkylated at position 4, where the substrate lysine would lie. The presence of a reactive group at this position is key to the production of covalently linked nucleosome/demethylase complexes, as explained in paragraph 6.1. The primers used were ordered from GeneArt (ThermoScientific) (Table 2) and used as previously described for a two-step mutagenesis PCR (see *LSD2 cloning*, paragraph 2.2).

Table 2. Nucleotide sequences of the DNA primers used for site-directed mutagenesis of *X. laevis* H3 wild-type gene.

K23M-K27M FW	**CCATGGCAGCCAGGATGTCC**
K23M-K27M RV	GGAGCGGACATCCTGGCTGC

3.2 Recombinant histones expression and purification

Recombinant *X. laevis* histones were expressed and purified according to published protocols (Dyer *et al.*, 2004) For each histone plasmid, a transformation was performed on *E. coli* BL21 DE3 pLysS strain (Stratagene). From each plate of transformation, one colony was taken and a pre-inoculum was grown O/N in a shaking incubator (200 rpm) at 37°C. Cells from the pre-inoculum were then poured, in a 1:20 dilution ratio, in the new auto-inducing medium 2xTY (16 g/L bacto-tryptone, 10 g/L yeast extract, 5 g/L NaCl) supplemented with 0.1% glucose. Usually after 2-3 hours the cells reached an Optical Density O.D.$_{600nm}$ of 0.4. Induction was performed with IPTG 0.4 mM at 37°C. After 2 hours cells were harvested by centrifugation at 10’000 rpm (round per minute) in a Beckman TA-10-250 fixed angle centrifuge rotor for 10 minutes at 15°C and 6 grams aliquots were prepared in 50 ml Falcons. Cells were then gently resuspended by pipetting in Wash buffer (50 mM Tris pH7.5 at 4C°, 100mM NaCl, 1mM β-ME, 1 mM PMSF), flash-frozen in liquid nitrogen and stored at -20°C.

The frozen pellet was thawed in a water bath at 37°C, and a homogenization step was required also to fluidize the solution. Cell extract was centrifuged at 25’000g for 20 minutes at 4°C. Supernatant was discarded , and the pellet was resuspended by gentle pipetting in an equal volume of Wash buffer (50 mM Tris pH 7.5 at 4C°, 100 mM NaCl, 1mM β-ME, 1mM PMSF), supplemented with 1% Tryton X-100 to release histones from insoluble inclusion bodies. Then the sample was centrifuged at 15’000 rpm for 10 minutes at 4°C. The pellet was then washed once again in Wash buffer supplemented with 1% triton X-100 and centrifuged as previously. Other two washes were performed in Wash buffer to remove the detergent from the pellet. Each pellet was incubated for 30 minutes at room temperature with 200 ul of DMSO each. Unfolding buffer (6 M GnHCl, 20 mM AcONa pH 5.2, 5 mM DTT) was added till a total volume of 15 ml, and the tubes were gently mixed with a spatula, and left for 1 hour at rt. Samples were then centrifuged at 25’000g for 20 minutes at 20°C, and the supernatant was collected and filtered. The sample was then loaded onto a Sephacryl S200 XK50 Gel Filtration column (GE Healthcare) previously equilibrated in Urea A Buffer (7 M Urea, 20 mM AcONa pH 5.2, 200 mM NaCl, 2 mM BME, 1 mM EDTA) (Figure 6).

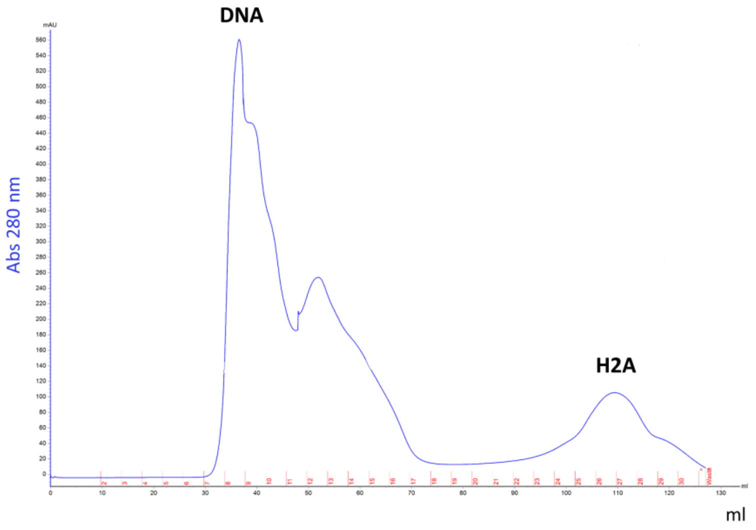


Figure 6. Sephacryl 200 XK 50 column elution of histone H2A, shown here as a representative of all histone preparations. The last peak is the one containing histone protein, whereas most of the DNA elutes in the void volume.

Among the fractions collected, those showing a ratio A260/A280 smaller than 1, were analysed by SDS-PAGE electrophoresis gel. These fractions should be less contaminated by DNA, since its absorption maximum is at 260 nanometers, while the maximum absorption wavelength for proteins is usually 280. After gel electrophoresis, those fractions containing histones were collected and concentrated in Amicon Ultra 30 kDa. Then the sample was loaded onto a SP FF16/10 cationic exchange column (GE Healthcare). We used this approach to clear the sample from DNA bound to basic residues of histones. DNA was washed away in the same Urea A buffer, since it cannot bind to the negatively charged resin, which displaces it from histones. The gradient elution was performed in 10 Column Volumes (CV), till 100% of Urea B Buffer (7 M Urea, 20 mM AcONa pH 5.2, 1 M NaCl, 2 mM BME, 1 mM EDTA) (Figure 7). Fractions were analysed by SDS-PAGE electrophoresis and agarose gel to check the presence of histones and DNA, respectively.

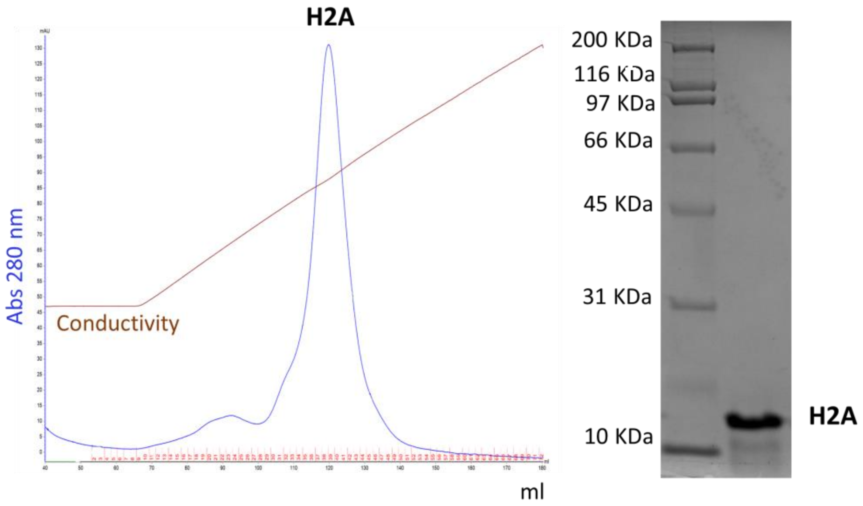


Figure 7. Elution profile of H2A histone from the cationic exchange SP FF 16/60 column (left panel). SDS-PAGE (right panel) stained with Coomassie blue showing the final histone preparation after SP FF passage.

Protein fractions were pooled and put in a dialysis membrane (CutOff 5-8 KDa). The dialysis was performed in 2 liters of 2 mM BME water solution at 4°C, in order to remove urea. The sample was then concentrated till 2 mg/ml, lyophilized and stored at -20°C.

3.3 Optimization of the histone purification protocol.

My research project on the mechanism of nucleosome recognition by LSD2 implied large production of semi-synthetic nucleosomes. Given the enormous amount of time invested in each single nucleosome preparation, I tried to optimize the protocol in order to optimize it and reduce the time required to perform it. I started from a paper published by Klinker and colleagues on PLOS one (Klinker *et al.*, 2014), and adapted it to the instruments available in the laboratory. Briefly, dry cell pellets, expressed following the standard protocol (Dyer *et al.*, 2004) were resuspended in SAU buffer (40 mM NaOAc pH 5.2, 1 mM EDTA, 5 mM β-ME, 10 mM Lys), supplemented with protease inhibitors, PMSF 1 mM and urea powder up to 6 M. The cell suspension was sonicated on ice for a total of two minutes (30% power) till urea powder was completely dissolved in buffer. The extract was cleared by centrifugation for 40 minutes at 56000 g. The supernatant was passed over a Capto-Q column and a SP FF 16/10 column in series, to bind DNA and histones respectively. After washing in 6 M urea SAU buffer, the Capto-Q was removed and elution was carried out from the SP FF column with NaCl gradient up to 1 M NaCl. Histones were eluting at 200-400 mM NaCl (Figure 8). Fractions were checked by SDS-PAGE and agarose gel for protein and DNA contaminations, pooled and dialyzed O/N against 2 liters of 20 mM Tris pH 8.0. The morning after, the sample was passed again over a Capto-Q column to remove remaining contaminant DNA. The unbound fraction was concentrated up to 10 mg/ml and checked by SDS-PAGE and UV-Vis absorbance spectra. Aliquots of 4 mg each were prepared, flash-frozen in liquid nitrogen and stored at -80 °C.

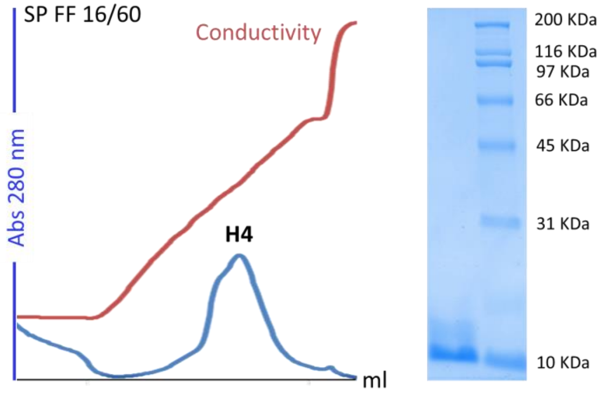


Figure 8. SP FF 16/10 elution profile of the H4 preparation following the new protocol from Klinker and colleagues (Klinker *et al.*, 2014). On the right, the SDS-PAGE analysis of the collected fractions is shown.

The main advantage of this optimized protocol, besides saving one day of work (Figure 9), is that it enabled me to produce 3-fold more histone protein from a single preparation.

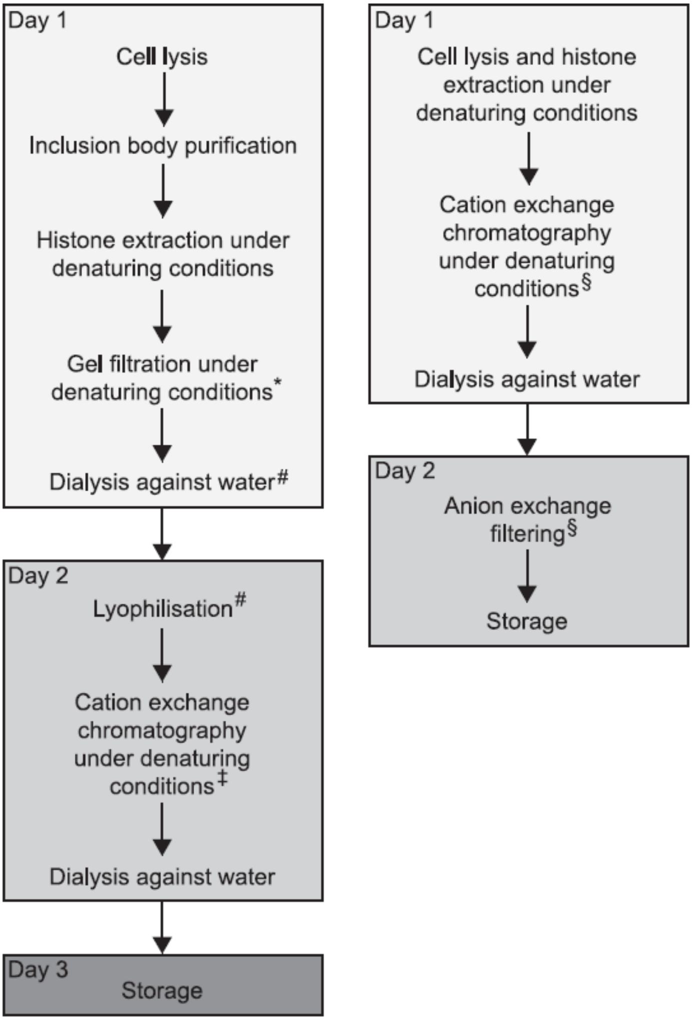


Figure 9. Comparison of the two histone purification strategies. On the left: the conventional histone purification method according to the Luger's protocol (Luger, Rechsteiner and Richmond, 1999). On the right: the new protocol I adopted according to Klinker and coworkers (Klinker *et al.*, 2014). Image from Klinker *et al.* (Klinker *et al.*, 2014).

3.4 H3 alkylation

The histone H3 protein used for the production of semi-synthetic nucleosomes carries two mutations: Lys4Cys-Cys110Ala. Mutation of the substrate lysine 4 to cysteine is key to provide the anchoring point to the propargylamine-carrying reactive, as explained in figure 9 through an acid-base reaction. Such a H3 histone exposing a highly reactive chemical moiety at the substrate residue 4, where the flavin-dependent enzymes carry out their demethylation reactions, can covalently react with the FAD cofactor of either LSD1 or LSD2, ending up in a covalent enzyme-substrate complex (for more details, see paragraph 6.1). The mutation of the cysteine residue at position 110 to alanine on the other hand, is necessary to avoid attachment of the propargyl to the wrong, non-substrate position. The alkylation reaction follows a similar protocol to that published by Simon and colleagues for the production of nucleosomes with methyl-groups at specific locations (Simon *et al.*, 2007).

The freeze-dried histone was dissolved in water bath at 40°C in alkylation buffer (1 M Hepes/NaOH pH7.8, 4 M guanidinium chloride, 10 mM L-Met, 10 mM DTT). Alkylation reaction was perfomed in the same buffer using a final 50 mM concentration of the N-(2-chloroethyl)-N-methylprop-2-yn-1-amine alkylating agent (Simon *et al*. 2007). Histone concentration was 10 mg/ml (655 µM), and DTT was added till 20mM concentration. The sample was let at room temperature for 2 h, then more DTT was added till a final molarity of 30 mM. after 30 minutes, more alkylating reagent was added till 75 mM. after 2 hours and 30 minutes, the reaction was stopped by adding 50 mM BME. After 30 minutes, the buffer was changed in desalting to MilliQ water. Then the sample was concentrated.

The product of reaction is a histone carrying a propargylamine analogue of dimethyl-lysine covalently bound with Cys4 (Figure 9). The final outcome was checked by ITMS-ESI mass spectrometry. The reaction involves the thiol group of the cysteine, which attacks the carbon carrying the chlorine substituent which is a good leaving group. The extremely reactive propargyl moiety does not interfere with this acid-base reaction.

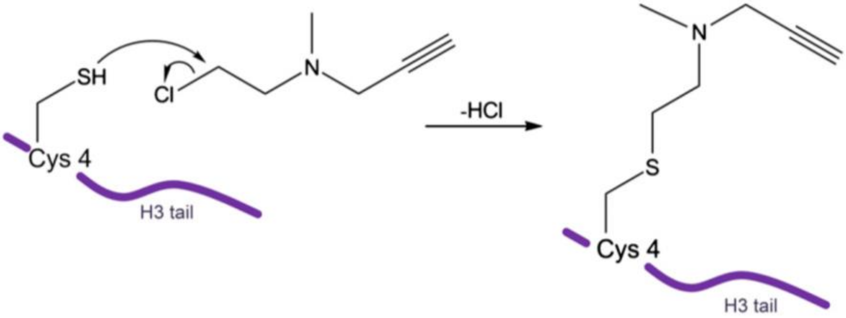


Figure 9. Mechanism of reaction between N-(2-chloroethyl)-N-methylprop-2-yn-1-amine and cysteine residue in position 4 on H3 histone tail. From Pilotto, Speranzini *et al*. (Pilotto *et al*., 2015).

3.5 Mass Spectrometry of Alkylated H3

ITMS-ESI technique was used to check the yield of the H3 histone alkylation reaction (Figure 11). Measurements were performed at the laboratory of Federica Corana, Centro Grandi Strumenti, Cascina Cravino, Pavia. ESI (ElectroSpray Ionization) is a technique to produce ions from macromolecules in solution. The liquid is ionized by a high-voltage field, and an aerosol is formed. Each drop carries a charged fragment of the macromolecule, so that it can be subjected to the voltage field applied and reach the detector. Ion Trap (IT) MS uses a combination of electric and magnetic fields to capture charged particles in vacuum. Since all charged fragments are retained in the “field” trap, through manipulation of magnetic field forces, subclasses of all possible ions present (with a particular charge/mass ratio) can be released and directed to the detector in a controlled manner, so that different ions can be distinguished according to their charge/mass ratios.

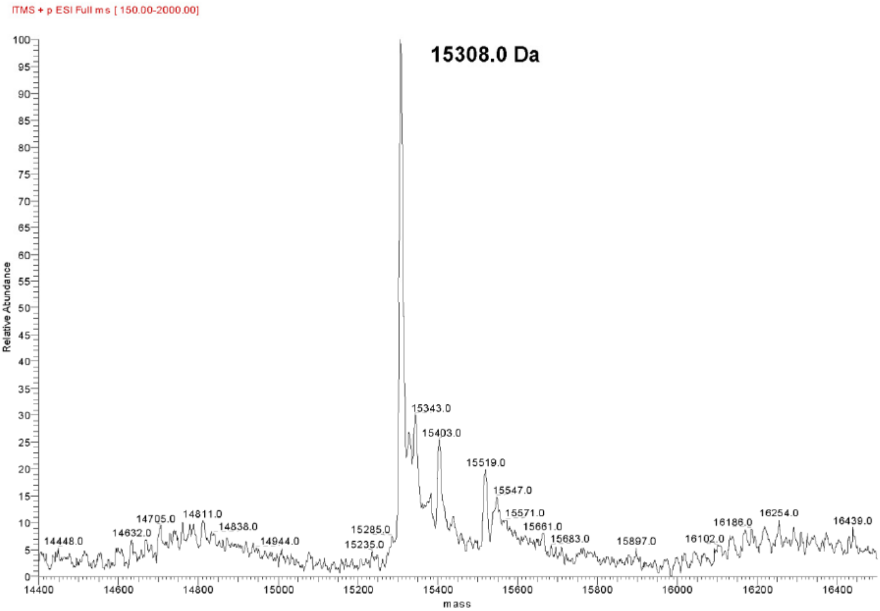


Figure 11. Mass spectrum of a typical alkylated H3 sample. The graph reports the relative abundance of the species at different calculated molecular masses. Calculations are performed starting from the number of positive ions hitting the detector, each with its charge-to-mass ratios. Ions were considered as formed by fragments of the protein plus one hydrogen or sodium atom. The graph clearly shows that Alkyl-K4C-C110A-H3 (MW 15308 Da) is the predominant species (MW of the non-alkylated histone K4C-C110A-H3 is 15216 Da).

3.6 Histone octamer refolding

Lyophilized histones are separately re-suspended in Unfolding Buffer (6M GnHCl, 20 mM AcONa pH 5.2, 5 mM DTT) and let at room temperature for 1 hour. The protein concentration was measured at NanoDrop Spectrometer, looking at 276 nm Absorbance. Histones were then mixed at equimolar ratio, and the sample was diluted in unfolding buffer till a final concentration of 1 mg/ml of total proteins and extensively dylaized at 4°C against 600 ml of Refolding Buffer (2M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, 5 mM BME), changing the dialysis solution three times, one every 6 hours. The slow replacement of the unfolding buffer with an high salt concentration buffer aims at correctly refold histones. The sample was then concentrated in an Amicon Ultra 10 kDa tube and loaded onto a Superdex 200 16/60 gel filtration column (GE Healthcare). The fractions collected were checked on a SDS-PAGE, and the ones containing the octamer were pooled and concentrated (Figure 12). After addition of glycerol till 50%, the octamers were stored at -20°C.

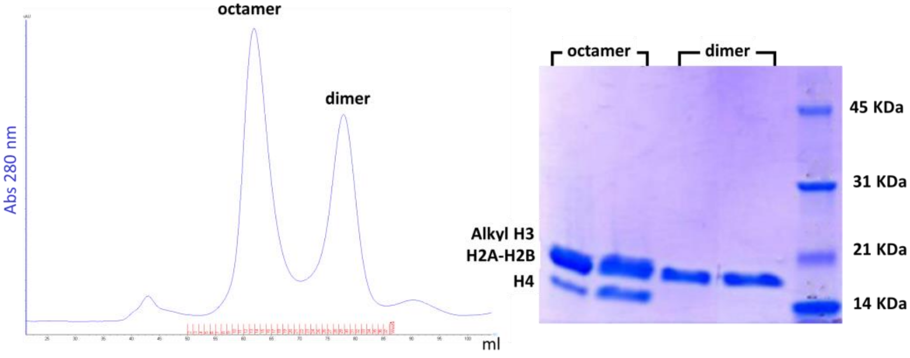


Figure 12. Elution profile of refolded octamer from Superdex 200 16/60 with its corresponding SDS-PAGE analysis. The column separates the refolded octamer (elution volume 62 ml, SDS-PAGE lanes 1, 2) from excess of H2A-H2B dimers (elution volume 77 ml, SDS-PAGE lanes 3-4).

3.7 147-bp Widom DNA sequence preparation.

The DNA sequence for nucleosome reconstitution is carried on a pUC vector (Ampicillin resistance) as an array of 15 head-tail multiple copies, separated one from the other by an EcoRV recognition site (16 restriction sites /plasmid). A single colony of *E. coli* DH5α cells was inoculated in 1 ml of LB at 37 °C for 2 hours. Then 4 ml more were added and the growth continued for 2 hours. 10 ml of LB were added and the cells were let grow for 4 h more. Two milliliters of the pre-inoculum were transferred into 1 liter of LB and the culture was left at 37°C O/N in a shaking incubator. Cells were harvested by centrifugation and stored at -20 °C. DNA was extracted with a MegaPrepKit (Qiagen). Restriction cleavage of the plasmid was performed for 24 hours at 37°C, at a DNA concentration of 1 mg/ml, calculating 30U Eco RV HF (NEB) for each nmol of restriction sites. To check digestion, few microliters were run on a 1.5% (w/v) agarose gel.

To separate the vector from the digested insert a precipitation step followed. PEG 6000 up to 6.5% (v/v) and NaCl up to 500 mM were added directly to the digestion reaction and the mix was let on ice for 1h. the sample was then transferred in centrifuge tubes, and centrifugation was done at 20’000g for 20 minutes at 4°C. The supernatants were collected and the 146bp DNA sequence was precipitated by addition of 2.5 volumes of ethanol at 4°C. As the precipitant was added, the solution was placed at -80 °C and let there for half an hour. A final centrifugation step, (25’000g, 20’, at 4°C), pelleted the DNA at the bottom of the tubes (Figure 1). The supernatants were discarded whereas the pellets were dried and dissolved in sterile water.

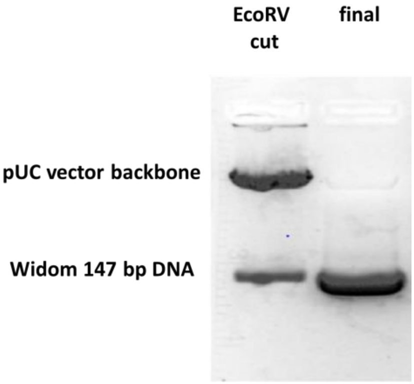


Figure 13. Agarose gel analysis of the DNA purified after EcoRV digestion. The first lane on the left is the precipitate after PEG 6K and NaCl addition. The right lane corresponds to the last pellet obtained.

3.8 Semi-synthetic nucleosome reconstitution.

Reconstitution of the Nucleosome Core Particle (NCP) occurs during gradual removal of salt from the octamer-DNA mixture. As salt ions are stripped off the solvation shell of octamers and DNA, the two species can interact and assemble as in nuclear chromatin. The speed of the reaction is controlled through progressive desalting of the solution. The whole procedure is carried out at 4°C to preserve proteins. First, the DNA-containing solution is adjusted to 2 M KCl concentration. Then, DNA, octamers, and H2 dimers are mixed at molar ratio of 1.2 : 1 : 0.5 respectively, reaching a final DNA concentration of 0.7 mg/ml using RB high Buffer (2 M KCl, 10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT). The sample is placed in a dyalisys membrane, within a high-salt buffer solution. Then, the high-salt buffer is gradually exchanged with a low-salt buffer (250 mM KCl, 10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT) by a peristaltic pump, working at 4°C, with a flow speed of 0.7 ml/min, for 36 hours. The sample was finally dialyzed in RB low-salt buffer for 3 hours more. To improve correct positioning of DNA on the octamer core, a thermal shift is required: 50°C in water bath for 45 minutes. The sample was placed on ice for cooling, and then loaded onto a CaptoDEAE anion exchange column (GE Healthcare). The basic free dimers and octamers, the reconstituted NCP particles and the negatively charged free DNA can be separated according to their electrostatic interactions with the weakly cationic resin. The column was pre-equilibrated in TES-250 buffer (10 mM Tris pH 7.5, 0.5 mM EDTA, 250 mM KCl). After loading and a first washing step in the same buffer, a second wash was peformed with 30% TES 600 buffer (10 mM Tris pH 7.5, 0.5 mM EDTA, 600 mm KCl), during which free dimers and octamers eluted. The elution gradient was performed from 30 to 100% of TES 600 buffer, in 3 column volumes. Nucleosomes were eluted during this step, while the DNA remained bound to the resin (Figure 15), and a final wash with 1 M KCl replenished the column. NCPs were immediately dialyzed overnight against TCS buffer (20 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT), in order to remove salt ions interfering with DNA wrapping around the octamer. SDS-PAGE and native PAGE were run to check purity and homogeneity of the final sample. Purified NCPs were concentrated in Amicon Ultra 30kDa and stored on ice, for 2-3 weeks at last.

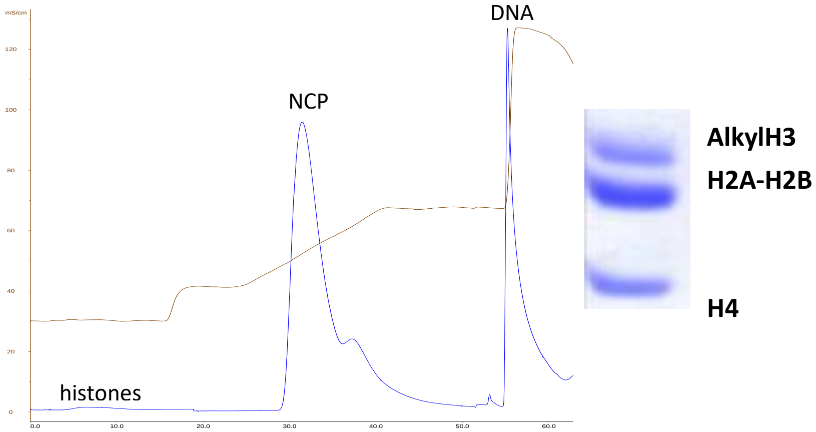


Figure 15. Elution profile of recombinant reconstituted nucleosomes from CaptoDEAE anionic exchange and SDS-PAGe gel of the final sample. After nucleosome elution from the CaptoDEAE column, a washing step with 1 M KCl is necessary to detach DNA from the resin(second peak). The 280 nm-absorbance profile is shown in blue, in brown the conducibility of the buffer. On the right, the SDS-PAGE experiment clearly reveals the presence of all four histones in the sample.

3.9 Chicken erythrocyte nucleosome preparation

Another method to obtain nucleosomes was purification from chicken blood. Chicken blood samples were collected at “Avicola Valtidone” (via Fontanino 6/A, Località Campo D’Oro, Castel San Giovanni) and kept on ice until filtered through 4 layers of cheesecloth, then centrifuged for 4 min at 1500g at 4°C. The clear supernatant (the serum) is discarded, and the erythrocytes are washed three times by centrifuging at 4°C, speed 1500g in 25 volumes of Buffer A (0.34 M sucrose, 15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 2 mM EDTA, 2 mM EGTA, 15 mM BME, 0.1 mM PMSF). Blood aliquots are then frozen at -80°C. All subsequent steps are performed at 4 °C. Pelleted erythrocytes are lysed with a mixer in 25 volumes of Buffer N (0.34 mM sucrose, 15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 2 mM EDTA, 2mM EGTA, 15 Mm BME, 0.1 mM PMSF, 0.5% Nonidet P-40). After centrifugation for 5 minutes at 1500g, two wash steps were executed with 50 volumes of Buffer N. After erythrocyte lysis, the sample was prepared for chromatin digestion: it was washed three times by centrifugation for 4 min at 1400g in 50 volumes of Buffer B (0.34 mM sucrose, 15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 15 mM BME, 0.1 mM PMSF). The pellet was resuspended in Buffer B to 20-30 ml final volume. The suspension was then warmed to 37°C and $CaCl_2$ was added till 1 mM concentration. *Staphylococcus aureus* nuclease was added to 45 U/mL, and the digestion was run for 5 minutes. Addition of EDTA 1.5 mM stopped the digestion, and the sample was cooled to 4 °C, and centrifuges again 5 minutes at 8000g. The pelleted nuclei are resuspended in 5 volumes of 0.25 mM EDTA, pH 7.5, and left in gentle stirring for 1 hour. After nuclei lysis, a centrifugation step (20 minutes at 8000g) separated the nuclear membranes from soluble chromatin components. The pellet is discarded and the supernatant kept. The sample is diluted in EDTA 0.25 mM, pH 7.5 till A_{260} around 50. Then 1% Tris pH 8.0, NaCl 75 mM were added along with Roche Complete Inhibitors 1X. To strip H1/H5 histones, pre-equilibrated CM Sepharose FF (GE Healthcare) was added up to 100mg/ml final concentration, and the sample was gently stirred O/N. The day after, centrifugation 25 minutes at 8000g further purified the supernatant, which is collected and spinned in Eppendorf for 10 minutes at 10’000rpm. A Second chromatin digestion was performed: in Amicon concentrator, CutOff 100KDa, spinning 10 minutes at 3000g, we changed the buffer to Buffer C (20 mM NaCl, 10 mM Tris pH 7.5, 2 mM BME). After dilution with the same buffer to reach A_{260} around 25, the suspension was warmed to 37 °C. Then $CaCl_2$ was added up to 1 mM final and *S. aureus* nuclease was added up to 50 U/ml final. The digestion was let to proceed for 10 minutes before being stopped with the addition of EDTA 1.5 mM final, and cooled again on ice. Through an Amicon concentrator (membrane Cut-Off 100kDa) the sample was concentrated till 500 μl, and loaded at 0.9 ml/min onto a Sephacryl S-400 (300 ml bed volume) in Buffer C supplemented with EDTA 0.2 mM. Fractions were checked for the presence of histones and DNA though SDS and agarose gel electrophoresis, and those with polynucleosomes were pooled and digested a third time. After dilution in Buffer C till a final A_{260} of 10, the suspension was warmed to 37 °C, and CaCl 0.8 mM, and *S. aureus* nuclease (50 U/ml) were added. 10 minutes later, the reaction was stopped by adding EDTA 1.7 mM and lowering of the temperature to 4°C again. Further concentration prepared the sample to be injected onto Sephacryl S-400 Gel Filtration column. Again, Buffer C with EDTA 0.2 mM was used for elution, with a flow rate of 0.7 ml/min. Fractions collected were checked through SDS-PAGE and agarose gels to select those with DNA lengths less than about 160 bp. Finally the nucleosomes were concentrated in Amicon Ultra Cut-Off 100kDa in TCS buffer (20 mM Tris pH 8.0, 0.25 mM EDTA, 3% glycerol). Nucleosomes were kept on ice or flash-frozen in liquid nitrogen and stored at -80°C.

4. NPAC protein analysis, cloning, expression and purification

4.1 NPAC bioinformatics analysis

NPAC sequence was the subject of various analysis with online bioinformatics tools in order to predict the function of the protein and the boundaries between protein domains (Jacobson *et al.*, 2014). Homologous sequences to NPAC were searched on UniProt (Bateman *et al.*, 2017) and PDB (Berman *et al.*, 2000) databases. PrDOS (Takashi Ishida1, 2007) output evidenced a long stretch of disordered amino acids starting from residue 100 to 260 (Figure 16).

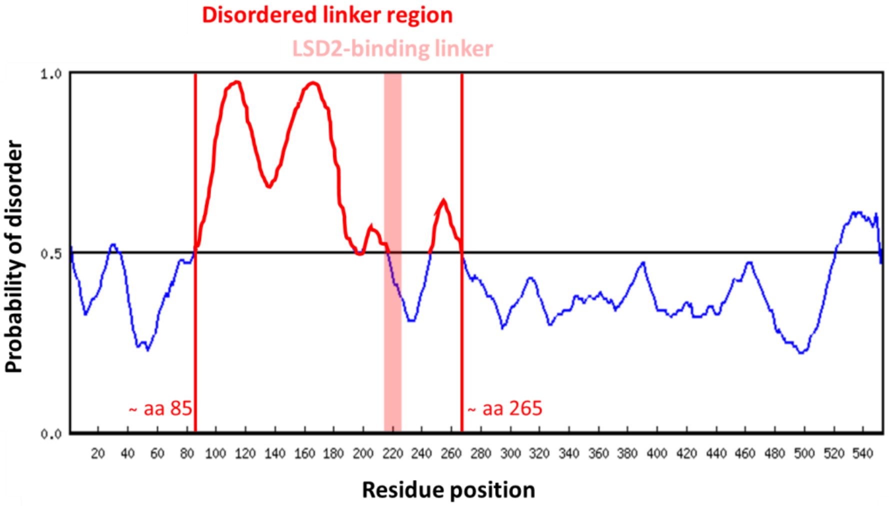


Figure 16. Disorder profile plot of NPAC, calculated by PrDOS software and arranged to evidence the clearly disordered region covering roughly 180 residues (evidenced by red lines) and the LSD2-binding linker within (pink area). Two ordered regions constitute the N-terminal and C-terminal portions.

Separate homology search of the N-terminal (105-aa long) and the C-terminal (residues 250-553) portions of NPAC were performed with BLAST (Altschul *et al.*, 1990), Clustal Omega (Li *et al.*, 2015) and ELM (Dinkel *et al.*, 2016)..Representative NPAC sequences were selected from Uniprot (Bateman *et al.*, 2017) and compared with Clustal Omega (Li *et al.*, 2015) as described in the paper attached (Marabelli *et al.*, 2019).

In order to identify the NPAC domain organization and design the most convenient NPAC construct suitable for expression as recombinant protein, NPAC sequence was analysed with PhophoSitePlus website (www.phosphosite.org) to look for the presence of any physiologically relevant residue.

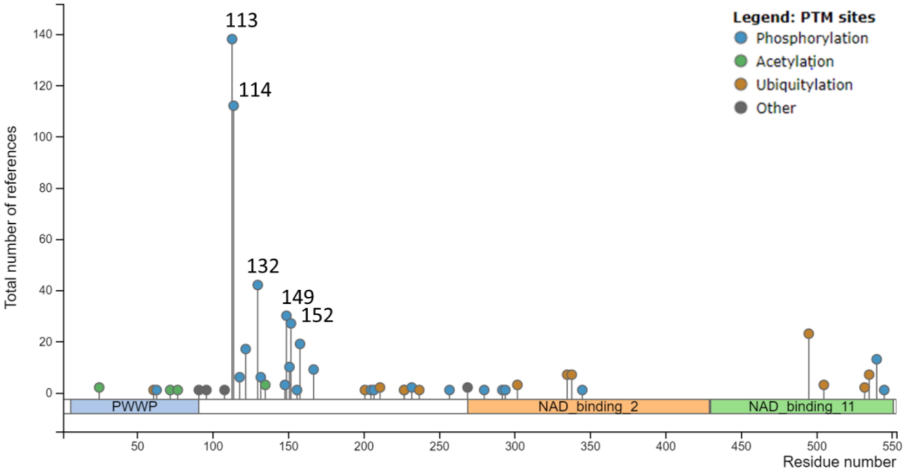


Figure 17. Phosphosite plus output for human NPAC. On the "y" axis is the total number of references for which each amino acid position of the NPAC protein (on the "x" axis) has been found to be phosphorylated. On top of the most prominent phosphorylation sites is reported the residue numbering for clarity. Most of the phosphorylated residues lie at the boundary between the PWWP domain and the flexible linker region. A legend for the PTMs here represented is on the top right of the image.

Given the high number of phosphorylated residues between residues 105 and 122 (Figure 17) and considering that sequence alignment among several orthologs revealed the existence of a 98-residue isoform (X3 in *Erinaceus* species; Figure 56), various NPAC constructs including different portions of the PWWP domain were tested to delineate the boundary between the N-terminal domain and the linker region (Table 3). For clarity, in Figure 18 only the PWWP 1-105 protein is shown (second line from the top).

Regarding the C-terminus of NPAC, two different constructs were designed to include the dehydrogenase domain (residues 261-553), with and without the LSD2-binding linker (residues 214-225; the two proteins are schematized at the bottom of Figure 18). A third protein was designed after the first experiments with NPAC full-length failed (the full-length NPAC is the first line from the top), which includes the N-terminal PWWP module (residues 1-98) along with the C-terminal portion containing both the LSD2-binding linker and the dehydrogenase domain (third protein from the top; Figure 18).

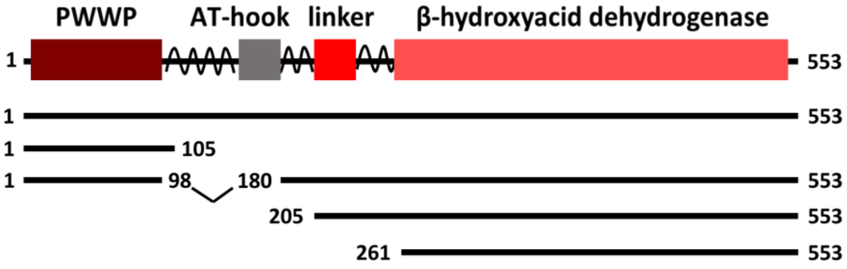


Figure 18. NPAC domain organization as evidenced from bioinformatics analysis. The constructs designed for expression and analysis are outlined below. N-terminal and C-terminal residues are indicated.

4.2 *NPAC-DH plasmid and mutagenesis*

The TEV-cleavable N-terminal 6xHis-tagged NPAC dehydrogenase gene (NPAC DH, residues 261-553, Figure 18) cloned in a pNIC28-Bsa4 vector, was a kind gift from Structural Genomics Consortium (SGC), Oxford (UK). NPAC DH mutants were obtained through site-specific mutagenesis (Table 3).

Table 3. Nucleotide sequences of the DNA primers used for site-directed mutagenesis of human NPAC dehydrogenase gene.

M437K FW	**AAGATGAAGCTGATCGTGAACATGGTCCAAGGGAGCTTC**
M437K RV	GATCAGCTTCATCTTGGCTGCATTGCCCACTTC
M437N FW	CAATGCAGCCAAGATGAATCTGAT
M437N RV	ACCATGTTCACGATCAGATTCATCT

4.3 NPAC-DH proteins expression and purification

E. coli BL21 codon-plus (DE3)-RP (Novagen; chloramphenicol resistant) cells were transformed and plated. A single colony was then picked and grown at 37° C, in a shaking incubator (200 rpm) O/N in LB medium supplemented with chloramphenicol and kanamycin at 37° C. The day after, the pre-inoculum was poured in the ratio of 1:100 v/v into a fresh TB (Terrific-broth) medium with Kanamycin and Chloramphenicol. The culture was grown till the optical density (OD_{600}) reached 0.8, and then induced at 17°C with IPTG 0.5 mM for 18 hours. Cells were collected by centrifugation at 5000 rpm for 10 minutes.

Cell pellet was resuspended in Ni-A buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 40 mM imidazole, 5% (v/v) glycerol), supplemented with protease inhibitors and 1 mM PMSF. Resuspended cells are sonicated with a 250 W Branson sonicator set at 30% Amplitude, till complete lysis and next centrifuged at 70,000 g for 35 minutes at 10° C. The following protein purification is performed at 20°C because of the tendency of the protein to aggregate at 4°C. The soluble fraction was filtered with a 2 µm filter to remove cell particulate and loaded onto a 5 ml HiTrap column (GE Healthcare), pre-equilibrated with Ni-A buffer. After extensive washing, the 6xHis-tagged protein is eluted in Ni-B buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 250 mM imidazole, 5% (v/v) glycerol). The eluate was incubated with TEV protease and dialysed against

2L Ni-A buffer without imidazole at 20° C O/N. The second purification step is a further affinity chromatography with a HiTrap Ni-column equilibrated in the dialysis buffer. The fractions containing NPAC DH are pooled and concentrated with an Amicon Ultra concentrator 30 kDa cut-off (Millipore), until a final volume of 5 ml. The sample was then loaded onto a Superdex 75 16/60 (Ge Healthcare) equilibrated in dialysis buffer. The fractions were anaylilsed by SDS-PAGE (Figure 19), collected and concentrated till 10 - 12 mg/ml. The concentration of the protein was determined through detection of the absorbance at 280 nm with a NanoDrop spectrophotometer (Thermo Scientific). The extinction coefficient of NPAC DH at 280 nm was calculated with the ProtParam tool (Wilkins *et al*., 2005) on the online SIB ExPASy Bioformatics Resources Portal (Artimo *et al*., 2012) and it is 16,305 M^{-1} cm^{-1} for the wild-type protein.

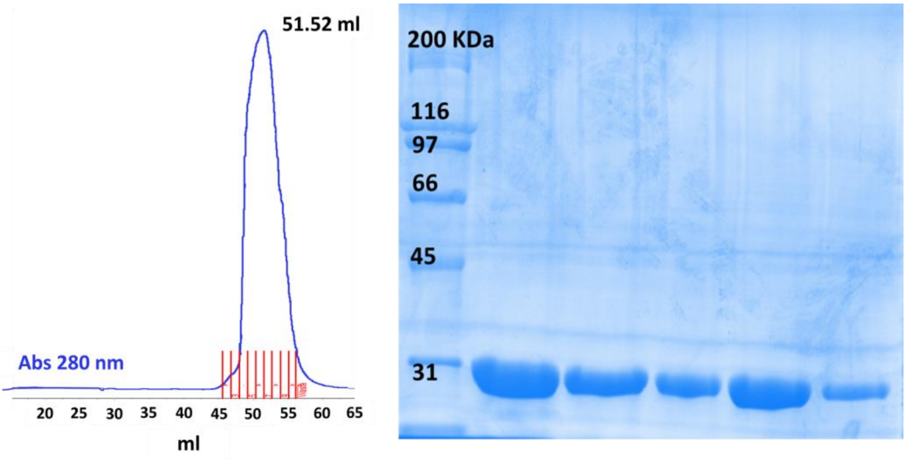


Figure 19. Elution profile of the purified NPAC-DH from a preparative Superdex 75 16/60 and relative SDS-PAGE analysis. The elution volume is 51.52 ml, which corresponds to the molecular weight of the NPAC-DH tetramer (about 120 kDa). On the right, SDS-PAGE of the eluted fractions confirms the identity of the purified construct (expected MW of the monomer: 34.053 kDa).

4.4 NPAC full-length cloning

The NPAC full length gene was ordered from GeneArt (ThermoScientific), with designed 3' and 5' 15-bp sequences complementary to the ends of the recipient vector pET24a digested with BamHI and XhoI. Recombinase reaction between the synthetic gene fragment and the open vector backbone was carried out with the Infusion cloning kit (Takara, Clontech). The DNA sample was then used to transform Stellar Competent cells (Takara, Clontech), which were subsequently used for DNA extraction and sequencing, according to the protocol previously described. The final vector contained NPAC full-length with an N-terminal Flag-8xHis-SUMO tag. This construct was used as the template for subsequent cloning experiments (Table 4).

Two-step PCR reactions were performed, as previously described (see *LSD2 cloning* paragraph), digested with DpnI and purified using the commercial kit NucleoSpin Gel and PCR Clean-up (Macherey-Nagel). Transformed TOP10 *E. coli* competent cells were plated onto LB agar plates. DNAs from single colonies were extracted using the Miniprep kit (Qiagen) and sequenced by GATC Biotech service.

Table 4. Nucleotide sequences of the DNA primers used for site-directed mutagenesis of human NPAC wild-type full length gene.

1- 98 FW	**GACCAGATAATAACGTCATCCCACAATTCTTCTGATGACAAG**
1-98 RV	TATTACTGGTCTTTCCCTTTGGCTCTCCTGAGGAACTCTTC
1-105 FW	CTTCTTAATAAGATGACAAGAATCGACGTAATTCC
1-105 RV	CATCTTATTAAGAAGAATTGTGGGATGACGTCTG
1-122 FW	AGGCCATAATAAAACTCAGGTGATGAGAAGCGCAA
1-122 RV	GTTTTATTATGGCCTACTTCTCTCCTCACTGGAATTACGTC
Δ99-179 FW	GGGAAAGACCAGGATCTCACCATCCCGGAGTCTAGTACCG
Δ99-179 RV	ATCCTGGTCTTTCCCTTTGGCTCTCCTGAGGAACTCTTCGAC
Δ205 FW	ATCCGATCCTCATTTCCATCATTTCC
Δ205 RV	ATCGGATCCACCACCAATCTG

4.5 NPAC full length, Δ99-179 and Δ205 proteins expression and purification

The same protocol was applied for all three proteins NPAC full length, Δ99-179 and Δ205. Representative preparations are showed in Figures 20 and 21 for the full-length and Δ205 NPAC proteins respectively. All proteins were expressed with a N-terminal 8xHis tag.

E. coli BL21 codon-plus (DE3)-RP (Novagen; chloramphenicol resistant) cells were transformed and plated. A single colony was then picked and grown in LB medium supplemented with chloramphenicol and kanamycin, at 37° C, in a shaking incubator (200 rpm) O/N. The day after, the pre-inoculum was poured in the ratio of 1:100 v/v into a fresh TB (Terrific-broth) medium with Kanamycin and Chloramphenicol. The culture was grown till the optical density (OD600) reached 0.8, and then induced at 17°C with IPTG 0.5 mM for 18 hours. Cells were collected by centrifugation at 5000 rpm for 10 minutes.

Cell pellet was resuspended in Ni-A buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 40 mM imidazole, 5% (v/v) glycerol), supplemented with protease inhibitors and 1 mM PMSF. Resuspended cells are sonicated with a 250 W Branson sonicator set at 30% Amplitude, till complete lysis and next centrifuged at 70,000 g for 35 minutes at 10° C. The following protein purification is performed at 20°C because of the tendency of the protein to aggregate at 4°C. The soluble fraction was filtered with a 2 μm filter to remove cell particulate and loaded onto a 5 ml HiTrap column (GE Healthcare),

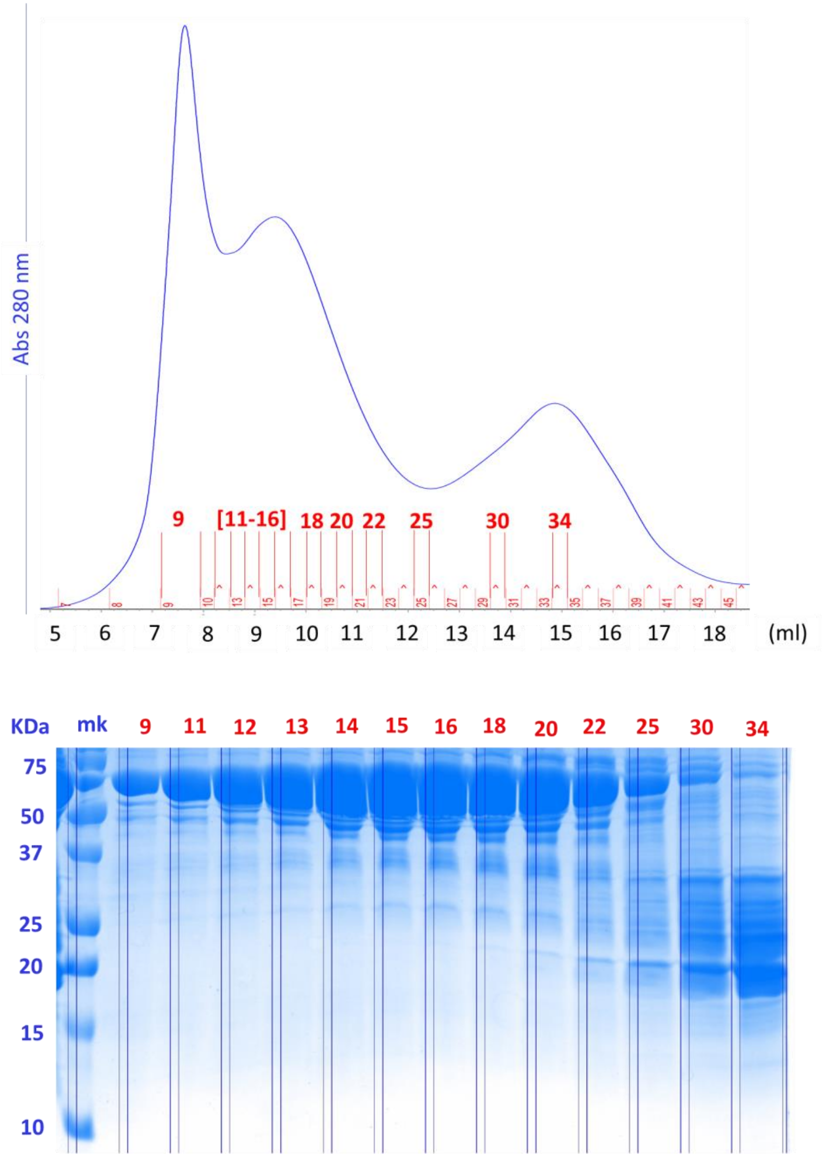


Figure 20. **Elution profile of the purified NPAC full length from a preparative Superdex 200 10/300 (top) and relative SDS-PAGE analysis (bottom). The elution volume of fractions [11-18] corresponds to the molecular weight of the full-length NPAC tetramer (about 300 kDa). SDS-PAGE of the same fractions confirms the identity of the purified construct (expected MW of the full-length monomer: 72 kDa).**

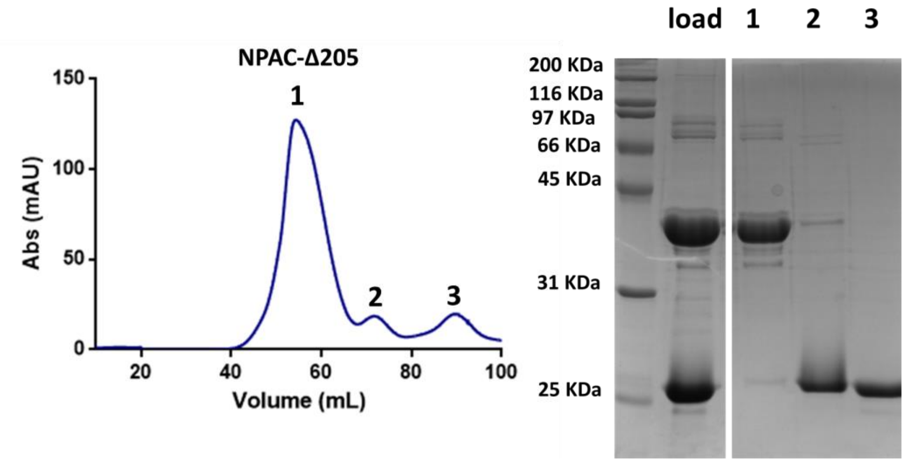


Figure 21. Preparative Superdex 200 16/60 elution profile and relative SDS-PAGE analysis of the cleaved SUMO-NPACΔ205 protein. The protein of interest (peak1) elutes separately from the SUMO tag (found in peaks 2 and 3). The protein behaves as a tetramer, with an estimated molecular weight of 140 kDa. SDS-PAGE confirms the expected mass of the monomeric polypeptide (37.54 kDa).

4.6 NPAC-PWWP constructs expression and purification

Three representative PWWP constructs were separately cloned and expressed in BL21(DE3) *E. coli* cells and analysed for the effect of different C-terminal peptide lengths. There are evidences indeed, about the strong effect on the stability of small chromatin reader proteins exerted by the length of their N-terminal or C-terminal peptides (Savitsky *et al*., 2016). Moreover, Phosphosite Plus output for the amino acid sequence of NPAC (Figure 17), showed a putatively important phosphorylation sites at residues 113, 114 and 122, very close to the C-terminus of the identified PWWP module sequence (residues 1-98). Small-scale expression tests clearly showed that the 1-98 and the 1-105 proteins are the best expressing ones (Figure 22).

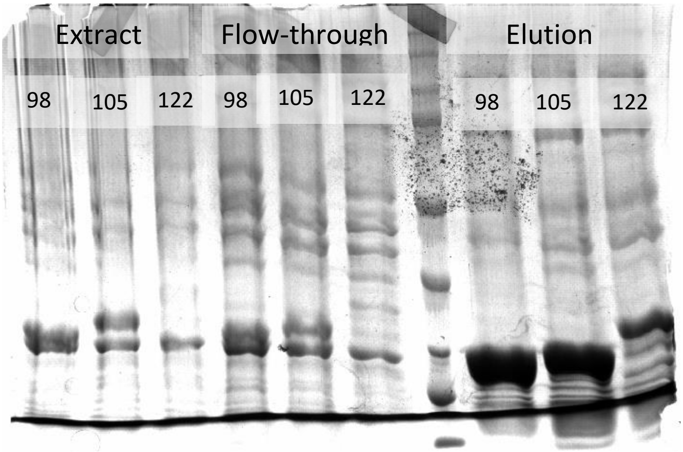
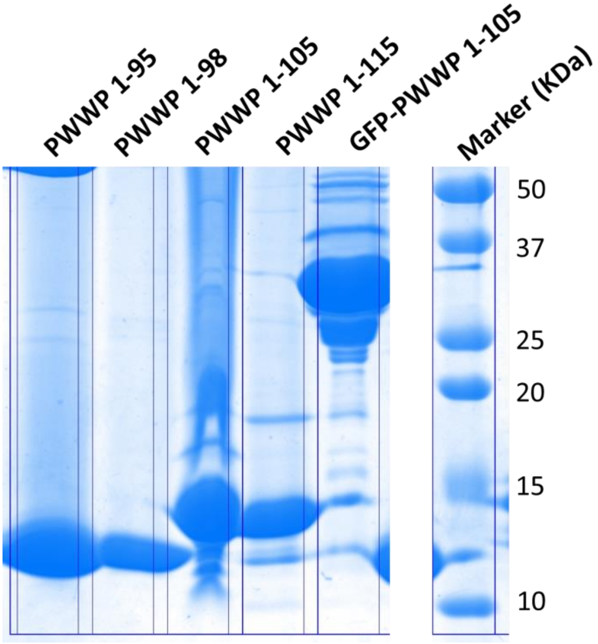


Figure 22. SDS-PAGE of the small scale purification trials of three NPAC-PWWP constructs 98, 105 and 122. The 1-122 construct degrades during the purification, whereas the 98 and 105 constructs are more stable.

NPAC-PWWP 1-95, 1-98, 1-105 and 1-115 were cloned and prepared for crystallization trials and biochemical experiments. An additional GFP-PWWP 1-105 protein was cloned, expressed and purified according to the same protocol with the aim to perform binding studies with the nucleosomes. Detection of the GFP absorbance signal at 500 nm indeed would facilitate identification of the nucleosome species both in analytical chromatographic experiments and in native PAGE assays. The same protocol for expression and purification was applied for the preparation of the GFP-PWWP 1-105 protein. (Figure 23). Proteins were expressed in BL21(DE3). Cells were grown in 2xYT medium supplemented with Kanamycin and 1 % (w/v) glucose at 37 °C until OD_{600} reached 1. Induction was performed with IPTG 0.1 mM, for 16 h at 17 °C. Cells were collected by centrifugation and stored at -20 °C.

Collected cells were resuspended in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5 % glycerol, 1 mM PMSF and sonicated. The suspension was centrifuged at 56,000 g for 1 h and the cell extract was loaded onto a His-Trap column. The resin was washed first with 50 mM NaH_2PO_4 pH 8.0, 1 M KCl, 5 % glycerol and then with 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5 % glycerol, 50 mM imidazole. Elution was carried out in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5 % glycerol, 250 mM imidazole. His-tagged Prescission protease was added and the sample was dialyzed O/N in resuspension buffer. The sample was then passed a second time through a His-Trap column, and the flow-through was collected, concentrated and loaded onto a Superdex 75 10/300 in 20 mM PIPES pH 6.5, 200 mM NaCl (Figure 23).



F

Figure 23. SDS.-PAGE analysis of the purified PWWP proteins used for crystallization trials.

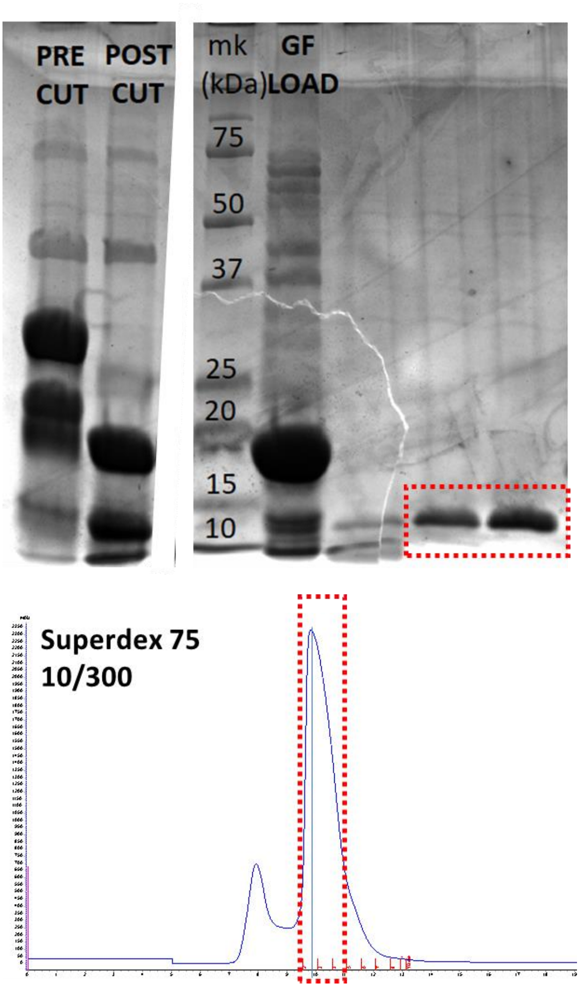
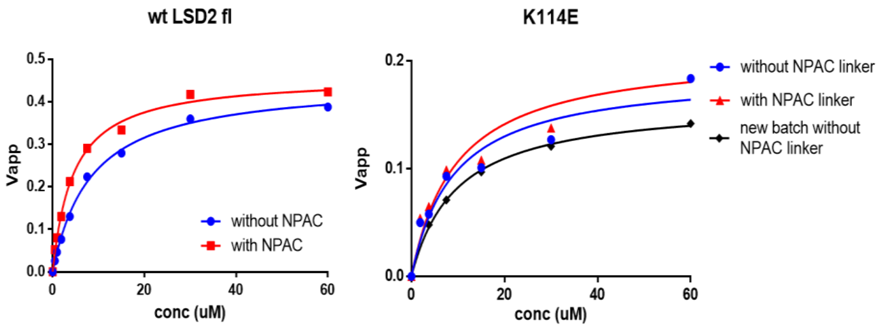


Figure 24. SDS-PAGE showing the last steps of NPAC-PWWP (1-105) preparation. From left to right, each lane refers to the sample eluted from the first His-trap column, the same sample after O/N incubation with 6xHis-Prescission protease, the molecular weight standards (in kDa), the sample loaded onto preparative gel filtration and the tested aggregate peak (eluting in the void volume) and the first and the third fractions of the monodispersed NPAC-PWWP peak. The elution profile of the preparative Superdex 75 10/300 is shown below. Collected fractions are highlighted in a red box.

5. Activity assays

5.1 LSD2 demethylase activity tests

All purified LSD2 proteins were tested for their demethylase activity prior to analytical SEC experiments aimed at studying the interaction with semi-synthetic nucleosomes. These experiments were performed using a 21-aa monomethylated H3K4 peptide (mimicking the first 21 residues of histone H3 tail) as substrate and the HRP/AR assay to measure the conversion rate as previously done with LSD1 (Binda *et al.*, 2010). Briefly, the FAD-mediated demethylation reaction produces H_2O_2 as side product, which is used by Horse Radish Peroxidase (HRP) to convert Amplex Red into resorufin. This ancillary reaction (stoichiometrically equivalent to the demethylation reaction) can be followed by measuring either absorbance or fluorescence of resorufin. The LSD2 reaction was performed in 20 mM HEPES pH 8.5 by measuring fluorescence using a Clariostar plate reader (Clontech). Although the purified wild-type and mutant full-length proteins were equally active, the K_M of the proteins expressed in *E. coli* appeared to vary depending on the preparation with respect to those expressed in *P. pastoris*, which was then selected as the standard sample for activity assays (in Figure 25, a representative mutant is chosen for all LSD2 preparations from *E.coli* expression system).



	Wt	Wt + linker	K114E	K114E + linker	K114E*
K_M (min^{-1})	8.40 ± 0.56	4.51 ± 0.36	9.08 ± 0.02	9.65 ± 0.03	9.67 ± 0.01
k_{cat} (µM)	0.45 ± 0.01	0.46 ± 0.01	0.19 ± 0.02	0.21 ± 0.03	0.16 ± 0.01

Figure 25. Apparent kcat values of the full length wild-type LSD2 expressed in *E.coli* (top) and a representative full-length mutant LSD2 (K114E, bottom), also expressed in *E.coli*. In both graphs the activity of LSD2 has been measured in presence and in absence of the small 12-aa long NPAC linker peptide, which had been reported to influence LSD2 activity (Fang *et al*., 2013). No significant difference was detected either between LSD2 with and without NPAC linker, neither among LSD2 wild-type and mutant proteins. In the K114E graph, samples from two different preparations are compared.

5.2 NPAC-DH dehydrogenase activity tests

To investigate the catalytic properties of NPAC-DH, spectrophotometric assays were performed with a Cary 100 UV-Vis spectrophotometer (Agilent). In a quartz cuvette, NPAC-DH 10 µM was incubated in Tris 20mM pH 7.5 buffer, and a molar excess of either NADH, NAD^+, NADPH or $NADP^+$ was added. The absorbance at 340 nm was monitored to detect any increase/decrease of cofactor concentration. Various dehydrogenase substrates were separately added at 1 mM final concentration: acyl-CoA, alanine, ascorbic acid, ascorbate, citrate, formate, glutamate, glyoxylate, malate, malonate, pyruvate, succinate, lactate, glycerol, glyceraldehyde, glyceraldehyde 3-phosphate, gluconate, glucose 1,6-phosphate, sorbitol, fructose, fructose 6-phosphate, L-arabinose, ATP, ADP, GTP, UDP, nicotinate. The reaction mixture was kept at 25 °C for 30 minutes while constantly monitored, but no activity was detected (data not shown).

6. Complex formation studies with semi-synthetic nucleosomes

6.1 LSD2 /NPAC-linker/nucleosome

In order to stabilize the otherwise short-lived LSD2/nucleosome complex, H3K4-alkylated nucleosomes were reconstituted (see H3 alkylation, paragraph 3.4).

The flavin cofactor of LSD2 is attached by the highly reactive propargyl group exposed by H3 K4C (Figure 10) and a covalent LSD2/nucleosome complex forms (Figure 26).

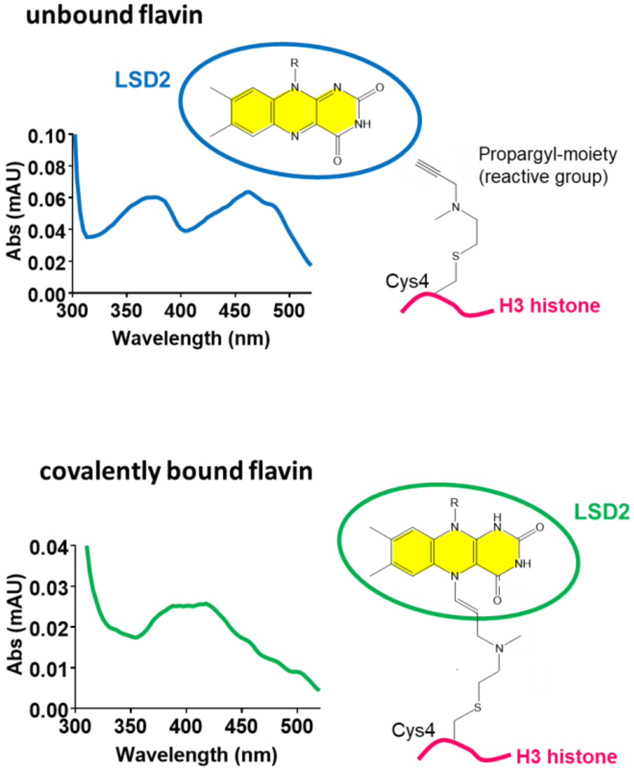


Figure 26. Alkylated semi-synthetic nucleosomes enable chemical trapping of the LSD2/nucleosome transient complex. Covalent binding of the propargyl unit to the FAD cofactor produces a shift in its absorbance profile, which can be monitored though spectrophotometric methods. Image adapted from Marabelli *et al.*, 2019.

The flavin absorbance spectrum changes upon covalent binding, and the maximum absorption shifts from 458 nm to 400 nm (Figure 26). If not differently stated, NPAC-linker is always mixed together with LSD2 and semi-synthetic nucleosomes, at a molar ratio five-fold higher with respect to LSD2. The reaction yield can be monitored by analytical-SEC experiments (Figure 27).

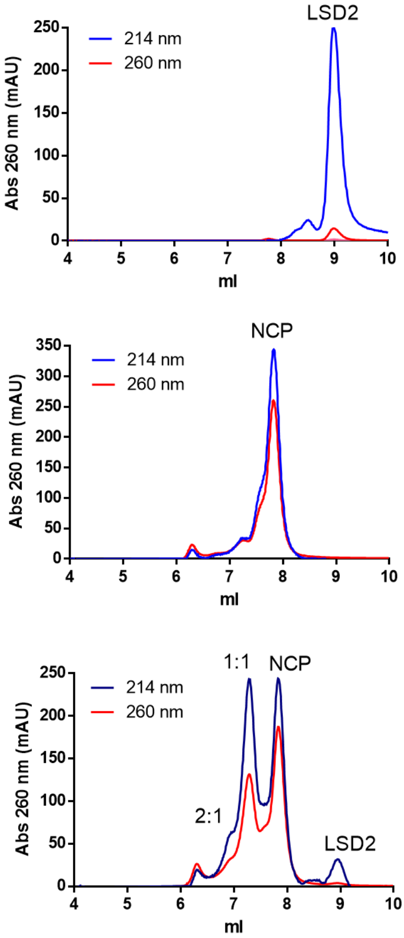


Figure 27. From top to bottom: elution chromatograms of LSD2 alone, semi-synthetic nucleosome and sample containing LSD2-NPAC and semi-synthetic nucleosomes. In the last chromatogram, four species are present. The identities of the last two peaks is gathered from the first two chromatograms, namely LSD2 (V_e: 8.92 ml) and alkyl-NCP (V_e: 7.73 ml). The first two eluting species are LSD2- NCP complexes at molar ratios of 2:1 and 1:1, respectively.

As LSD2 is exposed to unsaturating concentrations of the nucleosome substrate, this tool allows also evaluation of formation of the catalytically competent complex, that mirrors the K_M of LSD2 activity on nucleosome but not its product release rate (k_{cat}). It was then used to analyze the effect of various parameters on the recognition of the nucleosome by LSD2.

There is an intrinsic limit to the yield of the formation of the LSD2/nucleosome complex even in presence of the facilitator NPAC-linker. The free nucleosome peak is always present and a 100% yield on LSD2/NPAC-linker/nucleosome complex species formation cannot be achieved also after prolonged incubation times (Figure 28), nor with great excess of LSD2/NPAC-linker. The maximum achievable LSD2/nucleosome complex yield depends on the sample of semi-synthetic nucleosomes being used. Indeed, different nucleosome preparations tested with the same preparation of LSD2 yielded different chromatograms. For this reason, each analytical SEC experiment presented in this thesis is always accompanied by and compared to the chromatogram of the wild-type LSD2 tested onto the same nucleosome preparation in the standard conditions described above.

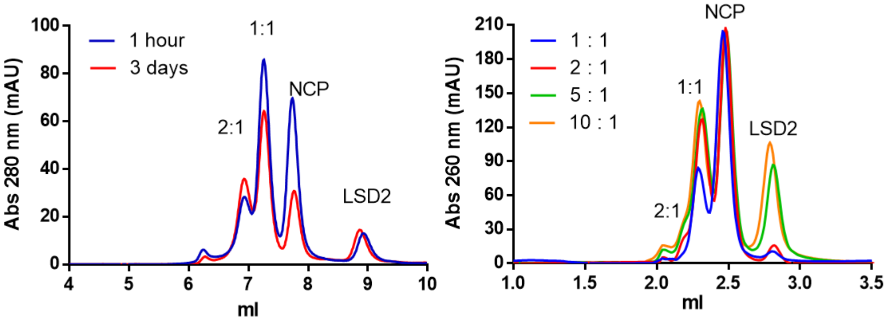


Figure 28. Effect of incubation time (left) and increasing LSD2/NPAC-linker to NCP ratio (right) on nucleosome recognition by LSD2. On the left, Prolonged incubation, up to 3 days leads to a small increase of the free nucleosomes/1:1 complex ratio, but at the expenses of partial precipitation of the sample, as can be deduced from the smaller area under the "3 days incubation" curve than that under the standard "1 hour incubation" curve. On the right, LSD2/NPAC-linker (1 to 5 molar ratio respectively) were incubated at increasing molar concentrations with the same amount of semi-synthetic nucleosomes. A significant increase of complex yield occurs when LSD2 concentration becomes twice the substrate concentration, but increasing amount of the demethylase enzyme do not further affect the final result.

If not differently stated, the protocol that was finally adopted on the basis of the best results is the following: 10 µM semi-synthetic nucleosomes are incubated for 1 hour with 20 µM LSD2 and 100 µM NPAC-linker in buffer 20 mM Tris pH 7.5 (4°C), 100 mM NaCl. Silica gel columns WTC-030N5 or WTC-030S5 (Wyatt Techonology) were used in running buffer 15 mM HEPES pH 7.3, 200 mM NaCl.

6.2 NPAC-DH analytical SEC

In order to assess the NPAC oligomerization state, a SEC step is performed with the Superdex 200 5/150 (GE Healthcare) using the ÄKTA Micro system (GE Healthcare) in analytical scale (Figure 29). Elution profile is monitored by UV absorbance at 280 nm, 260 nm and 348 nm. 348 nm corresponds to the wavelength to detect the reduced form of the nicotinamide cofactor, 260 nm to monitor the nucleic acid content (the ADP bound to nicotinamide), 280 nm corresponds to the wavelength at which tryptophan residues absorb. The column is equilibrated with the Dialysis-GF Buffer reducing NaCl to 100 mM. The salt content has been decreased to favour cofactor binding inside the enzyme active site.

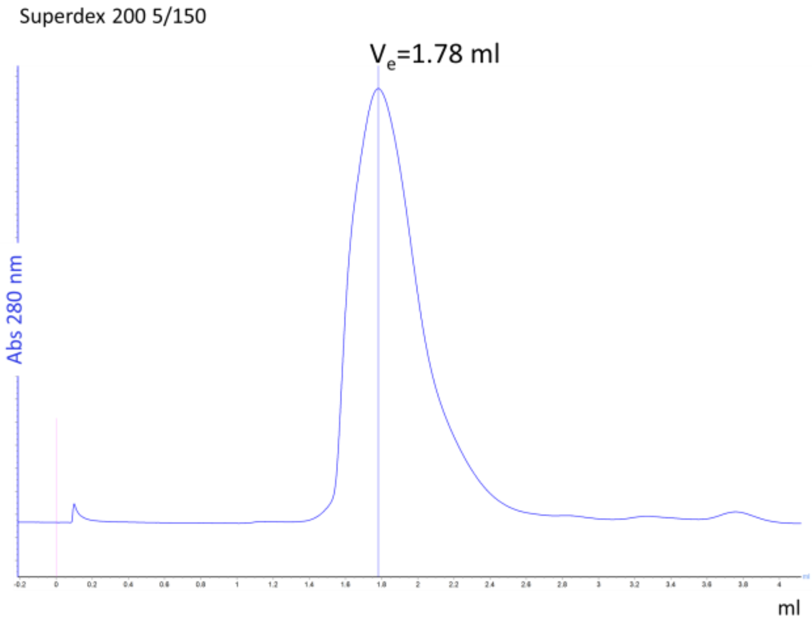


Figure 29. Analytical Superdex 200 15/150 of the purified NPAC-DH.

6.3 LSD2/NPAC Δ205

LSD2 Δ30 (in 20 mM Tris pH 8, 100 mM NaCl, 1 mM TCEP) was mixed with NPAC Δ205 (in 50 mM HEPES pH 7.5, 500 mM NaCl, 5 % glycerol, 0.5 mM DTT) at different molar ratios in 100 mM NaCl final concentration. LSD2 Δ30 was incubated with NPAC Δ205 and semi-synthetic nucleosomes for one hour in 20 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT. The mixtures were loaded onto silica gel columns WTC-030N5 or WTC-030S5 (Wyatt Techonology) equilibrated in 15 mM HEPES, pH 7.3 (25 °C), 200 mM NaCl. The elution profiles were recorded at 214, 260, and 280 nm.

7. Thermostability assays

7.1 NPAC DH Thermofluor protocol set up

Thermofluor (Biorad) is a useful technique to assess protein stability. The instrument actually consists of a thermocycler coupled with a fluorescence detector. The sample is warmed up from 20° C to 90° C, whereas the fluorescence emitted at a specific wavelength is detected and registered throughout the experiment. The protein of interest is incubated with a particular dye, able to couple absorption to a fluorescent emission only when bound to hydrophobic aminoacids. Such residues are usually hidden within the interior, hydrophobic core of folded proteins in solution, and they become exposed only upon protein unfolding. As heat denatures the protein, the dye has access to the hydrophobic patches of the polypeptide, and thus its fluorescence increases. In a standard Thermofluor experiment, the fluorescence intensity is plotted against the temperature (Figure 30). Through calculation of the first derivative, Temperature of melting (Tm) is assessed, which corresponds to the temperature at which the protein unfolds more rapidly, and is a measure of the protein stability. Thermofluor experiments allow characterization of the protein behavior in different conditions: buffer, pH, in presence of ligans, inhibitors or cofactors as well. The higher the Tm, the higher the stability of the protein. After extensive screening for the optimal conditions of protein and dye concentrations, I selected those samples exhibiting a melting curve with minimal influence from the Sypro Orange signal at around 58 °C (Figure 30). Therefore, the final protocol was set up as follows: 25 µM NPAC DH is mixed with the fluorescent dye SYPRO Orange (Invitrogen), provided 5,000x concentrated in DMSO and diluted 1,000 times till 5x final concentration, in a final volume of 20 µl in Dialysis-GF buffer. Once the microplate has been filled, it is sealed with optical-clear quality sealing caps (Biorad). Thermograms are recorded using the Opticon software (BioRad) which allows to monitor fluorescence of HEX which is compatible with SYPRO Orange maximal absorption at 470 nm, and its maximal emission at 569 nm. A temperature gradient is applied, starting from 20° C and increasing till 90° C, measuring the fluorescence signal every 0.5° C. NPAC DH thermostability was also measured in presence of the cofactor NADH or NADPH in a molar ratio of 1:2, with the aim to establish which of them is better retained by the protein. The analysis was performed in triplicate.

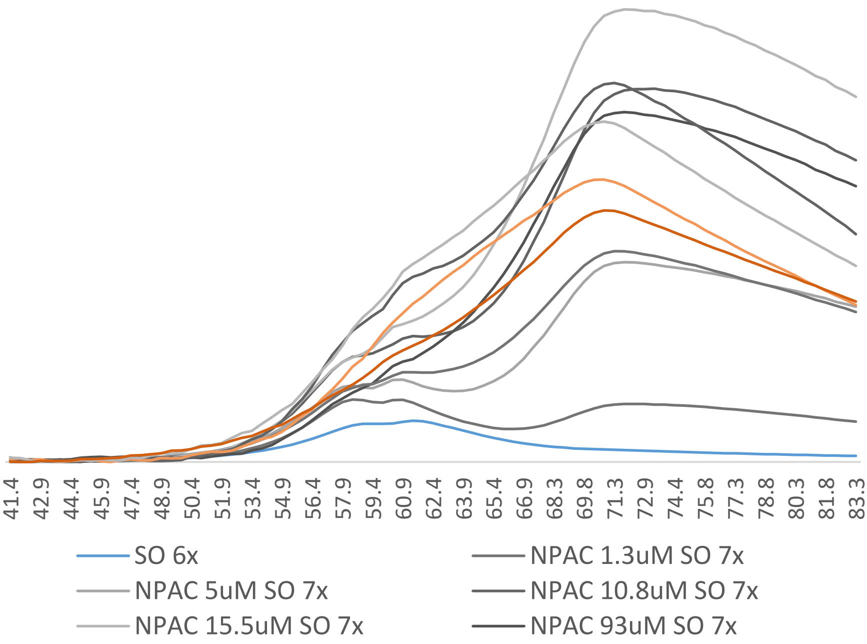


Figure 30. Thermal Denaturation Assays on NPAC DH using Thermofluor. Various protein concentrations were tested in presence of various dilutions of the fluorescent dye Sypro Orange and heated up to 90 °C. here are shown the most representative ones. Fluorescence signal is a sigmoidal curve against temperature (X axis). As shown by the image on the top, with increase in temperature protein's hydrophobic regions become gradually more exposed to the dye, wich regains its fluorescence once bound to the protein. Maximal fluorescence intensity is obtained when the protein unfolds completely. The two-step shape of the curves reflects in part the signal from the sypro orange dye alone (see SO 6x sample), and in part the quaternary conformation of NPAC DH. Indeed hydrophobic interactions between NPAC DH monomers are the first ones to be disrupted by temperature. Hence the hydrophobic monomer-monomer interface is exposed to the dye and a first increase of fluorescence is seen. The second slope detected is the one effectively correlating to the protein melting temperature, as no further increases in fluorescence are seen at higher temperatures.

8. Electrophoretic methods

8.1 Native page

Acrylamide gels were prepared with 6% (w/v) liquid acrylamide in TBE buffer (17.8 mM Tris pH 8.2, 17.8 mM Boric Acid, 0.4 mM EDTA). Otherwise, a Mini-PROTEAN TGX Precast gel (Any-KD, BIO-RAD) was run without samples for 3 hours at 4°C, at 150 V. Then the samples were loaded using as sample buffer a solution containing 5% (w/v) sucrose in TCS buffer and run for 1 hour and 30 minutes at 150 V for the pre-cast gels, and at 100 V for the 6% acrylamide gels at 4°C in TBE buffer (22.5 mM Tris/borate pH 8.3, 0.5 mM EDTA). To follow the run we also loaded few microliters of glycerol and Coomassie Blue in a separate well. As the run ended, the gel was stained with Sybr-Safe to check for DNA, and then with Coomassie Blue for proteins.

8.2 Western Blot

Western blots were performed to check the integrity of octamers and nucleosomes stored at -80°C. First, an SDS-PAGE was run with a coloured marker (we preferentially used precat gels, AnyKD from BioRad). With Turbo-Transfer (BioRad) apparatus, we transferred the proteins to PVDF membrane. The membrane was blocked in TBS-T buffer (Tris 100mM pH 7.5, NaCl 150 mM, Tween20 0.1%, using MIlliQ water) with 6% of milk, for 1 h at rt. Then the primary antibody was added, diluted in TBS-T + milk 5%. After one hour, three washes were performed in TBS-T, leaving the membrane for 10 minutes in the washing buffer, each time. Then the secondary Ab previously diluted in TBS-T + milk 5% was incubated for 30 minutes at rt. Three washes with TBS-T, 10 minutes each cleaned the membrane. For the photographic plate development, at darkness, the following steps were done: preparation of fresh reactive solutions Immuno-Star HRP Luminol/Enhancer and Immuno-Star HRP Peroxide Buffer. For each solution 1 ml was added directly on the membrane, possibly were the proteins (bound by the conjugate antibodies) should be. After 5 minutes the excess is rapidly removed and the membrane is covered by a plate for 10 seconds (more or less, varying with the efficiency of the luminol production). The plate is immediately put in developer Buffer, and, as the bands appear, washed in water and placed in Fixer Buffer.

9. SPOT-assay

9.1 Characterization of NPAC-PWWP histone-binding properties

In order to characterize the hypothetical chromatin-binding abilities of NPAC PWWP domain (see homologous sequence alignment in *NPAC PWWP cloning* section), we went for a high-throughput assay to detect typical protein-protein interactions between chromatin readers and histone peptides. Our collaborators, Panagis Filippakopoulos and Sara Picaud at the Structural Genomics Centre in Oxford, prepared an array of H3.1, H3.3 and H4 peptides, each one synthetized over the same amino-functionalized cellulose membrane (Whatman™ Chromatography paper Grade 1CHR, GE Healthcare Life Sciences #3001-878, Little Chalfont, UK). We selected H3 and H4 histone tails, as they were the most promising candidates for binding by NPAC, mainly because homologous PWWP domains were known to recognize H3K36me3 and H4K20me3 (Weaver, Morrison and Musselman, 2018). Once the freshly prepared membrane was ready, I went through the blotting protocol with the 6xhis-Flag-SUMO-NPAC PWWP (1-105) protein, according to sections 2.5 and 2.6 of the Materials and methods section published by our collaborators (Picaud and Filippakopoulos, 2015). Briefly, the His-tagged protein was blotted onto the membrane, pre-blocked with 5% BSA. After washing, an anti-his antibody, coupled to HRP was incubated on the membrane, washed, and eventually detected with a Pierce® ECL Western blotting Substrate (Thermo Scientific, distributed by Fisher Scientific). The resulting chemiluminescence signal (Figure 48) from each peptide-spot was quantified with the Kodak 1D V.3.6.2 Scientific Imaging System and the data were exported on Microsoft Excel.

10. Fluorescence polarization

10.1 Characterization of DNA binding by LSD2 and NPAC-PWWP

To measure the binding affinities of LSD2 for DNA sequences or histone peptides, FP assays were carried out on Clariostar plate reader (BMG Labtech), monitoring the change in polarization properties of fluorescently labelled DNA or H3(1-21) peptidewith tetramethylrhodamine (TAMRA). Fluorescent molecules are able to adsorb and re-emit photons in a characteristic polarization plane. The emitted photon has a lower energy, because of the fact that part of the energy is retained by the excited electron of the fluorophore. When fluorescent molecules are excited in solution, they continuously move and rotate, and that is why the emission spectra won’t be in the same lane of polarization, but it will be mostly depolarized. As the speed of molecular rotation decreases, because of size and shape of the molecule itself, scattering of the re-emitted light is lowered. On the contrary, excitation of small molecules, able to rapidly rotate and move with more freedom degrees, will produce a greater proportion of scattered light. Fluorescence polarization is based on the relative amount of scattered and polarized light. It is a measure of the freedom degrees of a fluorescent molecule. The rotation speed is directly dependent on both size and shape of the labelled compound, and they change after complexing of the fluorophore to another molecule. To perform DNA and histone binding assays, we used this quick and sensitive method, with a Clariostar instrument. All experiments were perfomed at room temperature and in the same buffer (15 mM KH_2PO_4 pH 7.2, 1 mg/ml BSA, 5% glycerol and 0-100 mM KCl), in order to have comparable results. Two DNA sequences, differing by a G/T mismatch at position 7, were covalently labelled with TAMRA at their 5’ at the peptide synthesis facility of Netherlands Cancer Institute, The Netherlands. Here are the sequences of the oligomers:

DNA 1: 5’-TAMRA-AGTCGCCAGGAACCAGTGTCA-3

DNA 2: 5’-TAMRA-AGTCGCCAGGGACCAGTGTCA-3’

To prepare the affinity assay with histone N-terminal tail, peptides we used the following primary sequence:

A R T Kme2 Q T A R K S T G G K A P R K Q L A

The peptide was purchased from Sigma, already covalently labelled with TAMRA fluorophore at its C-terminus. Briefly, experiments were carried out in 15 mM Tris pH 8, 0.01 % Tween 20, at 25°C,with at 0 mM and 100 mM– 100 - 150 mM NaCl and 5 nM DNA. for LSD2.

For the PWWP domain, the assay was performed with 5 nM labelled DNA in 15 mM Tris pH 8, 0.01 % Tween 20 at 25 °C, at 0 - 50 - 100 - 150 mM KCl using serial dilution of PWWP starting from 16 μM in the first well. All the experiments were done in triplets. Dissociation constants were calculated from the regression curve, measured by fixing fluorophore concentration value with GraphPad Prism.

11. Crystallographic studies

11.1 Introduction to the method

“*All the methods were then available for studying protein structure by X-ray crystallography were [thought] bound to fail*”. (*John C. Kendrew during his Nobel lecture, December 2, 1962)*

John Kendrew was the first to unlock the structure of a protein from X-ray diffraction data in 1957, and since then X-ray crystallography became widely used to gain insights onto a protein’s structure and hence its function. As X-rays cross an ordered array of atoms (a protein crystal for example), the electromagnetic wave sums up at particular positions of space and time, exactly where the distance between the encountered atoms equals an integer number of the wavelength pace. This results in a significant increase of the amplitude of the wave only along specific directions (angles of diffraction). The more the atoms encountered in this periodic array, the higher the amplitude of the sum. As the X-ray wavelength ranges between 0.7 to 1.5 Å, the diffracting objects have to be in the same order of magnitude, exactly as atoms, molecules and eventually proteins. Actually, it is not “atoms” giving rise to X-ray diffraction, but electrons. Hence the more electron-dense is a particular position of the crystal lattice (for example, a methionine sulfur atom, with respect to an arginine nitrogen atoms for example), the higher the contribution to diffraction spots. This proved to be particularly helpful in the reconstruction of unknown proteins and nucleic acid structures (the phosphate backbone of DNA diffracts stronger than the inner bases), for which the starting, electron-rich atom coordinates could have been set. At Kendrew’s time, the main problem of the technique was the solution of the diffraction pattern and the reconstruction of the diffracting object present in the crystal lattice.

Nowadays the bottleneck of crystallographic studies is the quality of protein crystal. This means that all proteins in a crystal lattice have not only to be arranged in an extremely regular three-dimensional pattern, but also they have to assume in the same, identical conformation and orientation. This can be really more difficult than expected, because even very small proteins are composed by hundreds of atoms, connected by hundreds of bonds, with hundreds degrees of freedom (Ramakrishnan, 2018). Only repetitive units in an ordered lattice, can give rise to the amplification of the diffracted X-ray (Smyth and Martin, 2000). In order to obtain the highest quality possible protein crystals, a drop of the sample protein in solution has to be dried, slowly enough to allow proper interaction and positioning of the single polypeptides with respect to each other, but faster than the lifetime of the protein stability itself. Moreover, it has to be taken into account that the rate at which the first crystal nuclei form is different for the various proteins and governed by thermodynamic forces that are not identical to those determining the rate at which these nuclei can grow to pure, diffracting protein crystals. The concentration of the protein in solution, as well as the presence of salts and other chaotropic agents, are the mainly used adjustable parameters for the crystallization process (Figure 31). Control of contaminants, ions, protein ligands and cofactors, other additives and protein stabilizing agents, along with pH, temperature and humidity is also part of the routine setting of a crystallization experiment. However, no clear rationale still is able to predict the crystallization conditions of a new protein. The widespread use of semi-automated robots for crystallization allows high-throughput screening of hundreds of crystallization conditions, which greatly slows down the time-consuming preparation and set up of crystallization trials. Once crystals have grown, they must be harvested from the droplets using a nylon loop, into which a drop containing the protein crystal is hold by capillarity. The crystal is protected from possible cryogenic damages through a brief immersion in a solution, identical to the crystallization mix, but containing 20% v/v cryo-protecting agents, such low molecular weight PEG or glycerol. Formation of ice around the crystal must also be avoided during following transfer step, for example during crystal mounting on the goniometer of the beamline (Karplus and Diederichs, 2012).

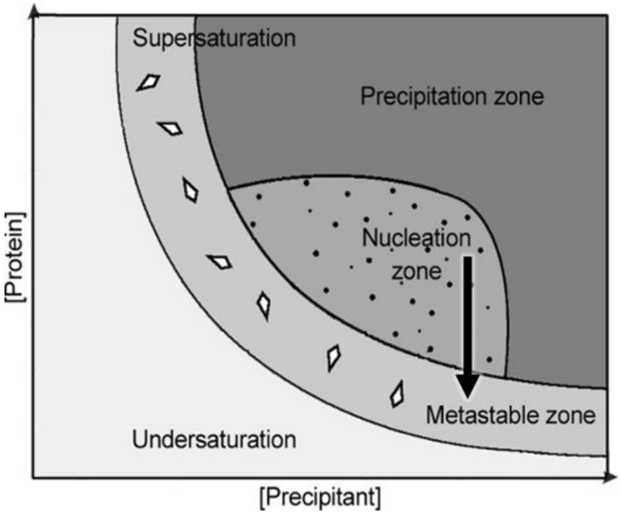


Figure 31. Crystallization phase diagram. The protein crystalizes at supersaturating concentration. Crystallization methods aim at placing the protein at the conditions of the nucleation zone, where crystal nuclei form. As the concentration of the monodispersed protein in solution decreases (black arrow), crystal growth occurs. At undersaturating conditions, the protein remains in solution. Adapted from Pichlo *et al.*, 2016.

11.2 NPAC-DH crystallization

Crystallization trials were set up with the purified protein at 10 mg/ml concentration. For each condition, a drop was also prepared with the buffer in which the protein was dissolved, as a control in case of salt crystals formation. The drops were checked the following days after their deposition for the first week and then once per week. Depending on the results, the most promising conditions were optimized by manual crystallization in vapour diffusion.

The best condition turned out to be 0.2 M $(NH_4)_2$tartrate, 19% PEG 3350. Crystals were long, thin, extremely fragile during fishing and poorly diffracting (Figure 32). Microseeding proved to be the right strategy to obtain a crystal diffracting at 3.6 Å resolution. Microseeding proved to be the right strategy to obtain a crystal diffracting at 3.6 Å resolution. Co-crystallization of NPAC-DH with either NADH or NADPH in molar excess did not give any promising result. The cryo-protectant solution was freshly prepared before fishing the crystals, and it contained 0.2 M (NH4)2tartrate, 19% (w/v) PEG 3350 and 20% (v/v) glycerol. X-ray diffraction data on crystals of NPAC-DH were collected on Dectris detectors of the Pilatus generation at the ESRF (Grenoble, France) and the SLS (Villigen, Switzerland) synchrotron beamlines.

Figure 32. NPAC-DH crystals in 0.2 M (NH4)2tartrate, 19% (w/v) PEG 3350 are needle-shaped, transparent and extremely fragile.

12. Cryo-electron microscopy

12.1 Introduction to the method

Cryo-EM is the new emerging method for structural analysis of macromolecular assemblies, which is expected to supersede X-ray crystallography in the next few years. Indeed, the number of cryo-EM structures published from since 2013 is already equal to those obtained by crystallography in the 90's (Savva, 2019). Cryo-EM generally describes any miscroscope experiment performed at low temperatures. Joachim frank, Jacques Dubochet, and Richard Henderson received the Nobel prize in Chemistry 2017 for their contribution to the current "resolution revolution" in microscopy. Actually cryo-EM developed from electron crystallography of 2D-crystals, no more than few unit cells thick. Differently from X-ray crystallography indeed, electron crystallography uses images as primary data, since the specimen is a weak phase object, with zero amplitude effects. The first alpha-helices were visualized in 1975 by Henderson and Unwin, for 2D-crystals of rhodopsin (Schertler, Villa and Henderson, 1993). As specimen damage and motion are greatly reduced in cryo-genic temperatures, a method for plunge-freezing protein samples on the grid while preserving their native states was implemented by Dubochet and colleagues (Adrian *et al.*, 1984). Joachim Frank studies on single particle alignment and superposition (Frank, 2010), further projected the cryo-EM method into the future of structural biology. In the last years, more powerful methods for single-particle image processing were developed, with a particular mention to Sjors work (Scheres, 2012). However, the breakthrough came with faster and more powerful CCD cameras and microcircuitry, which really paved the way to the current "resolution revolution" in cryo-EM. The advantage of cryo-EM over X-ray crystallography and NMR is that huge protein complexes can be studied. Moreover, sample preparation is significantly less expensive and time consuming, as only 4 microliters of a low-concentration (typically around 50nM for a 100 kDa protein) specimen are required (Table). Moreover, small impurities and disordered regions of the protein sample are not an issue as big as in crystallography and NMR. On the contrary, new techniques are capable of providing dynamic information based conformationally different structures. The only limitation for cryo-EM was thought to be the size of the target structure, as structures below 100 kDa were thought to be very difficult to align. In the particular case of our project on LSD1 and LSD2, since various attempts to co-crystallize either of these enzymes with NCP failed, we moved to cryo-EM to undertake structural studies. The size of the complex (about 300 kDa) and the nucleosome symmetry were considered advantages for particle picking and alignment. To this aim, the first step was to optimize the preparation of the sample as described in the next section.

12.3 Sample preparation.

Milligram quantities of the LSD2/NPAC-linker/NCP complex were obtained from incubation of 20 µM semi-synthetic nucleosomes and 30 µM LSD2Δ30/NPAC-linker (1:5 molar ratio) in their storage buffers for 2 hour on ice. The sample was then purified on Superdex 200 10/300 (three columns connected in series, for a total bed volume of 72 ml) equilibrated in 15 mM Tris pH 7.5, 0.4 mM EDTA, and 200 mM KCl. The elution profile was recorded at 260 nm, 280 nm and 400 nm using an AKTApurifier10 (GE Healthcare). Four species were eluting in the following order: (LSD2/NPAC-linker)$_2$/NCP; LSD2/NPAC-linker/NCP, NCP, LSD2. The identity of the peaks was obtained through comparison of the relative 260nm/400nm absorbance ratio, which reflects the DNA to covalently bound flavin of LSD2 ratio (Figure 33).

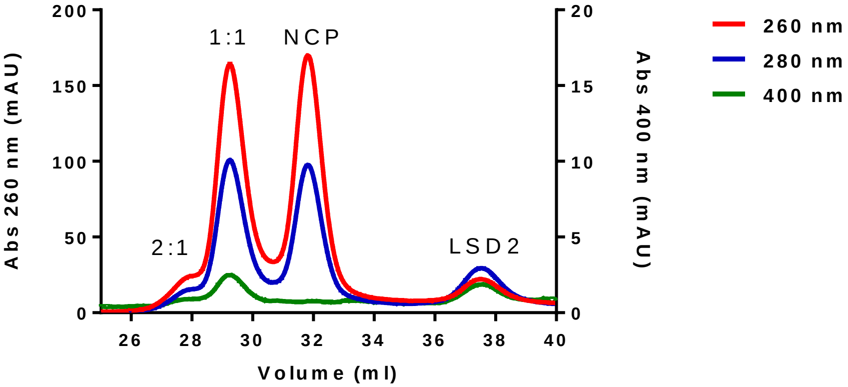


Figure 33. Elution profile of the LSD2/NPAC-linker/NCP sample used for cryo-EM studies. Protein, DNA and the flavin covalent adduct were detected monitoring the absorbances at 280, 260 and 400 nm respectively. Three Superdex 200 10/300 columns (GE healthcare) connected in series were used, in running buffer 20 mM tris/HCl pH 7.5 (4°C), 200 mM KCl.

We isolated the factions corresponding to the LSD2/NPAC-linker/NCP 1:1 peak. The running buffer of the preparative gel filtration is 15 mM Tris/HCl pH 7.5, 0.4 mM EDTA, 200 mM KCl. Eluting fractions of interest were immediately mixed with an equal volume of 15 mM Tris/HCl pH 7.5, 0.4 mM EDTA buffer. The sample was then concentrated by centrifugal ultrafiltration with an Amicon Ultra 30 KDa cut-off (Merck Millipore, Germany), and stored on ice for a maximum of 2-3 weeks (storage buffer is 15 mM Tris/HCl pH 7.5, 0.4 mM EDTA, 100 mM KCl).

12.4 Grid preparation.

Just before blotting of the cryo-EM grid, the sample was diluted in 15 mM Tris/HCl pH 7.5, 0.4 mM EDTA buffer to reach a final concentration of about 3 μM in buffer 10 mM KCl, 20 mM Tris pH 7.5 (4°C). The grids used for the high resolution data collection were prepared at the Grenoble Instruct Center (Grenoble, France). We used copper grids covered by a layer of holey carbon of the Quantifoil type (400 mesh, R2/1, Figure 34), glow-discharged for 5 minutes before use. Glow-discharging employs the reactivity of ionized water (present in the normal atmosphere, but also specific mixtures are used) over the surface of the grid to be cleaned. Hydroxy and Hydride ions react with the exposed grid surface to form aldehydes, alcohols, carboxylic acids and other types of hydrophilic moieties, also depending on the plasma composition. In this way, the surface of the carbon layer on the grid is rendered entensively hydrophilic, to allow the distribution of the sample in acqueous solution equally over the grid squares. On the contrary, the sample drop would minimize its contacts with the hydrophobic surface of the carbon, and subsequent drying with filter paper would remove the entire drop.

Drying of the blotted grid is a very important step to control the concentration of the particles on the grid, and the thickness of water in between carbon holes. Indeed, to control evaporation of water molecules and consequent concentration of the sample, we kept the grid in a 90% humidity camera. Blotting with a filter paper removes excess of solution, and this step also is quite delicate. Nowadays, robotization of the entire process between blotting and freezing of the grid made it a more reproducible process. In a Vitrobot (FEI), we placed 4 μl of sample onto the grid, blotted for 2 seconds in 100% humidity atmosphere at 20°C and plunged into liquid ethane. Liquid ethane or liquid propane are the elements of choice to rapidly cool down samples as they have a thermal capacity much higher than liquid nitrogen (10^6 degrees/second), despite being at the same temperature (-185°c for liquid helium, whereas liquid nitrogen is usually at – 190°C).

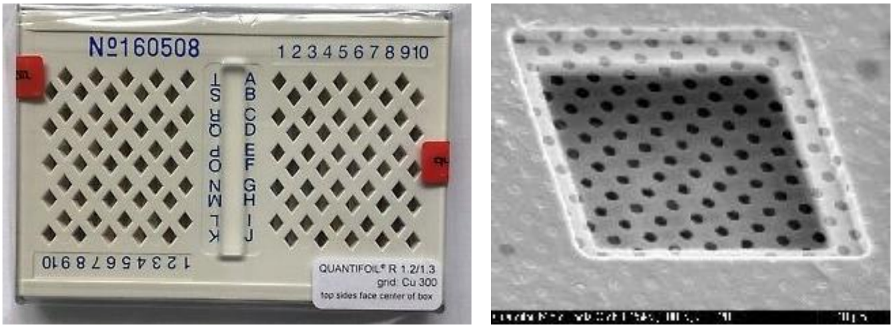


Figure 34. The holey carbon grids purchased from Quantifoil. On the left, the box containing the grids, which have to be handled carefully with special tweezers, in order not to kink the plane of the grid and not to scratch the holey carbon side. On the right, a zoomed view of one single square of the grid: the copper support holds a very thin layer (about 20 nm thick) composed roughly half of plastic and half of the carbon cover evaporated on it. Images taken from Quantifoil commercial website (www.quantifoil.com).

12.5 Data acquisition.

The screening and the first datasets collected to evaluate the feasibility of the project were carried out on a 200 kV microscope (Tecnai Arctica, FEI) equipped with a Falcon II camera (FEI) at the cryo-EM unit of the Nanyang Technological University (Singapore) and on a 300 kV microscope (Polara, FEI) equipped with a K2 Summit (Gatan) at the Grenoble Instruct Center (Grenoble, France). The frozen grids were clipped and loaded on a 300 kV TEM microscope (Titan Krios, FEI/Thermo Scientific) equipped with a K2 Summit (Gatan) at the eBIC of Diamond Light Source (Didcot, UK).

Once the grids were loaded onto the stable microscope, we selected those “holes” onto which promising images could be collected. Best grid squares and holes were selected among those without contaminants, ice crystals, and scratches, kinks or cuts on the grid support. For each of the holes (Figure 34, right) we collected three micrographs (an example micrograph is hown in Figure 35), taking care that in each of the micrograph was included also part of the carbon support. This strategy allows a more precise calculation of the experimental defocus for each image.

For the dataset with proposal number em16082, the images were recorded at 130 kX magnification in electron counting mode, pixel size of 1.06 Å, energy filter of 20 eV and defocus range between 0.7 and 3.05 µm. A total of 2078 40-frames movie stacks were collected with a flux of 50 e^-/A^2 over a total of 8 sec of exposure time (fluence 6.25 e^-/A^2 per sec). Electron counting detectors are the latest innovation in detector technology. Counting of single electron events at a specific pixel, instead of integrating the signal over the affected pixels, significantly increases the DQE (Detector Quantum Efficiency, that is the probability to detect any electron event), at any frequency till the Nyquist. The significant improvement of signal-to-noise ratio due to the new generation of electron detectors is thought to have been the main driver of the recent success of cryo-EM. Together with high-speed microcircuitry, which can collect many frames per image (up to 400 frames per second), the high sensitivity of a detector allows data collection at lower doses, that means lower sample damage.

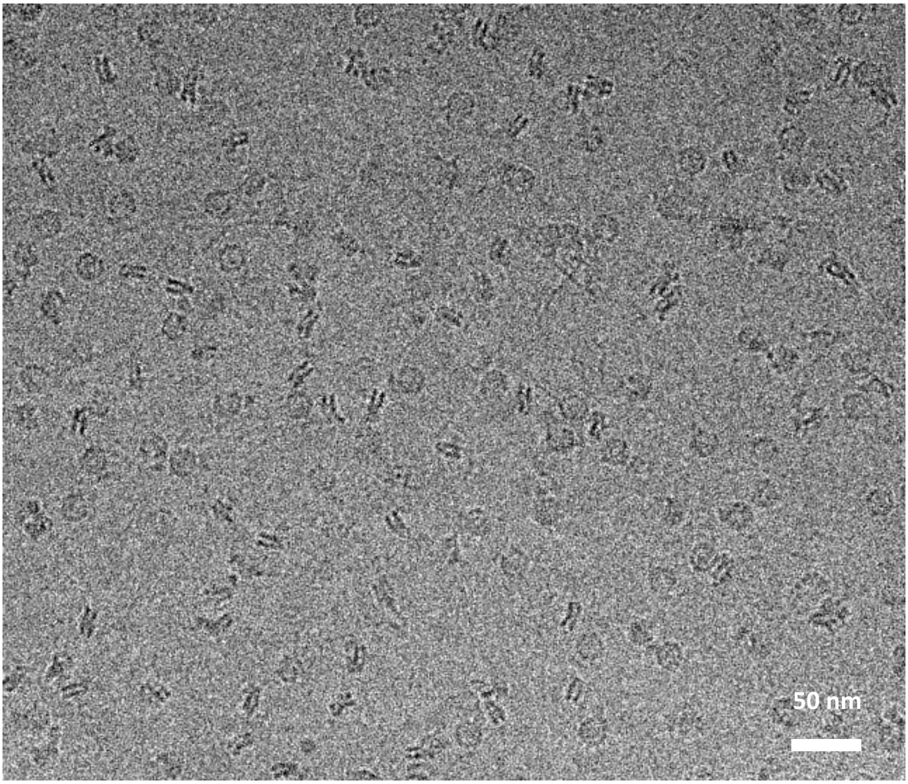


Figure 35. An example image used collected at the eBIC facility in Diamond under the proposal ID em16082. Single nucleosomes and complex particles in various orientations can be very easily detected by eye. Scale bar size is 50 nm.

12.6 Data processing.

The preliminary 840 images collected during sample optimization at the cryo-EM unit of the Nanyang Technological University in Singapore were converted in hdf files in EMAN2.12. Particles were picked up manually in e2boxer.py and the good micrographs were converted back to .mrc files (64MB) in e2proc2d.py. CTF was corrected with CTFfind3 (Rohou and Grigorieff, 2015) and further processing was performed in RELION 1.4 (Scheres, 2012). After 3D refinement, a low resolution map (17 Å) was obtained and used for design of mutational experiments. Thanks to our collaborators Sriram Subramaniam and Sagar Chittori, we could perform image processing of our tricky data, eventually in the most unbiased way.

All movie frames were aligned and corrected for beam motion using MotionCor2 (Zheng *et al.*, 2017). In particular, this software takes into account the initial beam-induced specimen motion (there is indeed a usual first "jump" of the specimen as it receives the first 2-4 e-/A2), and of the significant degradation of the sample, further motion and charge accumulation, in the last frames of the movie. By weighting the information of each frame, depending on its relative quality, over the final alignment, the final result is usually more reliable. As previously stated, images were collected in defocus, ranging from 0.7 to 3.4 micron. Actually all cryo-EM samples are studied in defocus, since the sample is a weak phase object, and no amplitude contrast would be present in a focused image. Defocusing allows to detect phase interference effects between the waves carrying the image of the object potential, and the unscattered wave. When certain ranges of object periodicities are present in the image, there are contrast enhancements relatively to the frequency of the wave, the defocus plane and the spherical aberration of the microscope optics. The amplitude of the phase contrast lowers with increasing frequencies, that means resolution. In essence, this means that the furthest the defocus, the higher the contrast, with major losses of high-resolution information (Getting started to cryo-Em. Online course by Grant Jensen at Coursera, CalTech). Viceversa, the closer to the focus an image is collected, the higher the proportion of high-resolution information, at the expenses of lowering the contrast.

For a typical cryo-EM experiment, images are collected in a range of defocus values, in order not to lose too much high-resolution information, while at the same time retaining the capability of visually detect the individual particles. The CTF (Contrast Transfer Function) is the Fourier transform of the Point Spread Function of the experimental image. The CTF describes how the information present on the image is affected by interference with the unscattered wave, defocusing and spherical aberrations. Other microscope optical aberrations among the Zernike polynomials could be taken into account, yet chromatic aberration, "coma", and astigmatism should be corrected before data collection, during the microscope alignment procedure (cryoEM course published by Cambridge LMB institute in 2017 on youtube. Classes by Sjors, Savva and Russo). CTF correction of the images is necessary to recover those negative contribution of the wave carrying the image of the object potential, to the detected signal, and correct the image accordingly. We also used the images of the Thon rings calculated from each micrograph, to select and discard those with evident spherical aberration (usually referred to as simply “astigmatism”), or ice present. We used CTFFIND4, implemented in Grigorieff lab (Rohou and Grigorieff, 2015).

After various attempts of manually picking particles, which resulted in ambiguous outcomes, we used a Gaussian blob as reference to proceed with automated particle picking in RELION 2.1 (Scheres, 2012). Proper particle picking and alignment are essential to obtain high-resolution data, as these processes are usually performed in Fourier space, with low-pass filtering. Hence, un-biased particle picking was definitely an important choice in our case, where the clearly recognizable nucleosome shape would have hampered recognition of all those complex particles in which the nucleosome low-frequency characteristics were not recognizable. Multiple rounds of 2D- and 3D-classification yielded the final five classes, as explained in the paper here attached (Marabelli *et al*., 2019).

Results and discussion

1. NPAC-linker facilitates nucleosome demethylation by LSD2

Investigation on the LSD2/NPAC-linker system through a combination of mutagenesis, fluorescence polarization and analytical SEC methodologies revealed a new mode for nucleosome recognition, as described in detail in the following sections and reported in the attached article (Marabelli *et al.*, 2019).

1.1. First cryo-EM studies of the LSD2/NPAC-linker/nucleosome complex

Thanks to our collaboration with the group of Daniela Rhodes and Sara Sandin at the Nanyang Technological University in Singapore, we had the possibility to optimize the sample purification and grid preparation conditions. A preliminary low-resolution (about 17 Å) cryo-EM map of the covalently attached LSD2/NPAC linker-semisynthetic nucleosome complex showed that LSD2 is lying at the edge of the nucleosome, where the only interaction possible could have been that with DNA (Figure 36). However, given the low resolution of the map, it was not possible to evaluate the orientation of LSD2/NPAC-linker with respect to the nucleosome.

Given the homology between LSD2 and LSD1, where LSD1-CoREST1 is known to recognize the nucleosome through DNA binding (Pilotto *et al.*, 2015), our hypothesis was contemplating a similar DNA-mediated mechanism for LSD2. LSD1/CoREST1 has higher nucleosome affinity for an NCP with protruding DNA strands (Kim *et al.*, 2015).

Thus, we firstly investigated LSD2/NPAC-linker binding of the more physiological 167-bp NCP. However, probably because of the DNA sticky ends, the LSD2/NPAC-linker/NCP complexes were not eluting separately from the free nucleosome sample (Figure 37). For the same reason, and also because of the variability among alkylated nucleosomes preparations, we could neither conclude anything on the effect of nucleosomal DNA on complex formation. We then focused our efforts on the study of the 147-by NCP recognition and binding by LSD2/NPAC-linker.

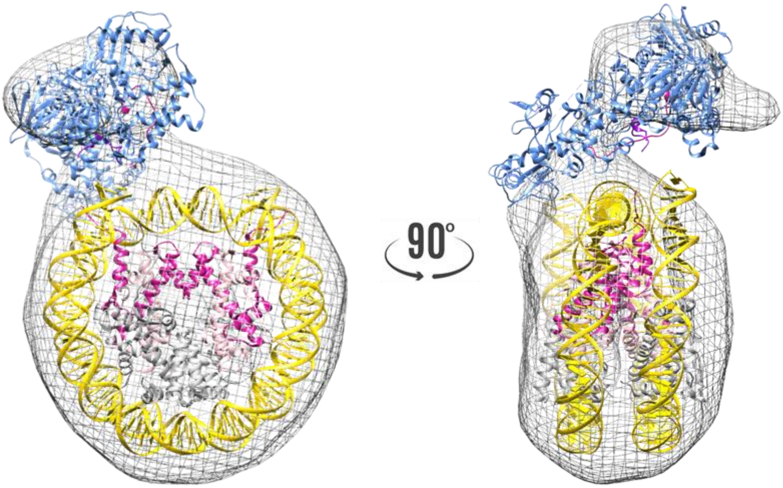


Figure 36. Cryo-EM map of the LSD2/NPAC-linker/NCP complex at 17 Å resolution. A tentative fitting of the PDB structures of the single 147-bp nucleosome (PDB: 6ESF), and of the LSD2/H3(1-26)/NPAC-linker (residues 214-225) crystal structure (PDB:4HSU), was performed in Chimera. Nucleosomal DNA is coloured in gold, histone proteins H2A, H2B, H3 and H4 are in dark grey, light grey, violet red and pink respectively. LSD2 is coloured in cornflower blue, NPAC linker segment is purple and the H3 tail is violet red for consistence with the nucleosome.

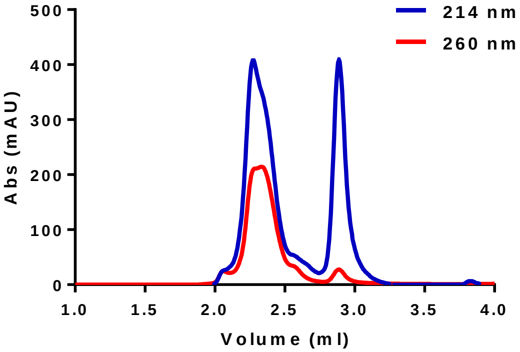


Figure 37. Elution profile from an analytical SEC column (Wyatt 030N5) of the mixture LSD2:NPAC-linker:NCP incubated at 2:10:1 molar ratio for 1 h on ice. Running buffer is 15 mM HEPES pH 7.3, 200 mM KCl. The peaks corresponding to LSD2/NPAC-linker/NCP and free NCP species overlap.

1.2 LSD2 alone binds DNA with nanomolar affinity

Binding of human full-length LSD2 to DNA and substrate peptides, with and without the NPAC-linker cofactor, were assayed through fluorescence polarization. In standard experiment conditions (i.e. without NaCl; see paragraph 10 in theMaterials and Methods section), LSD2 has ana dissociation constant (k_D)for DNA of about 133 nM, significantly higher than the affinity for the 21-aa long histone H3 N-terminal peptide (900 nM; Figure 38). In addition, experiments in presence and in absence of NPAC-linker revealed that there is no change in affinity for either DNA nor the H3 substrate tail peptide, differently from what had been stated before (Fang et al., 2013).

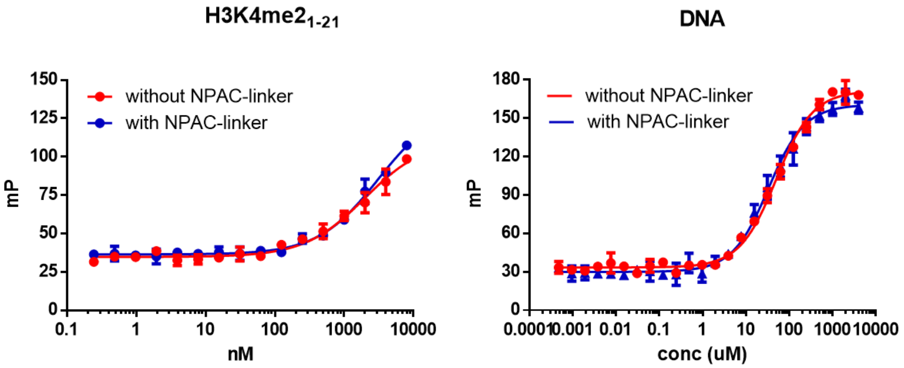


Figure 38. LSD2 FP binding profiles for short stretch of DNA or histone tail peptides at different concentrations. The calculated affinity of LSD2 for DNA oligos is 133 nM, whereas that one for H3 histone N-terminal 21-mer peptides is 900 nM.

1.3 LSD2 unspecifically interacts with nucleosomal DNA

To have a first indication of the surface patches which might be involved in DNA binding I then used the online software DNAbinder (Kumar, Gromiha and Raghava, 2007), whose output is a list of prediction scores for the DNA-binding ability of each protein amino acid. Among these, on the basis of LSD2 structure I finally selected 9 positively charged amino acids for mutagenesis studies, which likely form stabilizing interactions with the DNA phosphate backbone of the nucleosome. Most of these are Lys and Arg residues that locate on the rigid Zn-finger domain. Further side chains were selected in the surroundings of the H3/NPAC binding cleft. We also deleted the positively charged N-terminus of LSD2, and part of the structurally disordered but conserved loop connecting the SWIRM and the Zn-finger domains (Figure 39).

To dissect the molecular interactions between DNA and LSD2 surface electrostatics, I tested the retained DNA-binding ability of the mutants through fluorescence polarization (Figure 40, left; Table S1 in Marabelli *et al.*, 2019). At the standard assay conditions (no salt) the only significant difference in DNA affinity was detected for the N-terminally deleted protein (Figure 40, right). Then I tested the specificity of DNA recognition by the N-terminal tail, by increasing the ionic strength (Figure 40, right). It appears evident that DNA binding by LSD2 is strongly dependent on unspecific electrostatic interactions, mainly driven by the N-terminal tail and not by any specific patch on LSD2 surface.

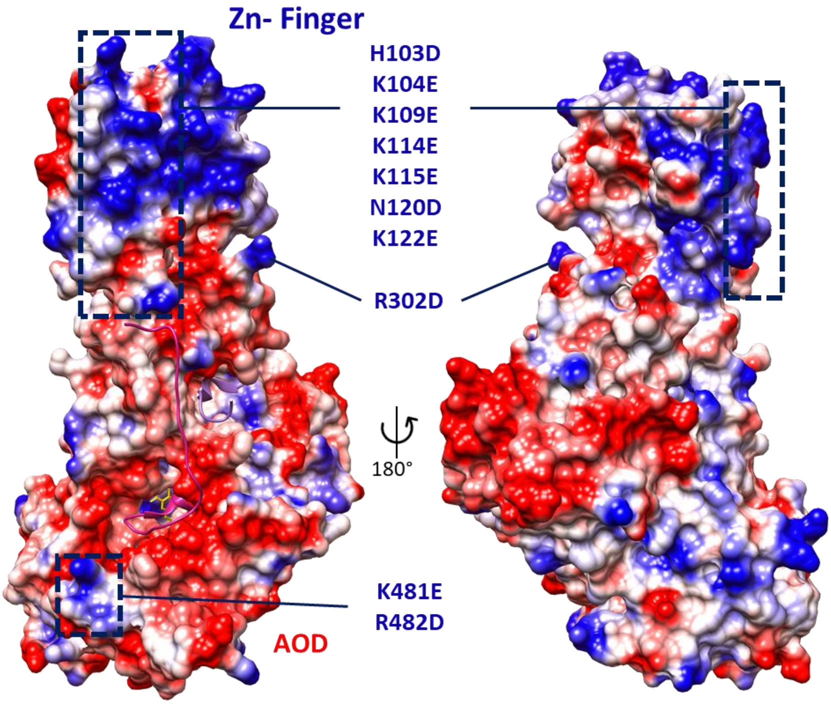


Figure 39. Overview of the selected target site for mutagenesis onto the LSD2 electrostatic surface, view from the substrate-binding site (left) and rotated by 180° (right). The ten mutagenized residues (whose alpha-carbons are evidenced as red spheres) are all exposed on the H3-tail binding side of LSD2 (left), mostly on the positively charged surface of the Zn-Finger domain. Red coloring refers to negatively charged atoms (-7 kcal/(mol*e)), whereas positively charged areas are coloured blue (7 kcal/(mol*e)). Various combinations of the mutations were tested: seven single point mutations, two double and two triple mutations, along with two deletions (evidenced in green).

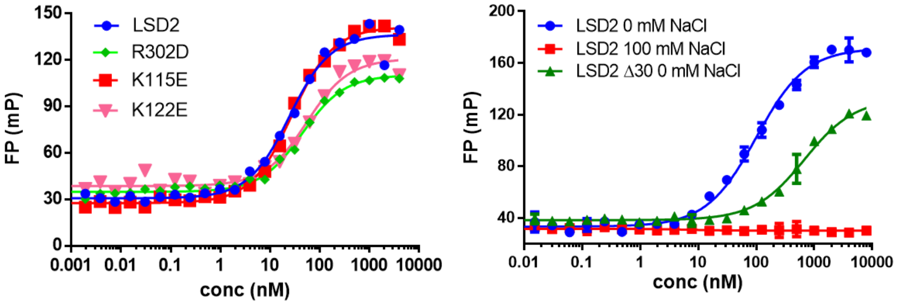


Figure 40. Fluorescence polarization DNA binding experiments. On the left, experimental curves are shown for few representative LSD2 mutants in comparison with LSD2 wild-type protein (blue). For a more detailed list of the resulting affinities for LSD2 mutants, please refer to the attached paper (Marabelli *et al.*, 2019). On the right, DNA-binding curves for the the Δ30 mutant in the standard assay condition and the wild-type protein at physiological salt concentration (100 mM NaCl) are shown in comparison to the wild-type LSD2 in the standard 0 mM NaCl assay condition (blue curve).

1.4 *LSD2 surface charges do not affect nucleosome recognition*

Analytical gel filtration chromatography of LSD2 proteins with nucleosomes was performed as described in the Materials and Methods section to dissect the mechanisms of NCP recognition of this enzyme. Briefly, each LSD2 mutant was incubated for one hour with NPAC-linker and alkylated nucleosomes. The mixture was then loaded onto Wyatt WTC-030S5 or WTC-030N5 columns and the elution profiles were compared to those of the wild-type, full-length protein (Figures 41, 42 and 43).

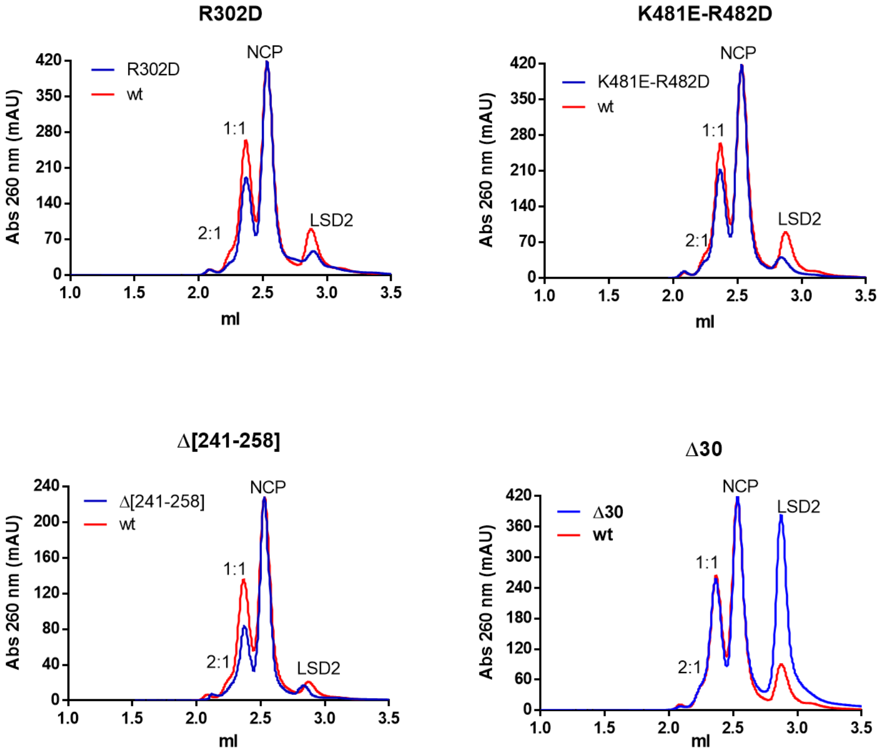


Figure 41. Effect of LSD2 mutations in protein domains other than the Zn-finger motif: Arg302 lies close to the H3-tail binding surface, Lys481 and Arg482 are on the Amino-oxidase domain, whereas deletions of residues 241-258 and 1-30 cover a flexible loop and the N-terminal tail of LSD2. For each graph, the comparison between the 260 nm absorbance elution profiles of the alkylated nucleosome complexes for LSD2fl wild-type (red) and each mutant (blue) is shown. The molar ratio of the mixture is 2: 1: 5 for LSD2fl : alkylated nucleosomes : NPAC-linker respectively.

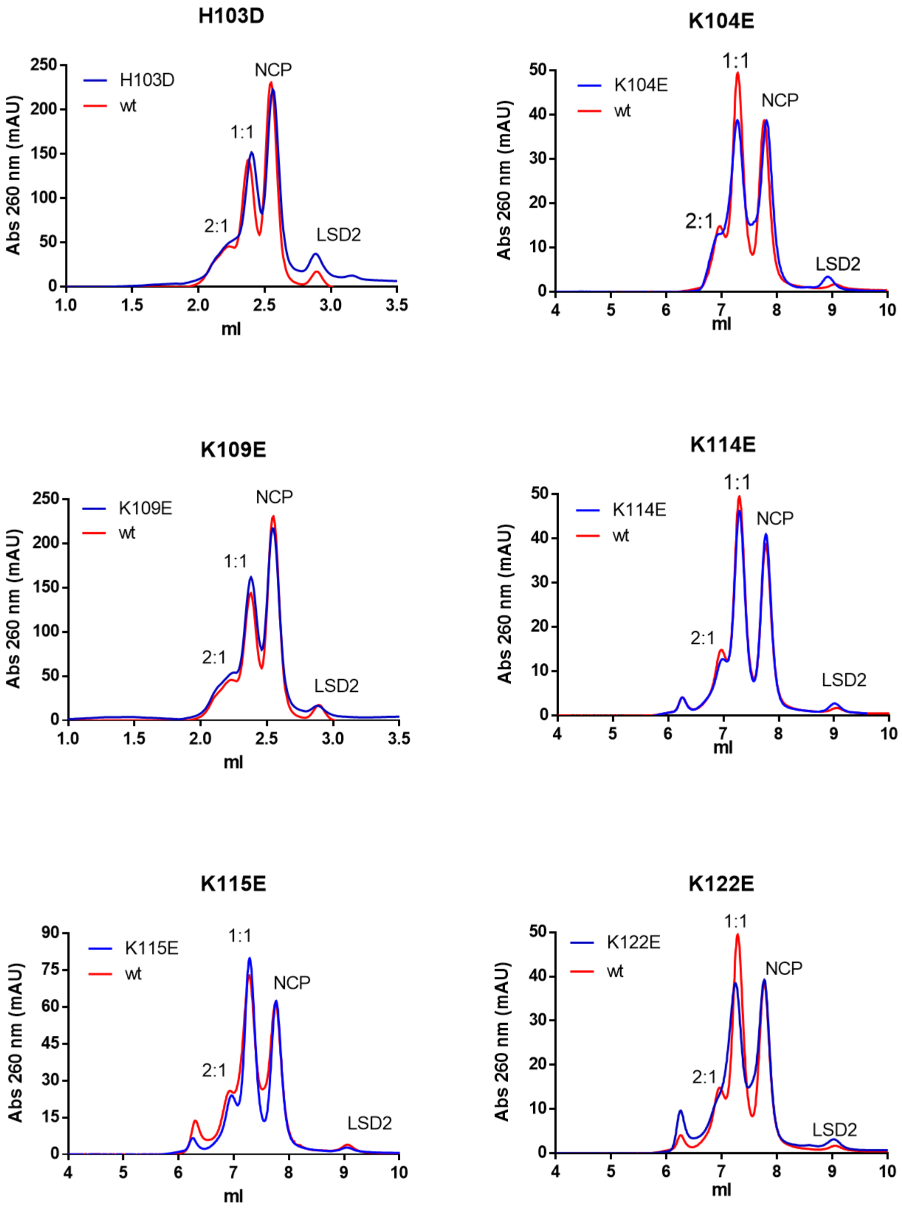


Figure 42. Effect of Zn-finger domain single point-mutations. **The 260 nm absorbance elution profiles of each experiment is compared to that with LSD2 wild-type. The molar ratio of the mixture is always 2: 1: 5 for LSD2fl : alkylated nucleosomes : NPAC-linker.**

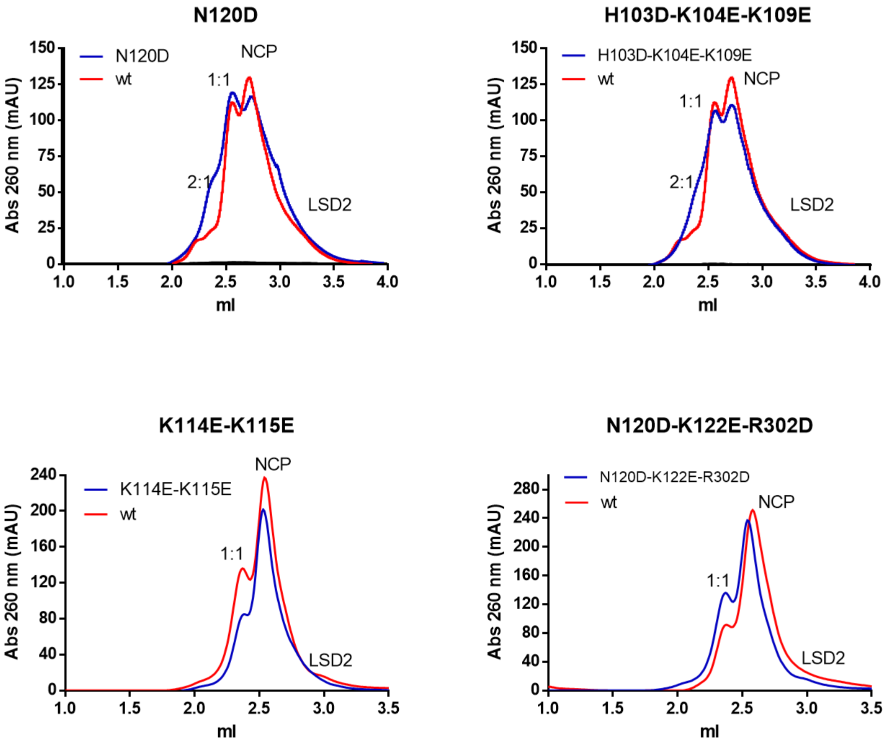


Figure 43. Effect of Zn-finger domain double and triple-mutations. For each graph, it is shown the comparison between the 260 nm absorbance elution profiles of the alkylated nucleosome complexes for LSD2fl wild-type and mutant. The molar ratio of the mixture is 2: 1: 5 for LSD2fl: alkylated nucleosomes: NPAC-linker respectively.

Differently from LSD1-CoREST system, for which single-point mutations on the DNA-binding surface of CoREST1 were sufficient to affect nucleosome recognition (Pilotto *et al.*, 2015), none of the tested mutants, either single-point mutations or double and triple mutations revealed any significant defect in nucleosome recognition (Figure 44). For comparison of the LSD2/nucleosome complex formation efficiencies, the ratio between the 260 nm absorbance value of the free nucleosomes peak to that of the 1:1 complex was calculated for each experiment, and then compared to that of the wild-type in the same conditions (incubated onto the same semi-synthetic nucleosomes sample and tested onto the same Wyatt SEC column on the same day, as described in the Materials and methods section).

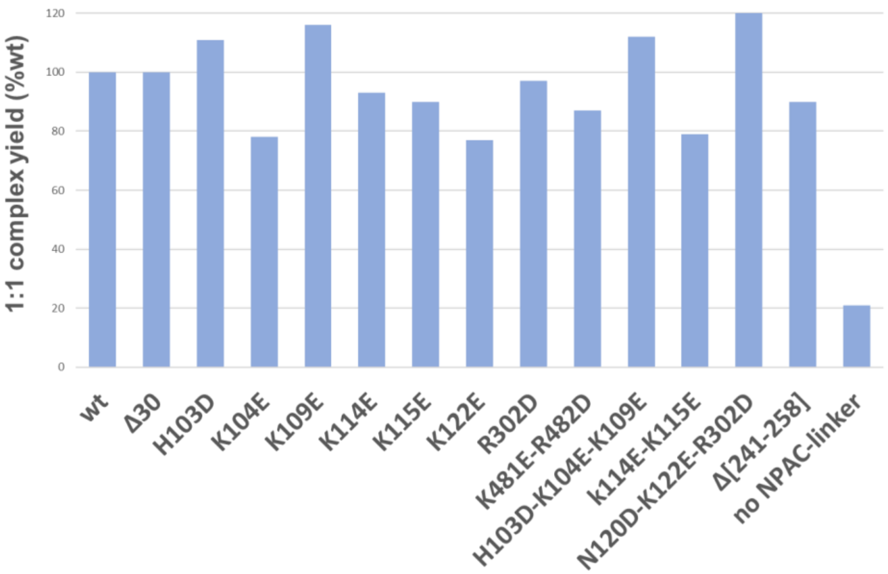


Figure 44. Bar-chart of the effect of surface LSD2 mutations on nucleosome complex formation efficiency as measured by analytical SEC experiments with semi-synthetic nucleosomes. Complex formation efficiency is calculated as the ratio between the 260 nm absorbance peak of the 1:1 complex eluting (refering to one LSD2/NPAC-linker to one alkylated nucleosome) to the 260 nm absorbance peak of free nucleosomes. The ratio is reported here as a percentage of the efficiency of the wild-type LSD2 with the same semi-synthetic nucleosome preparation, and tested on the same day.

1.5 NPAC linker is necessary for LSD2-nucleosome complex formation

As shown previously, all chromatographic experiments had been performed in presence of the synthetic 12-aa long linker peptide of NPAC, known to be essential for the demethylase activity on nucleosome substrate (Chen et al., 2013; Fang et al., 2013). As I tested the effect of NPAC-linker (12-aa long) on nucleosome recognition by wild-type LSD2, I discovered that it is essential for complex formation, as shown in Figure 45. The flavin absorption spectra of LSD2 after incubation with semi-synthetic nucleosomes with and without NPAC-linker were significantly different (Figure 26). The analytical SEC chromatograms of the LSD2/semi-synthetic nucleosome mixtures with and without NPAC-linker are also evidently different (Figure 45, right). If we consider the ratio of peak absorbance of the LSD1/NCP 1:1 complex to the one of unbound NCP as a measure of the yield of covalent complex, we can conclude that NPAC promotes NCP recognition by a factor 5 (Figure 44).

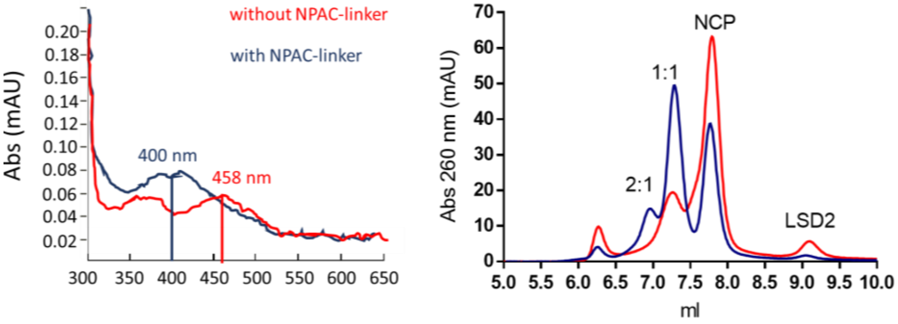


Figure 45. The effect of NPAC-linker on nucleosome binding by LSD2. Elution profiles of LSD2 and semi-synthetic nucleosome samples incubated with and without NPAC-linker.

NPAC-linker does not alter LSD2 abilities to bind either DNA oligos nor the H3 21-mer substrate peptide (Figure 38), and these data are also in accordance with previously published ITC experiments in which LSD2 affinity for H3 1-21 peptide (0.99±0.06 µM) increases only by a factor 1.5 in the presence of NPAC-linker (0.68±0.07 µM; Fang *et al*., 2013). Such a small increase in substrate recognition by LSD2 cannot account for the 5-fold NPAC-linker effect on nucleosome demethylation observed, and makes even more intriguing the NPAC-linker role in nucleosome recognition by LSD2.

To further exclude that the effect of NPAC-linker relies on the electrostatic interactions with either DNA or the positively charged H3 tail in the highly charged context of the nucleosome, complex formation efficiency was tested at varying ionic strengths, both in absence and presence of NPAC (Figure 46). The experiments definitely proved that the molecular interaction between LSD2 and its nucleosomal substrate, and the NPAC-linker effect as well, does not depend on electrostatics.

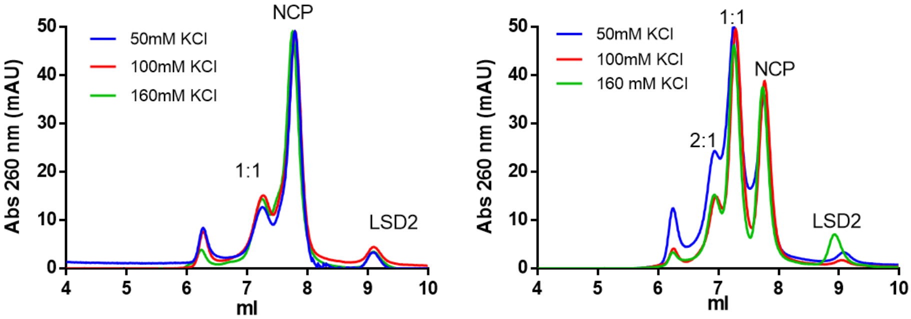


Figure 46. Comparison of LSD2/NPAC-linker/NCP complex formation efficiencies at low and high salt concentrations (see legends). The curves are standardized from the peak of free NCP. Incubation buffer was used as a running buffer too, which explains the different peak resolutions of the experiments.

Kinetic assays performed in our lab greatly helped in defining the role of the very short NPAC-linker in nucleosome recognition (see Table 1 in the attached paper: Marabelli *et al.*, 2019). Briefly, NPAC hydrophobic interaction with the H3 peptide is essential in minimizing the salt-dependence of positive substrate recognition by LSD2, and thus winning the competition for the H3 tail between the highly charged catalytic pocket of LSD2 and nucleosomal DNA.

2. Full length NPAC sustains LSD2 processivity

The multidomain protein NPAC hosts both chromatin-related modules as a PWWP and an AT-hook, together with a typically cytosolic dehydrogenase domain (Figure 1). We investigated the physiological function of NPAC at the single domain level, because of the impossibility to study a stable form of the full-length protein (data not shown). We kept a particular focus on its role within the LSD2 system and in the context of transcriptional elongation by RNA-Pol-II (Fang *et al.*, 2010; Fei *et al.*, 2018). Paragraphs 2.2 and 2.3 contain unpublished data, not presented in the attached paper (Marabelli *et al.*, 2019) and are therefore marked with an asterisk (*).

2.1 NPAC-PWWP unspecifically binds DNA and the H3 histone tail

The N-terminal PWWP domain of NPAC has been found at H3K36me3-rich chromatin loci, and in particular within actively transcribed regions (Vermeulen et al., 2010), but it is not able to specifically recognize H3K36me1/2/3-nucleosomes over unmodified ones (Sankaran et al., 2016). A more recent study on full-length NPAC also states no significant preferences were observed for nucleosomes carrying particular PTMs, and that it facilitates nucleosomes disassembly upon direct interaction (Fei et al., 2018). In my hands, NPAC-PWWP exhibited a very high affinity for DNA oligos, in the low nanomolar range. The ionic strength-dependency mirrors an electrostatic type of interaction (K_D at 0 mM NaCl is 0.4 nM, at 50 mM NaCl is 39 nM, at 100 mM NaCl is 676 nM whereas at 150 mM NaCl is almost 4 µM). NPAC-PWWP reveals thus strong DNA-binding properties, differently from LSD2 whose DNA-binding ability was significantly affected at physiological salt concentrations (Figure 47).

The methodology developed by Panagis Filippakopoulos and Sarah Picaud at the Structural Genomics Consortium in Oxford, was key to complete the scenario on the characteristics of the putative chromatin reader NPAC-PWWP. The SPOT-assay is a high-throughput analysis of the binding efficiencies between a proband protein and histone peptides with up to 700 different combinations of PTMs (for further details refer to paragraph 9 of the Materials and Methods section,). NPAC-PWWP exhibited no marked affinity for any histone PTM in particular. On the contrary, its binding pattern over a membrane blotted with H3 and H4 peptides revealed a “sticky” behavior with all histone fragments and PTMs combinations (Figure 48). In particular, NPAC-PWWP bound on the most positively charged patch, which on membrane corresponds to residues 40-45 of histone H3 (Figure 48 and Table S2 in Marabelli *et al.*, 2019).

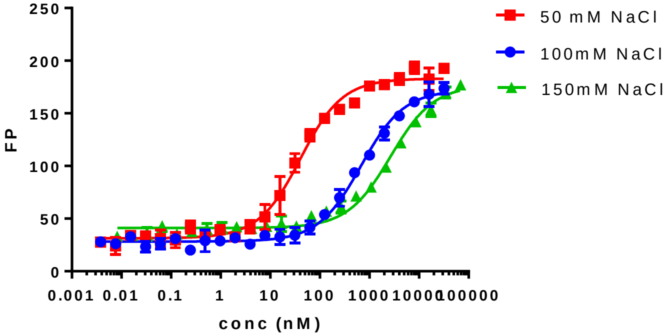


Figure 47. Fluorescence polarization curves of DNA-TAMRA oligo bound by NPAC-PWWP at varying salt concentration.

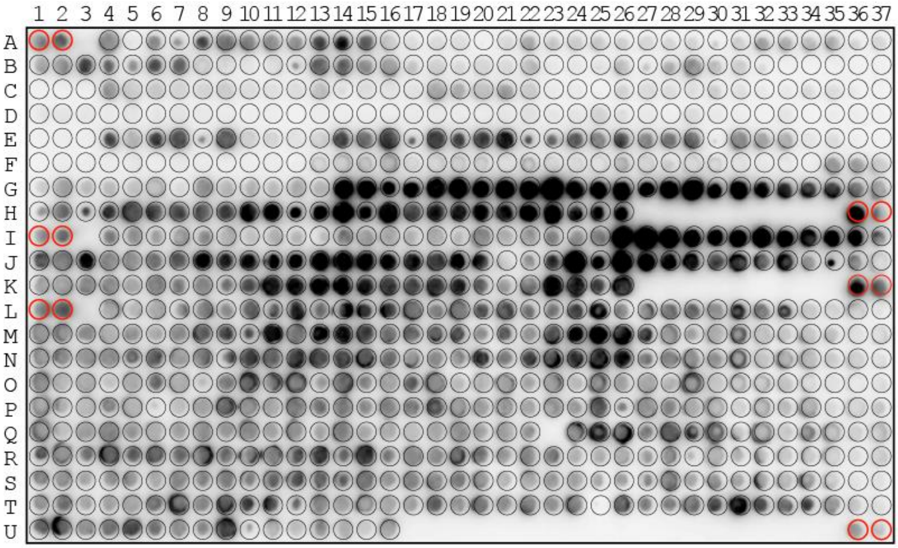


Figure 48. Peptide-SPOT assay on NPAC PWWP (residues 1-105) on H3.1, H3.3 and H4 peptides carrying various combinations of epigenetic moieties.

2.2 NPAC-PWWP binds to and destabilizes nucleosomes*

Electrophoretic mobility assay in native conditions (native-PAGE) revealed interestingly that nucleosomes are somehow destabilized and aggregate in presence of increasing concentrations of GFP-tagged NPAC-PWWP (Figure 49).

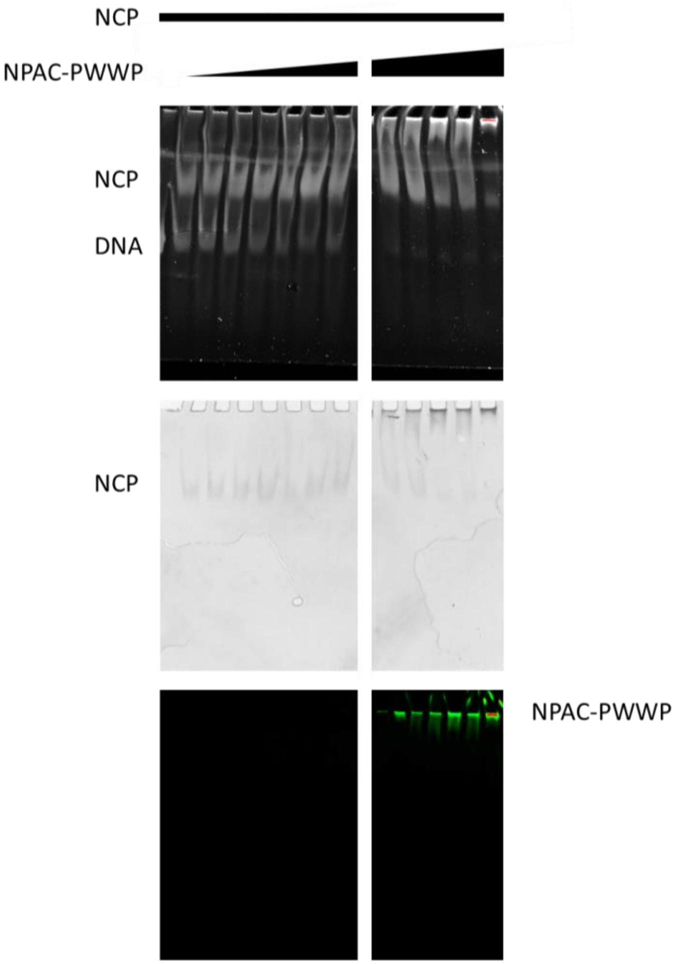


Figure 49. EMSA experiment on the effect of increasing concentrations of GFP-tagged NPAC-PWWP on chicken erythrocyte nucleosome. From top to bottom, the same gel was: stained for and imaged for SyBr-safe, whose fluorescence mirrors the presence of DNA; imaged for the presence of aromatic amino acids, which highlight the positions of proteins ad protein complexes, imaged for the intrinsic fluorescence of GFP, which colocalizes with the NPAC-PWWP protein.

Analytical SEC experiments on the same species, incubate for one hour on ice, also confirmed the PWWP-driven destabilization of nucleosomes, along with the fact that higher ionic strength values amplify this effect (Figure 50).

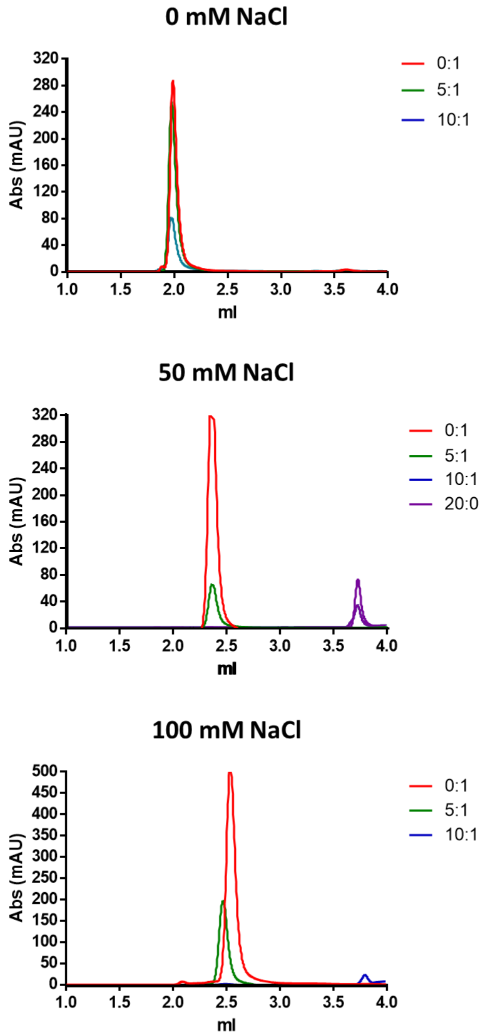


Figure 50. Elution profiles of mixtures containing NPAC-PWWP and chicken erythrocyte nucleosomes at varying molar ratio (0:1, 5:1, 10:1 and 20:1 respectively). The experiments were repeated at increasing NaCl concentrations (from top to bottom). Curves are referred to the 260 nm absorbances, except for the 10:1 and 20:0 NPAC-PWWP: chicken NCP experiments, for which the 260 nm curve absorbs too little and the 214 nm absorbance profile is showed instead.

However, none of the previous experiments showed the presence of any PWWP-NCP complex. To determine the presence of such a species, milligram-scale quantities of GFP-NPAC-PWWP were incubated for one hour with chicken erythrocyte nucleosomes at 2:1 stoichiometric ratio, in 100 mM NaCl final concentration buffer. The sample was then loaded onto a cation exchange Capto-DEAE column (Figure 51). The cationic exchange Capto-DEAE allows separation of the histone/DNA species according to their surface charge: therefore histones, tetrasomes, hexasomes, nucleosomes and free DNA can be separated. Increase or decrease of any of the detected species because of the presence of the NPAC-PWWP can be monitored through comparison of the NPAC-PWWP/nucleosome mixture elution profikle to that of GFP-NPAC-PWWP and chicken erythrocyte nucleosomes alone (Figure 51) IN presence of the NPAC-PWWP a new species appeared, eluting at lower ionic strength than nucleosomes, but at a higher ionic strength than that of free histones.

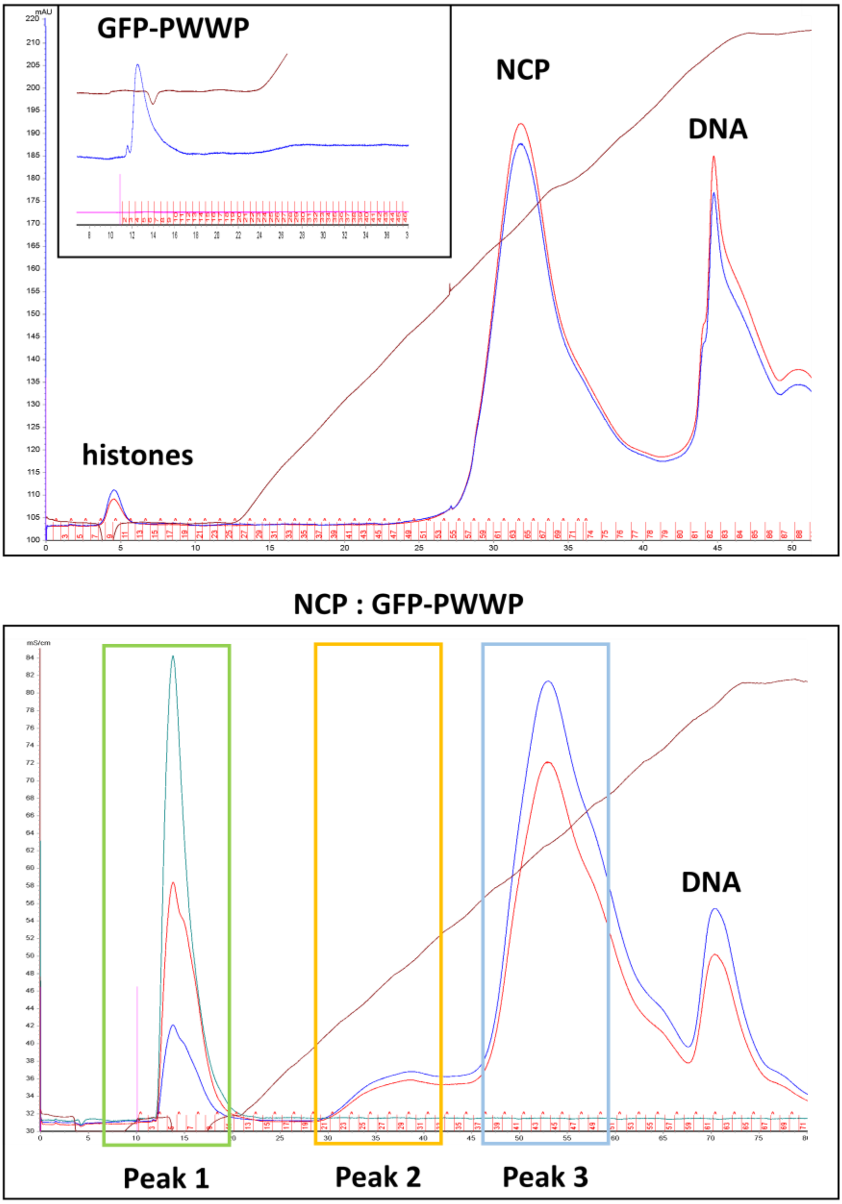


Figure 51. Capto-DEAE elution profile of the chicken erythrocyte NCPs incubated with the GFP-tagged NPAC-PWWP. The stoichiometric ratio is 1: 3 for the NCP and the PWWP species respectively. For comparison, the above graphs show the elution profiles of the single species alone; GFP-tagged NPAC-PWWP elutes in the flow through, as histone proteins.

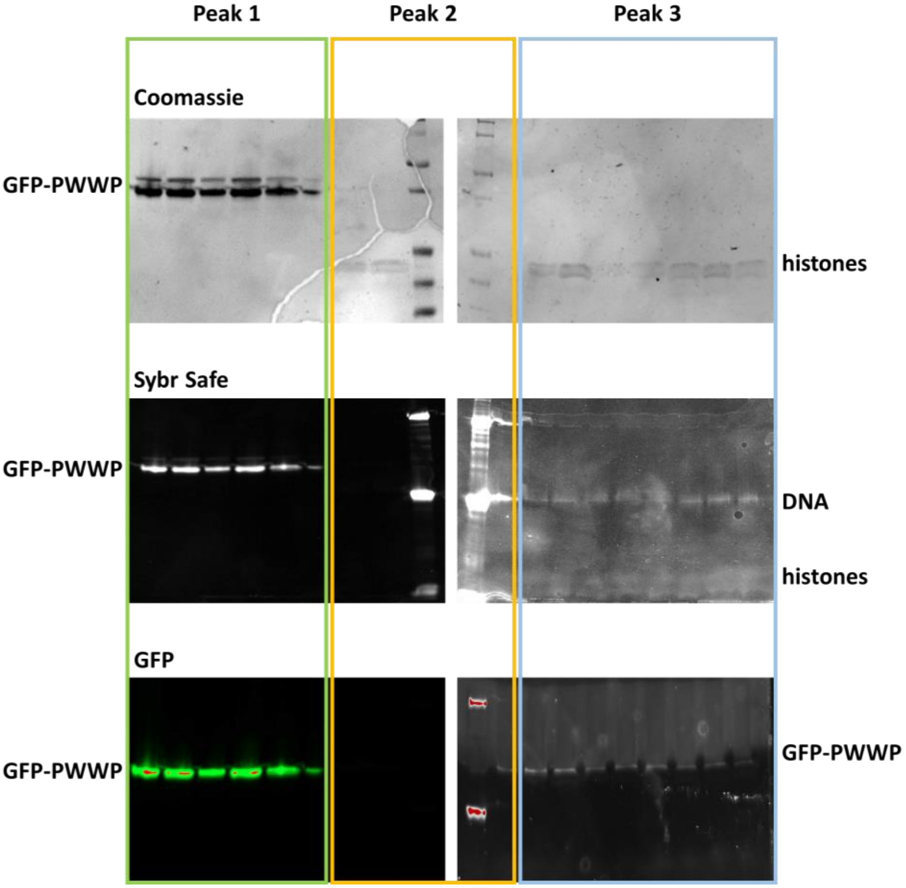


Figure 52. SDS-PAGE analysis of the fractions elution from the captoDEAE showed in Figure 51. The identity of the peak for each group of samples is highlighted by coloured boxes. From top to bottom, the same gel is stained with and imaged for Coomassie blue, for protein detection; Sybr Safe, whose fluorescence mirrors the presence of DNA; intrinsic fluorescence of GFP, which colocalizes with the NPAC-PWWP protein.

SDS-PAGE analysis of the eluting peaks confirmed, as expected, that GFP-NPAC-PWWP mostly elutes in the flow-though, whereas the nucleosome and DNA species eluted at high ionic strength (Figure 52). Peak 2 contained histones but not detectable quantities of nucleosomal DNA, which is in line with the fact that elution at lower ionic strength than nucleosomes imply a lower negative charge of the species. The most interesting, and unexpected, signal was the GFP fluorescence of NPAC-PWWP within peak 2 and 3 fractions (Figure 51).

2.3* NPAC-PWWP effect on nucleosomes is not hampered by the DH domain

Due to the intrinsic instability of NPAC full-length, I tested the effect of the NPAC-PWWP on nucleosomes in presence of the LSD2-binding linker and the DH domain. I incubated the NPAC Δ99-179 protein (containing the PWWP, the LSD2-binding linker and the dehydrogenase domain, see Figure 18) with increasing concentration of chicken erythrocyte nucleosomes, to test for the nucleosome destabilization effect by a longer construct of NPAC than the already tested PWWP domain (Figure 49). It appears that the presence of the LSD2-binding linker and the dehydrogenase domain do not affect nucleosome destabilization by NPAC-PWWP (Figure 53).

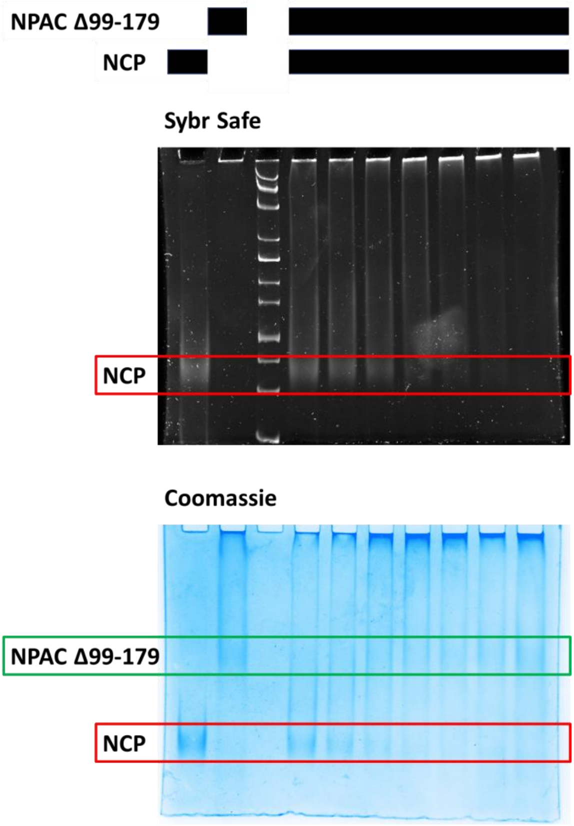


Figure 53. EMSA assay on nucleosome destabilization by NPAC Δ99-179. Chicken erythrocyte nucleosomes were incubated for one hour on ice with increasing concentrations of NPAC Δ99-179. Both Coomassie Blue and Sybr Safe stains reveal the fading of the nucleosomes (red box) band with increasing NPAC Δ99-179 (green box).

2.4 NPAC-DH is a NADPH/NADP$^+$-binding tetramer

Based on structural homology, the C-terminus of NPAC belongs to the β-hydroxyacid dehydrogenase (β -HAD) family (residues 261-553, from now on named NPAC-DH). NPAC-DH crystal structure has been already resolved at 2.5 Å of resolution by the SGC Toronto (PDB 2UYY, Figure 54). Similarly to the other β-HAD family members, NPAC-DH is a tetramer, where each subunit is essentially composed of two globular domains connected by a long α-helix (Figure 54).

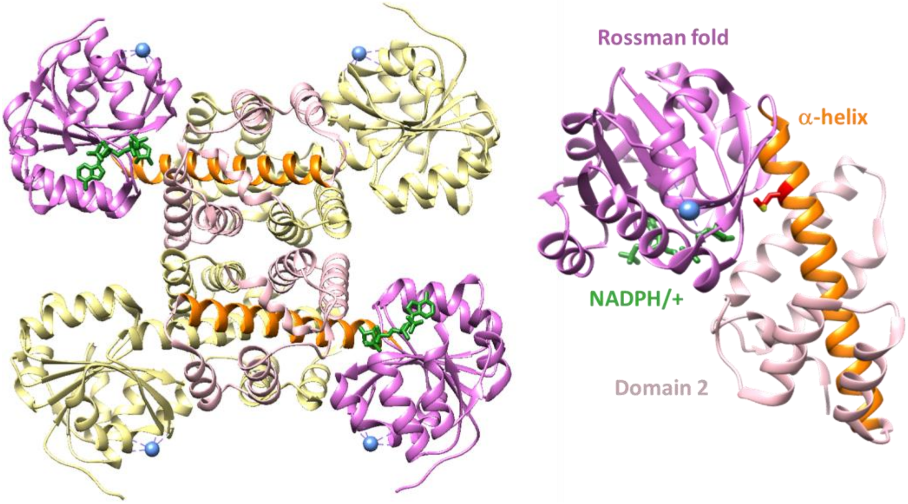


Figure 54. NPAC-DH crystallographic tetramer (on the left) and monomer (on the right, PD code 2UYY). Colors are referring to the distinct regions of the monomer NPAC: the Rossman fold (purple), the bound NADPH/+ cofactor (green), the long interconnecting alpha-helix (orange) and the second globular domain (pink).

One of the two domains stabilizes the quaternary structure of the dehydrogenase, whereas the second is a Rossman fold, which typically binds nucleotides such as ATP or NAD$^+$ cofactors. Evidences revealed that cofactor binding also has a stabilizing effect on the typically tetrameric structure of β-HAD enzymes (Njau, Herndon and Hawes, 2001). Differently from similar dehydrogenases, NPAC-DH crystal structure shows a low-resolution map for a NADPH-like molecule. Missing electrondensity for the nicotinamide ring and the impossibility to stabilize the oxidative state of the cofactor might imply weak cofactor binding. However, the experiment was reported to be performed in co-crystallization with the NADH cofactor, and there are no clues about the origin the bound molecule.

To clarify the identity of the cofactor bound to the unpublished structure (PDB 2UYY) of NPAC dehydrogenase domain, I moved to reproduce the crystals of NPAC-DH. The protein was purified from *E.coli* cells at room temperature as described in the Materials and methods section. SDS-PAGE confirmed the identity of the NPAC-DH with its expected MW of 31 kDa, but SEC analysis showed a unique monodispersed species with a molecular mass of 120 kDa, corresponding to that of the NPAC-DH tetramer. Also, UV-Vis absorption spectra of the purified construct revealed a peak absorbing at 348 nm, indicating that a nicotinamide-containing cofactor, either NADH or NADPH is bound (Figure 55).

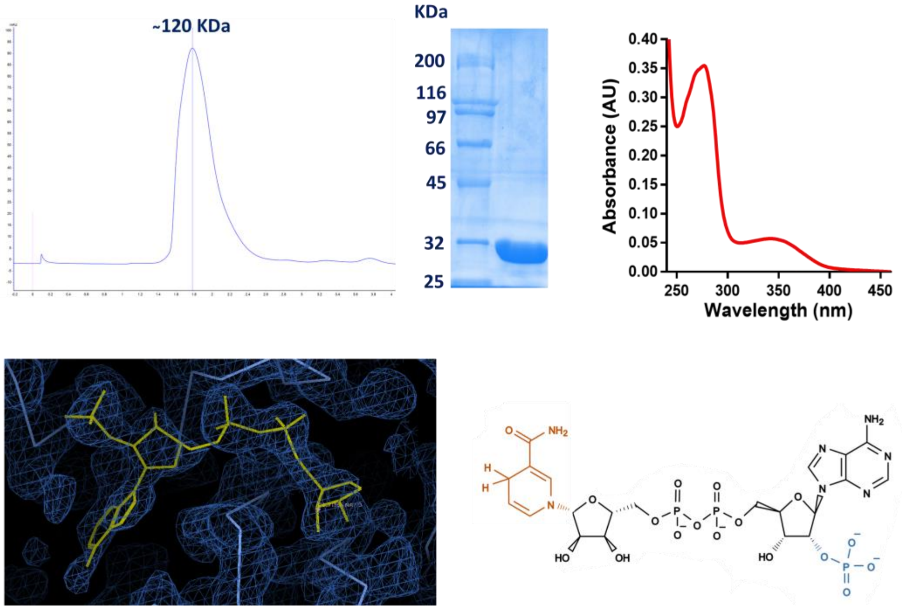


Figure 55. Top left) Pure NPAC-DH behaves as an oligomer of 120 kDa in solution, as revealed by analytical chromatography. The SDS-PAGE gel shows the band of the sample used, which corresponds to the monomeric NPAC-DH (31 kDa). NPAC-DH in solution is then a tetramer (31 kDa x 4 = 124 kDa). On the top right) Spectra of the purified NPAC-DH. The absorbance peak of a nicotinamide cofactor is visible at 348 nm. On the bottom left) electrondensity of the cofactor bound within the NPAC-DH structure on PDB site (code 2UYY). The nicotinamide moiety is not present, as it can be seen from the comparison with the molecular structure of the NADPH cofactor on the bottom right.

I tried to identify the proband substrate both through thermal denaturation experiments and spectrophotometric activity assays with no results. To clarify the identity of NPAC cofactor, I then moved to reproduce the crystals of NPAC-DH, both with and without the addition of possible cofactors (see Material and Methods). I obtained an electron-density map at 3.6 Å for the native protein, for which the refined structure perfectly fitted with the map from SGC Toronto (PDB code 2UYY). The unit cell hosts four monomers, interacting with each other through hydrophobic interactions. Two opposite subunits of the native NPAC tetramer bind $NADP^{+-}$ molecule, whose electron-density for the nicotinamide moiety is still missing (Figure 55).

2.5 NPAC-DH is an inactive enzyme

During the analysis of the solved structure, we noticed that a conserved lysine, crucial for the catalysis in all β-HADs (Hoover *et al.*, 2013; Zhang *et al.*, 2014), is mutated to methionine. The mutation lies on the long α-helix connecting the domains 1 and 2 (Figure 54). This helix is very conserved among β-HADs. It always positions exactly in front of the cofactor where it shapes the substrate-binding pocket. In the crystal structure of NPAC the cleft in between the $NADP^+$/NADPH cofactor and the helix appears very narrow because of the steric hindrance of the long protruding amino acid side-chains.

Intriguingly, bioinformatic analysis revealed that the catalytic lysine mutated during evolution since the appearance of a chromatin-related module within NPAC, being it either the PWWP, the AT-hook, or the LSD2-binding linker. Plant sequences in plants (*A.thaliana*, *S. microadriaticum*, *A. Amnicola*), which are missing the NLS and all nuclear domains conserve the catalytic lysine and their cytoplasmic activity on glyoxylate has been demonstrated (Hoover *et al.*, 2007, 2013; Simpson *et al.*, 2008; Allan *et al.*, 2009). The insect NPAC orthologues (*A. glabripennis*, *B. dorsalis*, *C. capitata*) carry a lysine-to-asparagine mutation but not the LSD2-binding linker and their role has not been investigated yet (Figure 56).

To better understand the physiological role of a nuclear inactive dehydrogenase, we characterized the effect of the residue 437 mutations observed among NPAC homologs. The purification and study of the two NPAC-DH mutants M437K and M437N was performed in our lab, with very interesting results, explained in the paper here attached (Marabelli *et al.*, 2019). The presence of a methionine instead of a lysine, at that particular position on the α-helix between the two globular domains, strongly stabilizes the tetrameric structure of NPAC, independently on the presence of the cofactor (Marabelli *et al.*, 2019).

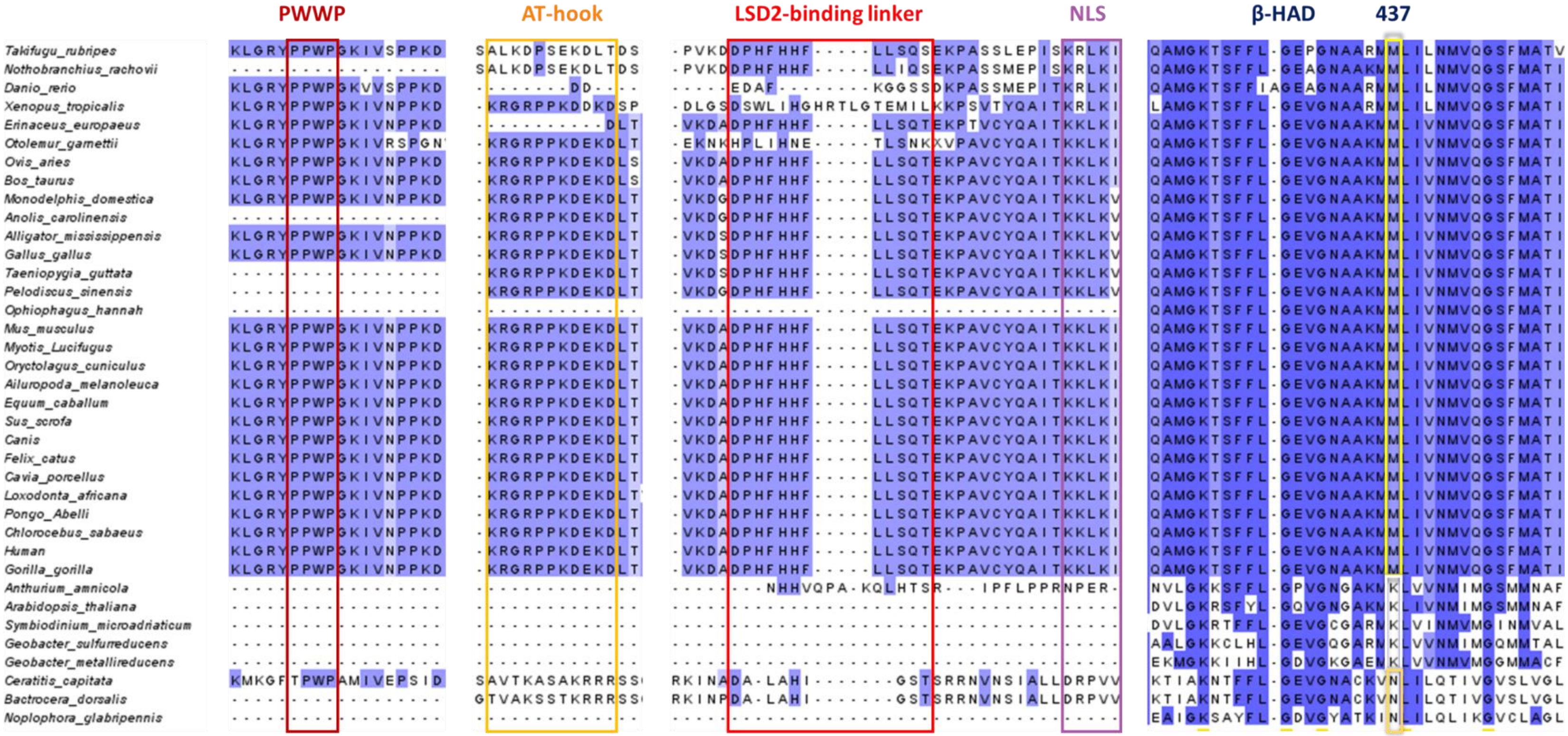


Figure 56. Multiple Sequence Alignment of NPAC orthologs from different species (named on the left column) with conserved residues evidence in blue through Jalview software. Residues are coloured with different grades of blue depending on their percentage identity scores (the darker the higher). Portions of the sequence alignment are shown, corresponding to the characteristic features of the PWWP domain(dark red), the AT-hook (yellow), the LSD2-binding linker (bright red) with its nearby nuclear localization signal (NLS, violet) and the portion of the dehydrogenase domain containing the lysine-to-methionine mutation at position 437 (methionine mutations are evidenced in yellow, arginine mutations typical of insect homologous proteins are evidenced in gold).

Taken together, these observations suggest that the evolution of NPAC-DH from a cytosolic enzyme, as it is in *A. thaliana* (Hoover *et al.*, 2007; Simpson *et al.*, 2008; Allan *et al.*, 2009) to a nuclear factor has surprisingly been driven by a hyper-stabilizing single-point mutation.

It is also interesting to observe that within the DH domain the most conserved residues lie at the interfaces between α-helices, and in particular in the inter-subunit interface of the crystallographic tetramer (Figure 57). The weaker selective pressure on specific cofactor-binding residues also is evident, which seems to confirm that NPAC has been selected to be a NADPH-independent stable tetramer.

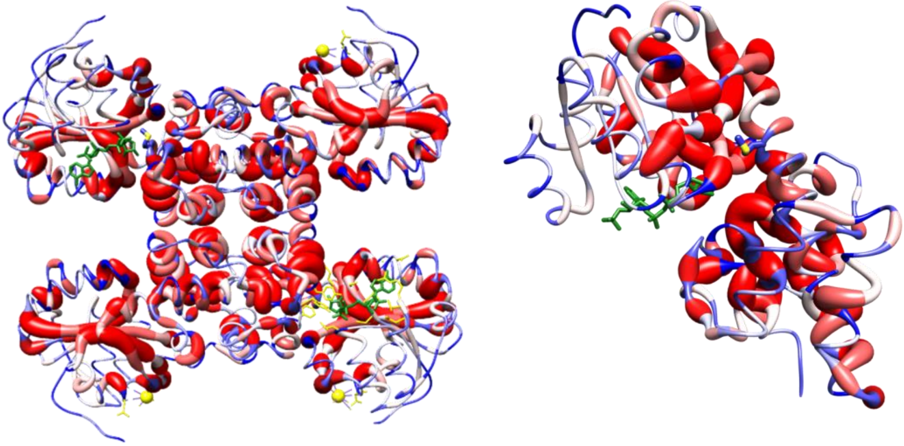


Figure 57. NPAC-DH crystallographic tetramer (left) and monomer (right) represented with round ribbons, with increasing thickness depending on the conservation of the amino acid, and depicted in gradient of color (from blue to red) to further evidence the most conserved residues (red). The NADPH binding pocket is the least conserved element of the NPAC-DH structure.

Dead enzymes are usually thought to be evolutionary relics, whose role is limited to their ability to sequester a particular substrate molecule, and in turn affect the catalysis of active homologs. NPAC is not only confined to a completely different cellular compartment than its counterparts, but it also assumed a new role, with seemingly no direct correlation with the metabolic pathways of cytosolic dehydrogenases: it is an allosteric modulator of nucleosome demethylation by LSD2.

The NPAC-DH evolution is a case study in the analysis of functional innovation in enzymes and proteins (Tawfik, 2014; Jackson *et al.*, 2015). In this perspective, the interesting dynamic effect of the NPAC-DH lysine-to-methionine single-point mutation on its quaternary structure has been beautifully investigated in collaboration with Professor Giorgio Colombo of the University of Pavia. It appeared that the single lysine-to-methionine mutation, in the middle of the inter-domain α-helix, destabilizes the interaction between the two globular domains of NPAC-DH (Figure 50). This in turn affects the quaternay structure of the pseudo-dehydrogenase. Results have been published in the paper here attached (Montefiori *et al*., 2019).

2.6 The NPAC tetramer hosts multiple LSD2

To investigate the function of NPAC as oligomerization centre, NPAC proteins containing both the LSD2-binding linker and the dehydrogenase domain were cloned, expressed and purified as described in paragraphs 4 of the Materials and Methods section. These NPAC proteins were incubated with LSD2 and semi-synthetic nucleosomes at different molar ratios and complex species formation were analysed through measurement of the absorption spectrum and analytical-SEC experiments. NPAC-DH does not inhibit the ability of NPAC-linker to bind LSD2 and promote its activity towards the nucleosome, because the 400 nm characteristic absorption spectrum of the covalently bound flavin was observed after incubation of LSD2 with both nucleosomes and either NPAC Δ205 or Δ99-179, but not after incubation of LSD2 and the nucleosome alone (data not shown).

Even more interesting, we found that varying concentrations of NPAC Δ205 lead to formation of different complex species with very high yet distinct molecular weights (Figure 58). The impossibility to separate the peaks allowed us just a merely qualitative evaluation of the species. Based on the ratio between the absorbance measured at 214 nm (proportional to peptide bonds) and 260 nm (for the DNA content), we speculate that the peak at 2.2 ml corresponds to a complex composed by one tetramer of NPAC, one LSD2 and one nucleosome, whereas the peak at 1.98 ml, with a lower protein content, is compatible with one NPAC tetramer, two LSD2 copies and two nucleosomes in complex.

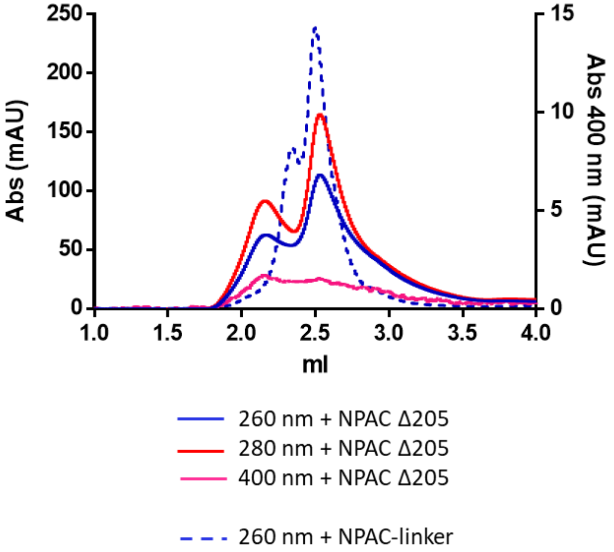


Figure 58. Analytical SEC analysis on LSD2/nucleosome complex formation in presence of NPAC Δ205. Complex species are monitored by the 400 nm absorbance (right Y axis), and elute at significantly smaller volumes than the LSD2/nucleosome/NPAC-linker complexes. The experiments were performed with identical quantities of LSD2, nucleosomes and NPAC-linker or NPAC Δ205.

Conclusions

1. A tail-based mechanism for nucleosome recognition by LSD2

It is evident from the presented data that, differently from its homologous LSD1, the DNA-binding capacity of LSD2 is not the main player in nucleosome recognition. The very short NPAC-linker also does not have an influence on nucleosomal DNA-binding properties, neither it significantly affects recognition of the substrate H3 peptide, as revealed by fluorescence polarization experiments. The NPAC-linker 5-fold enhancement of complex formation might thus be due to its positive effect on the acquisition by the H3 tail in the catalytically competent orientation and position. Indeed, similarly to LSD1, the substrate-binding cleft of LSD2 recognizes the first 16 amino acids through electrostatic interactions, whereas the H3 residues exposed to the solvent (in particular at positions 16-21, as shown in Marabelli et al., 2019) are oriented through salt-sensitive, hydrophobic interactions. In the highly-charged nucleosome context, it appears evident that the stabilizing effect of the NPAC-linker on this specific portion of the substrate peptide might greatly influence its orientation and proper stabilization within the active pocket.

Surface mutagenesis experiments with semi-synthetic nucleosomes further revealed that none of the surface exposed LSD2 patches is specifically involved in nucleosome approach and/or docking. Cryo-EM experiments confirmed that LSD2 can indeed approach its substrate with multiple orientations, all seemingly equally competent, as the LSD2/nucleosome structures all were obtained through site-specific crosslinking, requiring formation of a catalytically reactive encounter complex. It appears then that only the extensive and high-affinity interaction between the demethylase and the full 40-aminoacids long H3 N-terminus is sufficient to trigger complex formation and catalysis. In this context, the DNA-binding capacity of LSD2 might facilitate formation of the encounter complex, which in turn facilitates the capture of the histone tail by the demethylase. The key aspect of the cryo-EM experiments is that it allowed in silico isolation and analysis of different complex conformations, which would have been unfeasible with other techniques. Moreover, it is the first time that a nucleosome modifier is found not to stable dock onto the nucleosome through recognition of specific structural/chemical decorations (Zhou and Luger, 2018).

2. NPAC regulates nucleosome processing by multiple chromatin modifiers

Given the strong positive effect of NPAC-linker on LSD2 demethylation activity of nucleosome substrates, it was intriguing to discover that other domains of the LSD2 physiological partner also influence the characteristics of H3 demethylation within the same context.

The dehydrogenase domain of NPAC, at the C-terminal side of the LSD2-activating peptide, was selected during evolution for an inactivating mutation and became a tetramerization module. SAXS analysis of LSD2/NPAC stable complex indicated a complex stoichiometry between the two, as they form huge oligomers in presence of nucleosomes. The N-terminal PWWP chromatin-binding domain of NPAC probably has also an effect on the demethylation reaction, probably through stabilization of a more "open" state of the nucleosome substrate, which increases both H3 tail and DNA availability.

Taking into account the context in which the LSD2/NPAC system operates, which is during gene trasnscription elongation by RNA Pol-II, the effect of both N- and C-terminal domains of NPAC appear even more intriguing. The stable NPAC tetramer, determined by the dehydrogenase domain, holds multiple copies of active demethylases, thus increasing the avidity and processivity of the system. The fact that such a stable core has been selected by evolutive pressure in eukaryotes further sheds light on the importance of this oligomerization module, whose flexible linkers might function as a docking platform for many chromatin factors other than LSD2 and p38. Moreover, the nucleosome destabilizing effect of the NPAC-PWWP also might favour chromatin processing in actively transcribed gene bodies, were DNA needs to be more accessible by RNA Pol-II.

It is inevitable to do a comparison between the mechanisms of nucleosome recognition by the two homologous demethylases LSD1 and LSD2. Despite sharing the same catalytic properties, recognition of the same substrate occurs through different mechanisms, both of which involving a protein partner. DNA recognition by CoREST1 allows stable clamping of the nucleosome and consequent silencing of gene promoters, whereas LSD2/NPAC-linker only rapidly and flexibly interacts with the nucleosome core, of which the only recognized feature is the protruding substrate H3 tail. These opposite strategies for demethylation of the same residue H3K4me1/2 might reflect the diverse needs of the accompanying machineries and/or the epigenetic and accessibility states of the substrate nucleosomes within the two contexts: silencing of gene promoter versus actively transcribed gene bodies.

References

Adrian, M. *et al.* (1984) ‘Cryo-electron microscopy of viruses’, *Nature*, 308(5954), pp. 32–36. doi: 10.1038/308032a0.

Allan, W. L. *et al.* (2009) ‘Role of plant glyoxylate reductases during stress: a hypothesis’, *Biochemical Journal*. doi: 10.1042/BJ20090826.

Altschul, S. F. *et al.* (1990) ‘Basic local alignment search tool’, *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(05)80360-2.

Bateman, A. *et al.* (2017) ‘UniProt: The universal protein knowledgebase’, *Nucleic Acids Research*, 45(D1), pp. D158–D169. doi: 10.1093/nar/gkw1099.

Berman, H. M. *et al.* (2000) *The Protein Data Bank*, *Nucleic Acids Research*. Available at: http://www.rcsb.org/pdb/status.html.

Binda *et al.* (2010) Biochemical, Structural, and Biological Evaluation of Tranylcypromine Derivatives as Inhibitors of Histone Demethylases LSD1 and LSD2, :*J. Am. Chem. Soc. Vol. 13 (219), 6827-6833.*

Chen, F. *et al.* (2013) ‘Structural insight into substrate recognition by histone demethylase LSD2/KDM1b’, *Cell Research*. doi: 10.1038/cr.2013.17.

Dinkel, H. *et al.* (2016) ‘ELM 2016 - Data update and new functionality of the eukaryotic linear motif resource’, *Nucleic Acids Research*, 44(D1), pp. D294–D300. doi: 10.1093/nar/gkv1291.

Dyer, P. N. *et al.* (2004) *[2] Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA*.

Fang, R. *et al.* (2010) ‘Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 Methylation’, *Molecular Cell*, 39(2), pp. 222–233. doi: 10.1016/j.molcel.2010.07.008.

Fang, R. *et al.* (2013) ‘LSD2/KDM1B and Its Cofactor NPAC/GLYR1 Endow a Structural and Molecular Model for Regulation of H3K4 Demethylation’, *Molecular Cell*, 49(3), pp. 558–570. doi: 10.1016/j.molcel.2012.11.019.

Fei, J. *et al.* (2018) ‘NDF, a nucleosome-destabilizing factor that facilitates transcription through nucleosomes’, *Genes and Development*, 32, pp. 1–13. doi: 10.1101/gad.313973.118.

Fonfría-Subirós, E. *et al.* (2012) ‘Crystal structure of a complex of DNA with one AT-hook of HMGA1’, *PLoS ONE*. doi: 10.1371/journal.pone.0037120.

Frank, J. (2010) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford University Press. doi: 10.1093/acprof:oso/9780195182187.001.0001.

Fu, J. *et al.* (2006) ‘Nuclear protein NP60 regulates p38 MAPK activity’, *Journal of Cell Science*, 119, pp. 115–123. doi: 10.1242/jcs.02699.

Hoover, G. J. *et al.* (2007) ‘Characteristics of an *Arabidopsis* glyoxylate reductase: general biochemical properties and substrate specificity for the recombinant protein, and developmental expression and implications for glyoxylate and succinic semialdehyde metabolism in planta’, *Canadian Journal of Botany*. doi: 10.1139/B07-081.

Hoover, G. J. *et al.* (2013) ‘Identification of catalytically important amino acid residues for enzymatic reduction of glyoxylate in plants’, *Biochimica et Biophysica Acta - Proteins and Proteomics*. doi: 10.1016/j.bbapap.2013.09.013.

Jackson, B. C. *et al.* (2015) ‘Dead enzymes in the aldehyde dehydrogenase gene family: role in drug metabolism and toxicology’, *Expert Opinion on Drug Metabolism & Toxicology*. doi: 10.1517/17425255.2016.1108406.

Jacobson, M. P. *et al.* (2014) ‘Leveraging structure for enzyme function prediction: Methods, opportunities, and challenges’, *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2014.05.006.

Karplus, P. A. and Diederichs, K. (2012) ‘Linking crystallographic model and data quality’, *Science*. American Association for the Advancement of Science, 336(6084), pp. 1030–1033. doi: 10.1126/science.1218231.

Klinker, H. *et al.* (2014) ‘Rapid purification of recombinant histones’, *PLoS ONE*, 9(8). doi: 10.1371/journal.pone.0104029.

Kumar, M., Gromiha, M. M. and Raghava, G. P. S. (2007) 'Identification of DNA-binding proteins using support vector machines and evolutionary profiles', *BMC Bioinformatics*. doi: 10.1186/1471-2105-8-463.

Li, W. *et al.* (2015) 'The EMBL-EBI bioinformatics web and programmatic tools framework', *Nucleic Acids Research*, 43(W1), pp. W580–W584. doi: 10.1093/nar/gkv279.

Luger, K., Rechsteiner, T. J. and Richmond, T. J. (1999) 'Expression and Purification of Recombinant Histones and Nucleosome Reconstitution', in *Chromatin Protocols*. doi: 10.1385/1-59259-681-9:1.

Marabelli, C. *et al.* (2019) 'A Tail-Based Mechanism Drives Nucleosome Demethylation by the LSD2/NPAC Multimeric Complex', *Cell Reports*, 27, pp. 387–399. doi: 10.1016/j.celrep.2019.03.061.

Njau, R. K., Herndon, C. A. and Hawes, J. W. (2001) 'New developments in our understanding of the β-hydroxyacid dehydrogenases', *Chemico-Biological Interactions*, 130–132, pp. 785–791. doi: 10.1016/S0009-2797(00)00234-9.

Picaud, S. and Filippakopoulos, P. (2015) 'SPOTing Acetyl-Lysine Dependent Interactions', 4, pp. 370–388. doi: 10.3390/microarrays4030370.

Pichlo, C. *et al.* (2016) 'Production, Crystallization and Structure Determination of C. difficile PPEP-1 via Microseeding and Zinc-SAD', *Journal of Visualized Experiments*. doi: 10.3791/55022.

Pilotto, S. *et al.* (2015) 'Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation', *Proceedings of the National Academy of Sciences*, 112(9), pp. 2752–2757. doi: 10.1155/2014/349065.

Qin, S. and Min, J. (2014) 'Structure and function of the nucleosome-binding PWWP domain', *Trends in Biochemical Sciences*. Elsevier Ltd, 39(11), pp. 536–547. doi: 10.1016/j.tibs.2014.09.001.

Ramakrishnan, V. (2018) *Gene machine : the race to decipher the secrets of the ribosome*.

Reeves, R. (2001) 'Molecular biology of HMGA proteins: hubs of nuclear function.', *Gene*, 277(1–2), pp. 63–81. doi: 10.1016/s0378-1119(01)00689-8.

Rohou, A. and Grigorieff, N. (2015) 'CTFFIND4: Fast and accurate defocus estimation from electron micrographs', *Journal of Structural Biology*, 192(2), pp. 216–221. doi: 10.1016/j.jsb.2015.08.008.

Sankaran, S. M. *et al.* (2016) 'A PWWP Domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated lys-36 of histone H3 and regulates NSD2 function at chromatin', *Journal of Biological Chemistry*, 291(16), pp. 8465–8474. doi: 10.1074/jbc.M116.720748.

Savitsky, P, et al. (2016). Multivalent Histone and DNA Engagement by a PHD/BRD/PWWP Triple Reader Cassette Recruits ZMYND8 to K14ac-Rich Chromatin, *Cell Rep. 2016 Dec 6; 17(10): 2724–2737.*

Savva, C. (2019) 'A beginner's guide to cryogenic electron microscopy', *Portland Press Limited under the Creative Commons Attribution License 4.0 (CC BY-NC-ND)*, pp. 46–52.

Scheres, S. H. W. (2012) 'RELION: Implementation of a Bayesian approach to cryo-EM structure determination', *Journal of Structural Biology*, 180(3), pp. 519–530. doi: 10.1016/j.jsb.2012.09.006.

Schertler, G. F. X., Villa, C. and Henderson, R. (1993) 'Projection structure of rhodopsin', *Nature*, 362(6422), pp. 770–772. doi: 10.1038/362770a0.

Simon, M. D. *et al.* (2007) 'The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones', *Cell*. doi: 10.1016/j.cell.2006.12.041.

Simpson, J. P. *et al.* (2008) 'Identification and characterization of a plastid-localized Arabidopsis glyoxylate reductase isoform: Comparison with a cytosolic isoform and implications for cellular redox homeostasis and aldehyde detoxification', *Journal of Experimental Botany*. doi: 10.1093/jxb/ern123.

Smyth, M. S. and Martin, J. H. (2000) 'x ray crystallography.', *Molecular pathology : MP*, 53(1), pp. 8–14. doi: 10.1136/mp.53.1.8.

Takashi Ishida1, * and Kengo Kinoshita (2007) 'PrDOS: prediction of disordered protein regions from amino acid sequence', *Nucleic Acids Research*, 35, pp. W460–W464.

Tawfik, D. S. (2014) ‘Accuracy-rate tradeoffs: How do enzymes meet demands of selectivity and catalytic efficiency?’, *Current Opinion in Chemical Biology*. Elsevier Ltd, 21, pp. 73–80. doi: 10.1016/j.cbpa.2014.05.008.

Vermeulen, M. *et al.* (2010) ‘Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers’, *Cell*, 142, pp. 967–980. doi: 10.1016/j.cell.2010.08.020.

Vezzoli, A. *et al.* (2010) ‘Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1’, *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.1797.

Weaver, T. M., Morrison, E. A. and Musselman, C. A. (2018) ‘Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains’, *Molecules*, 23(2614). doi: 10.3390/molecules23102614.

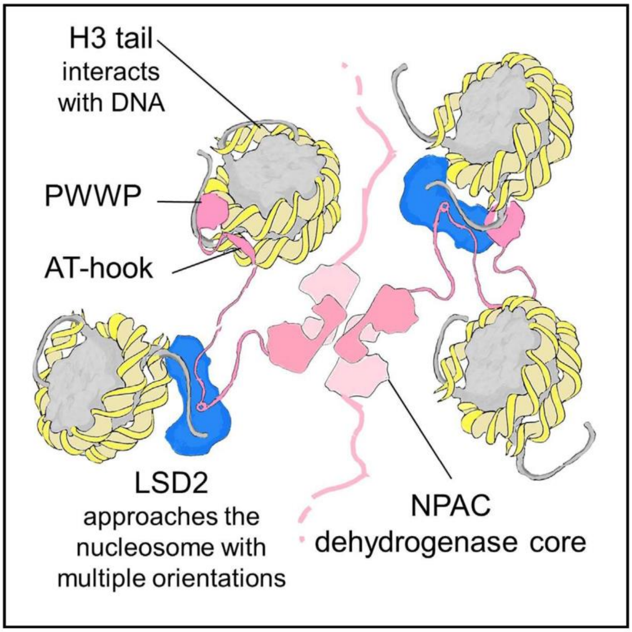
Zhang, Y. *et al.* (2014) ‘Structural characterization of a β-hydroxyacid dehydrogenase from Geobacter sulfurreducens and Geobacter metallireducens with succinic semialdehyde reductase activity’, *Biochimie*. doi: 10.1016/j.biochi.2014.05.002.

Zheng, S. Q. *et al.* (2017) ‘MotionCor2 - anisotropic correction of beam-induced motion for improved cryo-electron microscopy’, *Nature Methods*, 14(4), pp. 331–332. doi: 10.1186/s40945-017-0033-9.

Cell Reports

A Tail-Based Mechanism Drives Nucleosome Demethylation by the LSD2/NPAC Multimeric Complex

Graphical Abstract



Highlights

- LSD2 demethylase binds nucleosomes by discrete promiscuity at multiple sites
- Nucleosomal DNA represents an intrinsic barrier to histone demethylation
- NPAC has a rigid and catalytically inert oligomerization core and flexible arms
- The NPAC/LSD2 multifunctional enzyme complex modifies transcribed chromatin

Authors

Chiara Marabelli, Biagina Marrocco, Simona Pilotto, ..., Daniela Rhodes, Sriram Subramaniam, Andrea Mattevi

Correspondence

sriram.subramaniam@ubc.ca (S.S.),
andrea.mattevi@unipv.it (A.M.)

In Brief

Through biophysical, biochemical, and structural studies, including cryo-EM, Marabelli et al. describe how NPAC promotes LSD2 productive interaction with the nucleosome in a rapid and flexible manner. Their findings provide a molecular mechanism for LSD2 activity in the context of H3K4me2 demethylation during Pol II transcriptional elongation.

Marabelli et al., 2019, Cell Reports *27*, 387–399
April 9, 2019 © 2019 The Author(s).
https://doi.org/10.1016/j.celrep.2019.03.061

CellPress

Cell Reports
Article

OPEN
ACCESS
CellPress

A Tail-Based Mechanism Drives Nucleosome Demethylation by the LSD2/NPAC Multimeric Complex

Chiara Marabelli,[1,11] Biagina Marrocco,[1,11] Simona Pilotto,[1,9,11] Sagar Chittori,[2,11] Sarah Picaud,[3] Sara Marchese,[1]
Giuseppe Ciossani,[1,10] Federico Forneris,[1] Panagis Filippakopoulos,[3] Guy Schoehn,[4] Daniela Rhodes,[5,6,7]
Sriram Subramaniam,[8,*] and Andrea Mattevi[1,12,*]
[1]Department of Biology and Biotechnology "Lazzaro Spallanzani," University of Pavia, via Ferrata 9, 27100 Pavia, Italy
[2]Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, USA
[3]Structural Genomics Consortium, Nuffield Department of Clinical Medicine, University of Oxford, Headington, Oxford OX3 7DQ, UK
[4]Institut de Biologie Structurale (IBS), University Grenoble Alpes, CEA, CNRS, 38044 Grenoble, France
[5]Institute of Structural Biology, Nanyang Technological University, 59 Nanyang Drive, Singapore 636921, Singapore
[6]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore
[7]Lee Kong Chian School of Medicine, Nanyang Technological University, 59 Nanyang Drive, Singapore 636921, Singapore
[8]The University of British Columbia, 2215 Wesbrook Mall, Vancouver, BC V6T 1Z3, Canada
[9]Present address: Darwin Building, University College London, Gower Street, London WC1E 6BT, UK
[10]Present address: Department of Mechanistic Cell Biology, Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany
[11]These authors contributed equally
[12]Lead Contact
*Correspondence: sriram.subramaniam@ubc.ca (S.S.), andrea.mattevi@unipv.it (A.M.)
https://doi.org/10.1016/j.celrep.2019.03.061

SUMMARY

LSD1 and LSD2 are homologous histone demethylases with opposite biological outcomes related to chromatin silencing and transcription elongation, respectively. Unlike LSD1, LSD2 nucleosome-demethylase activity relies on a specific linker peptide from the multidomain protein NPAC. We used single-particle cryoelectron microscopy (cryo-EM), in combination with kinetic and mutational analysis, to analyze the mechanisms underlying the function of the human LSD2/NPAC-linker/nucleosome complex. Weak interactions between LSD2 and DNA enable multiple binding modes for the association of the demethylase to the nucleosome. The demethylase thereby captures mono- and dimethyl Lys4 of the H3 tail to afford histone demethylation. Our studies also establish that the dehydrogenase domain of NPAC serves as a catalytically inert oligomerization module. While LSD1/CoREST forms a nucleosome docking platform at silenced gene promoters, LSD2/NPAC is a multifunctional enzyme complex with flexible linkers, tailored for rapid chromatin modification, in conjunction with the advance of the RNA polymerase on actively transcribed genes.

INTRODUCTION

Nucleosomes are the tunable elements of chromatin. Multiple signaling pathways converge to build histone post-transcriptional modification patterns and in turn modulate the dynamics of the nucleosome by either directly affecting nucleosome stability or recruiting specific chromatin remodelers. Nucleosome assembly, disassembly, and modification are carried out by enzymatic subunits embedded within large macromolecular complexes (McGinty and Tan, 2015; Kim et al., 2019; Ricketts et al., 2019; Zhou et al., 2019). The histone demethylases LSD1 and LSD2 represent particularly important examples for exploring fundamental molecular mechanisms underlying these processes. These two enzymes share a highly homologous catalytic domain and employ an identical flavin-dependent reaction mechanism to process the same substrate: mono- and dimethyl Lys4 of histone H3 (H3K4me1 and H3K4me2) (Fang et al., 2013; Forneris et al., 2005; Karytinos et al., 2009; Lee et al., 2005; Shi et al., 2004). Despite their similar catalytic cores, LSD1 and LSD2 feature distinct non-catalytic domains, which form distinguishable binding platforms for different partners and accessory subunits. The elongated tower domain of LSD1 is bound tightly to CoREST (co-repressor of RE1 silencing transcription factor [REST]), which is necessary to grasp the histone tail from DNA (Kim et al., 2015; Pilotto et al., 2015). LSD1/CoREST also forms a bifunctional complex with histone deacetylases (HDAC1 and HDAC2) to establish chromatin repression at gene promoters (Marabelli et al., 2016; Shi et al., 2004). In contrast, LSD2 does not interact with HDACs and CoREST (Chen et al., 2013; Ciccone et al., 2009; Fang et al., 2010, 2013; Nagaoka et al., 2015; Zhang et al., 2013). The biological partner of LSD2 is NPAC (a cytokine-like nuclear factor, also named GLYR1 or NP60), a multidomain protein comprising an N-terminal PWWP domain and a conserved C-terminal dehydrogenase domain (Figure 1A). The two are connected by a long stretch of roughly 170 residues comprising two protein-protein interaction motifs. An AT-hook module (residues 168–180) that binds and activates the mitogen-activated protein kinase (MAPK) p38α is followed by a linker segment (residues 214–225) that binds LSD2 with submicromolar affinity (Figure 1A; Fang et al., 2013; Fu et al., 2006). This short NPAC segment was shown to stimulate the nucleosome-demethylase activity of LSD2 by binding to a

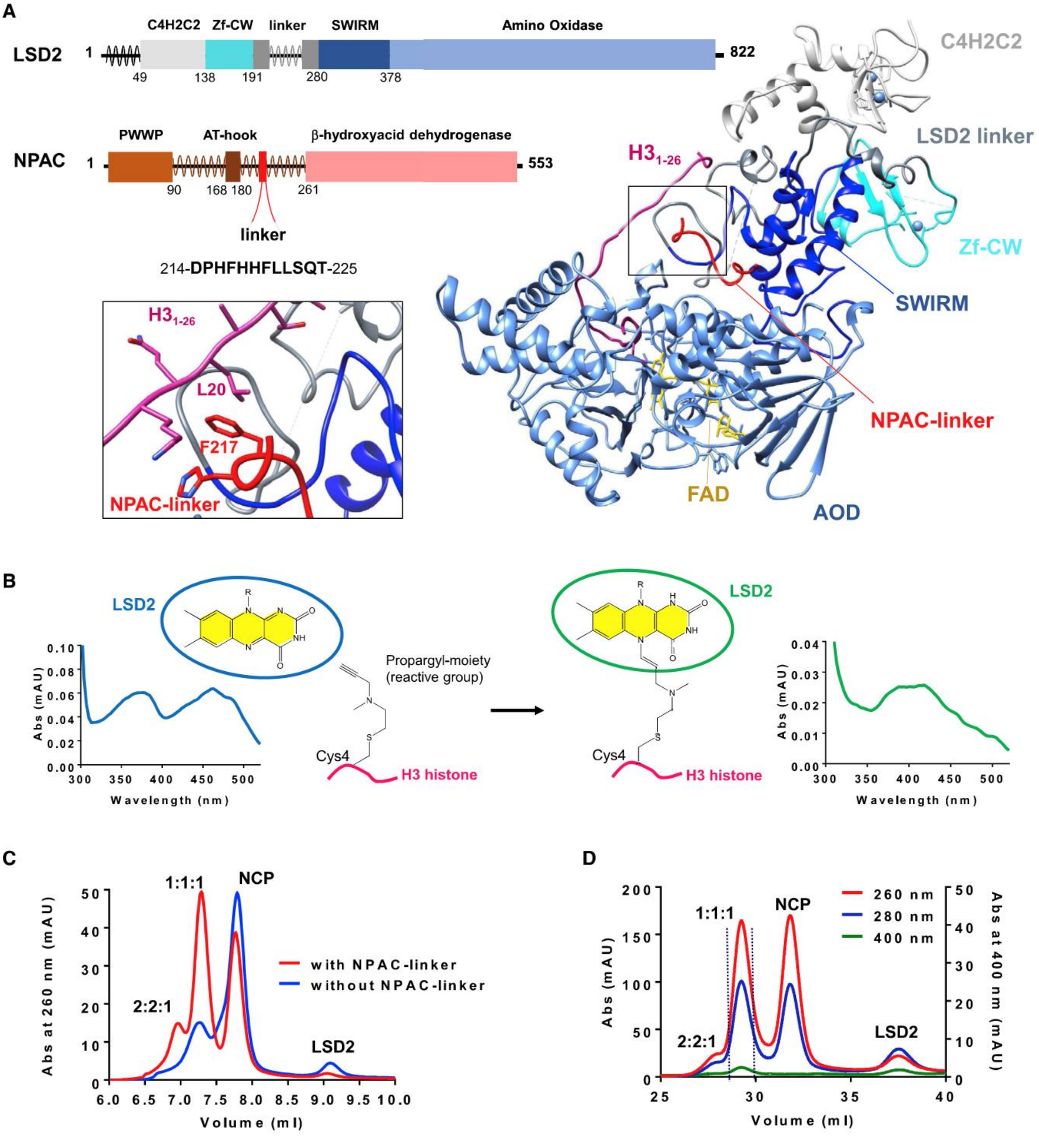


Figure 1. NPAC Linker Sustains Productive Nucleosome Recognition by LSD2

(A) Domain organization of LSD2 and NPAC. Disordered regions are in wavy lines. LSD2 (PDB: 4hsu) is colored according to its multidomain architecture, with the flavin adenine dinucleotide (FAD) in gold and the zinc ions in light blue.

(B) Our semisynthetic nucleosomes form a covalent adduct with the FAD of LSD2 (absorbance peak at 400 nm rather than at 458 nm, as for the oxidized enzyme).

(C) Elution profile (WTC-030S5 column, Wyatt) of semisynthetic NCP (10 μM) and LSD2 (20 μM) after 1 h of incubation with or without the NPAC linker (100 μM). Respectively, 2:2:1 and 1:1:1 stand for $(LSD2/NPAC\text{-}linker)_2/NCP$ and LSD2/NPAC-linker/NCP complexes. See also Figure S1.

(D) Elution profile of the LSD2/NPAC-linker/NCP sample used for cryo-EM (dashed lines; three 10/300 columns of Superdex 200 in series, GE Healthcare). Protein, DNA, and the flavin-H3 covalent adduct were detected by monitoring the absorbance at 280, 260, and 400 nm, respectively.

hydrophobic groove between the amine oxidase and the SWIRM domains of the demethylase (Figure 1A; Chen et al., 2013; Fang et al., 2013).

Several studies consistently demonstrate that both LSD2 and NPAC localize within trimethylated Lys36 of histone H3 (H3K36me3)-rich chromatin regions of actively transcribed gene bodies (Fang et al., 2010; Fei et al., 2018; Vermeulen et al., 2010). Depletion of either LSD2 or NPAC leads to similar inhibitory effects on gene transcription elongation (Fang et al., 2010; Fei et al., 2018). This seems to indicate that the LSD2/NPAC system processes chromatin before the passage of the RNA polymerase II (Pol II) elongating complex along DNA. Consistent with this hypothesis, the H3K4me2 signal decreases downstream of Pol II, and it is reconstituted immediately upstream (Barski et al., 2007). However, it remains elusive how a repressive epigenetic signal such as H3K4 demethylation would be important for transcriptional elongation.

The LSD2/NPAC complex represents a canonical example of chromatin regulation by the interplay between catalytic and non-catalytic modules within a single multisubunit nucleosome-modifying complex, but numerous fundamental questions remain to be answered. What is the role of the characteristic non-catalytic domains of LSD2? Do they promote the engagement of the histone tail within the active site? How does the short linker sequence of NPAC stimulate nucleosome demethylation? What is the role of DNA and the histone octamer in substrate recognition? Above all, how do the different architectures of LSD1/CoREST and LSD2/NPAC give rise to their opposing biological outcomes: gene repression by LSD1 versus transcriptional activation by LSD2? Here, we addressed these questions by cryoelectron microscopy (cryo-EM) analysis of the structure and conformational plasticity of the LSD2/nucleosome complex, as well as by dissection of the molecular properties of the multidomain protein NPAC. Our studies lead to the unexpected conclusion that instead of being a nucleosome-docking unit like LSD1/CoREST, LSD2/NPAC is a multimeric enzyme complex exquisitely tailored for the efficient processing of nucleosomal histone tails.

RESULTS AND DISCUSSION

H3K4me1 and H3K4me2 Demethylation Is More Efficient with Longer H3-Tail Peptides but Not Affected by the NPAC Linker

Our first experiments aimed to dissect the binding and kinetics of the human LSD2 demethylase reaction using the recombinant human enzyme expressed in *Pichia pastoris* and several mono- and dimethylated H3 N-terminal peptides (Table 1). First, we confirmed the previously reported faster turnover rates for H3K4me2 over H3K4me1 substrates (Table 1; Chen et al., 2013). Second, we showed that increasing the lengths of the H3 peptides improves LSD2 catalytic efficiency, with as much as an 8-fold lower K_M for the dimethylated 40-residue over the 21-residue substrate. Third, the enzymatic activity was markedly affected by the ionic strength: at 100 mM NaCl, K_M for the $H3_{1-21}$ peptide is 8–10 times higher compared to the value measured in the absence of salt. This effect was less pronounced on the longer $H3_{1-40}$ peptides, for which an increase in K_M was only detectable at 200 mM NaCl (Table 1).

These data indicated that residues 20–40 of the H3 tail feature improved catalytic activity. To investigate their actual contribution to binding, we studied a charge-removing double mutation, targeting two Lys residues in the 21- to 30-residue segment. We found that the K23M-K27M $H3_{1-40}$ peptide is as effective as a substrate as the wild type. Moreover, we tested LSD2 activity at 100 mM NaCl in the presence of 100 μM $H3_{16-40}$ or $H3_{21-40}$ peptides. No inhibition was observed, indicating that the standalone $H3_{16-40}$ and $H3_{21-40}$ segments bind, at best, too weakly to hamper substrate recognition by LSD2 (Table 1). These findings highlight residues 1–20 as the main drivers of H3-tail binding to the demethylase, with the subsequent amino acids enhancing binding, especially under conditions of high ionic strength.

Next, we studied how H3 peptide demethylation is affected and potentially regulated by the NPAC linker (residues 214–225 of NPAC) (Figure 1A; Fang et al., 2013). The outcome of these studies was consistent: the NPAC linker barely influences the steady-state kinetics of H3 peptide binding and demethylation. The only noticeable effect was that NPAC tends to mitigate the worsening (increasing) of peptide K_M values at high NaCl concentrations (150–200 mM) (Table 1).

NPAC Sustains Productive Nucleosome Recognition by LSD2

To focus on the recognition and demethylation mechanisms of the nucleosome, we reconstituted semisynthetic nucleosome core particles (NCPs) with a chemically modified recombinant H3 K4C-C110A double mutant exposing a propargylic group instead of the physiological dimethyl Lys4 side chain (Pilotto et al., 2015). These semisynthetic NCPs function as a suicide substrate: their H3 tail assumes a catalytically competent conformation within the enzyme active pocket and thereby covalently attacks the flavin prosthetic group (Figure 1B). Formation of the covalent NCP/demethylase complex can be easily visualized and quantified by analytical size-exclusion chromatography (Figure 1C).

Incubation trials with LSD2 and the semisynthetic NCPs resulted in poor, albeit clearly detectable, LSD2/NCP complex formation, but the addition of the NPAC-linker peptide led to a marked 7-fold increase in formation of the complex (Figure 1C). This value matches the reported NPAC effect on nucleosome demethylation by LSD2 as measured by published western blot assays (Chen et al., 2013; Fang et al., 2013). We systematically observed the presence of two peaks in the chromatograms. Based on their elution volumes and DNA content, gathered by the relative absorbance at 260 nm, we attributed the major peak to a 1:1:1 LSD2/NPAC-linker/NCP complex and the minor peak to a 2:2:1 complex (i.e., one LSD2/NPAC bound to each H3 of the same NCP). Neither varying the ionic strength nor performing charge-removing mutations of H3 (K23M-K27M) affected nucleosome binding with or without the NPAC linker (Figure S1). Therefore, our semisynthetic NCPs proved to be effective, useful tools for studying nucleosome recognition by LSD2 and the activating role of NPAC.

Table 1. H3K4me1 and H3K4me2 Demethylation Is More Efficient with Longer H3-Tail Peptides but Not Affected by the NPAC Linker

			Monomethyl Lys4[a]		Dimethyl Lys4[a]				
	NPAC[b]	NaCl (mM)	1–21	1–30	1–21	1–30	1–40	1–40 K23M K27M	1–40 + DNA[d]
K_M (μM)	–	0	5.89 ± 0.69	1.34 ± 0.13	1.91 ± 0.18	2.90 ± 0.27	0.99 ± 0.07	0.89 ± 0.05	no activity
	+	0	6.20 ± 0.69	1.18 ± 0.15	2.11 ± 0.19	1.85 ± 0.25	1.09 ± 0.06	0.68 ± 0.08	no activity
	–	50	ND	ND	ND	ND	ND	ND	46.02 ± 7.44
	+	50	ND	ND	ND	ND	ND	ND	>80.00
	–	100	>60.00	5.11 ± 0.79	16.93 ± 3.08	6.18 ± 0.63	2.06 ± 0.25	2.51 ± 0.42	7.00 ± 0.87
	+	100	>60.00	2.36 ± 0.21	16.29 ± 3.59	6.04 ± 0.93	1.55 ± 0.12	1.02 ± 0.11	6.24 ± 0.62
	–	150	ND	ND	ND	13.50 ± 1.54	3.22 ± 0.35	ND	ND
	+	150	ND	ND	ND	8.15 ± 1.11	2.03 ± 0.18	ND	ND
	–	200	ND	ND	ND	ND	6.46 ± 0.58	ND	ND
	+	200	ND	ND	ND	ND	4.94 ± 0.65	ND	ND
k_{cat} (min^{-1})	–	0	0.44 ± 0.01	0.56 ± 0.01	1.01 ± 0.02	1.39 ± 0.03	1.30 ± 0.02	1.24 ± 0.20	no activity
	+	0	0.66 ± 0.02	0.48 ± 0.01	1.02 ± 0.02	1.15 ± 0.04	1.18 ± 0.02	1.04 ± 0.02	no activity
	–	50	ND	ND	ND	ND	ND	ND	1.84 ± 0.15
	+	50	ND	ND	ND	ND	ND	ND	2.42 ± 0.38
	–	100	0.23 ± 0.02	0.41 ± 0.02	0.72 ± 0.05	1.46 ± 0.04	0.83 ± 0.03	0.92 ± 0.04	1.38 ± 0.06
	+	100	0.17 ± 0.03	0.34 ± 0.01	0.70 ± 0.06	1.30 ± 0.07	0.67 ± 0.01	0.64 ± 0.02	1.32 ± 0.04
	–	150	ND	ND	ND	1.01 ± 0.04	1.15 ± 0.03	ND	ND
	+	150	ND	ND	ND	1.19 ± 0.05	1.05 ± 0.02	ND	ND
	–	200	ND	ND	ND	ND	1.23 ± 0.03	ND	ND
	+	200	ND	ND	ND	ND	1.39 ± 0.05	ND	ND
K_d[c] (μM)	–	0	ND	ND	1.99 ± 0.30	ND	ND	ND	ND
	+	0	ND	ND	3.08 ± 0.38	ND	ND	ND	ND

Data are represented as mean ± SEM. All assays were carried out with a peroxidase-coupled method. ND, not determined.

[a] The number of amino acids of the histone H3 N-terminal peptide is indicated.

[b] NPAC-linker sequence 214–225 (Figure 1A).

[c] K_d values were measured by fluorescence polarization using C-terminally tetramethylrhodamine (TAMRA)-labeled histone H3 peptide with sequence ARTKme2QTARKSTGGKAPRKQLA.

[d] The effect of nucleosomal DNA on histone peptide demethylase activity: the 147-bp 601 DNA sequence was used at a final concentration of 0.5 μM.

DNA Interferes with LSD2 Catalysis

Collectively, the experiments with the H3 peptides and the intact nucleosomal particles led to the surprising observations that LSD2 is inhibited by the nucleosome and this inhibition is rescued by the NPAC linker. To investigate the source of this nucleosome-induced inhibitory effect on histone demethylation by LSD2, we first established whether non-substrate histone tails might compete with H3 for the LSD2 active site. We probed the kinetics of peptide demethylation in the presence of high concentrations (100 μM) of $H2A_{1-19}$, $H2A_{117-129}$, $H2B_{4-19}$, or $H4_{1-30}$ peptides. No inhibition was detected, suggesting that the H2A, H2B, and H4 tails do not interfere with LSD2 enzymatic activity (Figure S2). Given this result, we next assayed LSD2 in the presence of the isolated 147-bp 601 Widom DNA to probe its effect on catalysis. At low ionic strength (no NaCl added) activity was hardly detectable in the presence of DNA, both with and without the NPAC linker. The demethylase activity was recovered at 100 mM NaCl, though with a significant 3-fold increase of K_M with respect to the experiments performed in the absence of DNA (Table 1). These findings lead to the fundamental conclusion that a main source of the NCP-exerted inhibition is DNA.

The notion that DNA forms an intrinsic obstacle to nucleosome binding and modification is well documented, and it has been generally ascribed to the pronounced charge density of DNA (McGinty and Tan, 2015; Stützer et al., 2016; Gatchalian et al., 2017; Morrison et al., 2018; Weaver et al., 2018). With reference to LSD2, this phenomenon is functionally crucial, because it is at the heart of the regulatory effect exerted by NPAC. The specific electrostatic interactions between the N-terminal, highly charged, 15 amino acids of H3 and the inner catalytic pocket of the demethylase are inevitably weakened or even compromised by the contiguous crowded-with-charge nucleosomal DNA (Table 1). Conversely, the binding of the $H3_{16-26}$ segment to the outer active-site region of the LSD2/NPAC-linker complex depends less on charged groups (see Table 1 and the K23M-K27M mutation in Figure S1C). Instead, the binding of the $H3_{16-26}$ segment to the LSD2/NPAC-linker complex involves several hydrophobic contacts, especially with Phe217 of NPAC (Figure 1A; Chen et al., 2013; Fang et al., 2013). These interactions are expected to become increasingly more relevant at higher ionic strengths, closer to the physiological conditions, enforcing the influence of NPAC on the LSD2 demethylase activity (Table 1). This NPAC dependency is fully exhibited in the nucleosomal microenvironment, where not only the availability but also conformational flexibility of the tail is limited by the neighboring DNA (Morrison et al., 2018; Weaver et al., 2018).

Single-Particle Cryo-EM Reveals that Multiple Conformations and Few Interactions Characterize the LSD2/NPAC/NCP Complex

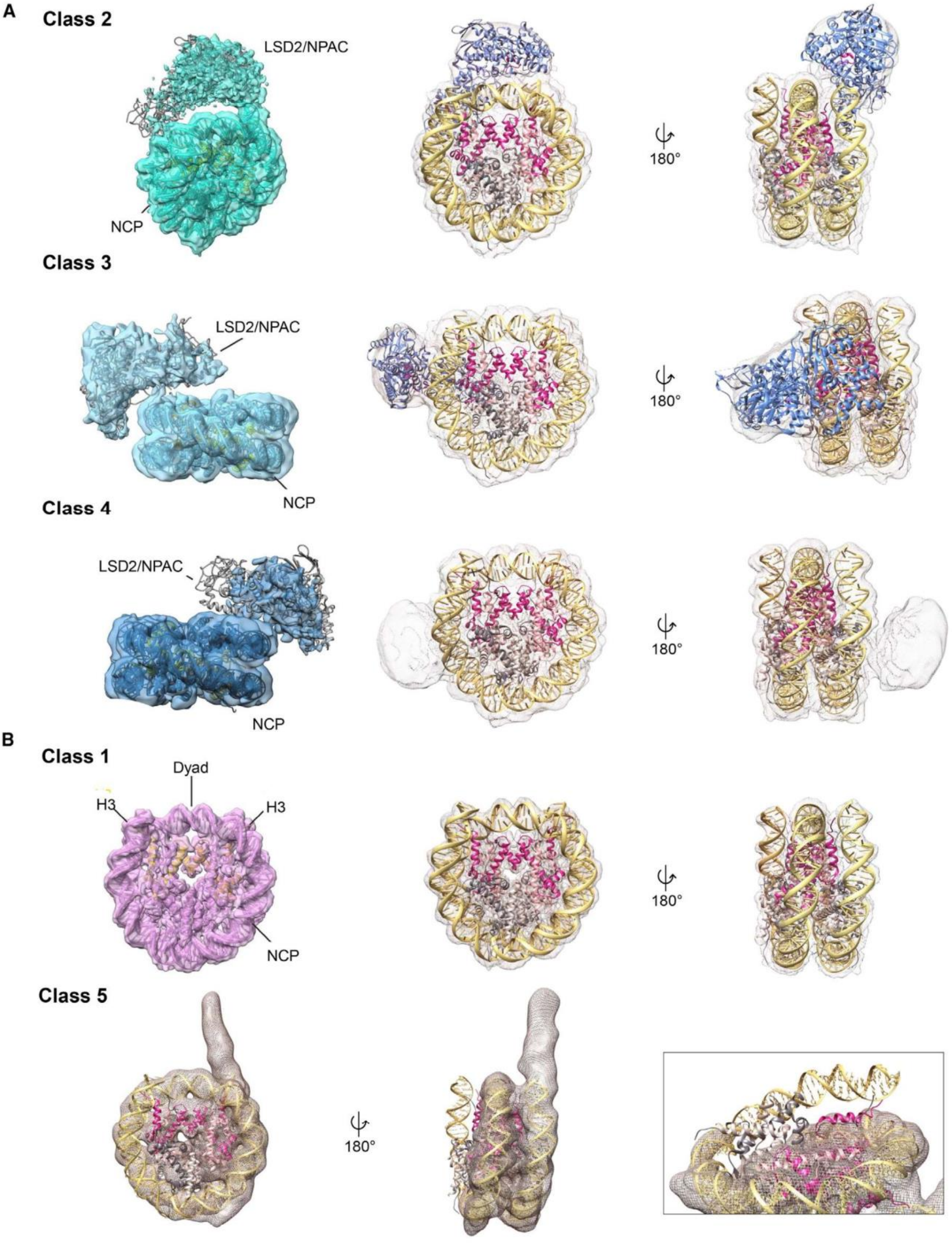
To structurally describe the mechanism of nucleosome binding by LSD2/NPAC, we carried out single-particle cryo-EM. We initially employed the semisynthetic nucleosomes reconstituted with the 167-bp 601 Widom sequence that comprises the DNA-linker segments, but the resulting LSD2/NPAC-linker/nucleosome complex could not be separated from the unbound nucleosomes. The semisynthetic NCPs reconstituted with the 147-bp 601 DNA gave excellent chromatographic profiles and enabled the preparative milligram-quantity purification of the covalent LSD2/NPAC-linker/NCP complex (Figure 1D). The specific cross-linking between the LSD2 flavin and the modified H3K4 enabled the purification of an otherwise inherently short-lived enzyme-substrate complex, without the need for external cross-linking agents (Figure 1B). The cryo-EM analysis revealed five distinct structures, and three of them (classes 2–4) correspond to clearly defined LSD2/nucleosome complexes (Figure 2). The demethylase region of the map is always less resolved, suggesting local variability of the LSD2 orientation, but the structures are still fully interpretable (Figures S3 and S4).

In class 2, the body of LSD2 arches over the nucleosome dyad (Figure 2A). The contacts between LSD2 and NCP are confined to two small areas (Figure 3A). In the first one, the H3-tail-binding loop (residues 478–482) of the LSD2 amine oxidase domain contacts the two base pairs of the nucleosomal DNA entry-exit point. In the second contact region, a loop on the C4H2C2 domain of LSD2 (residues 101–107) interacts with a short segment of the H3 α1 helix (around Glu73) and the N-terminal tail of H4 (around Asp24). The trace of the H3 N-terminal segment between the LSD2 catalytic pocket and the nucleosome exit point cannot be revealed at the resolution of our density maps. However, structural analysis predicts that both H3 histones of the NCP can be engaged by LSD2. The first possibility is in line with the LSD2/NPAC-linker/$H3_{1-26}$ co-crystal structure: residues 23–35 of H3 extend over a distance of ~33 Å from the rim of the LSD2 catalytic cleft to the site where the tail protrudes from the nucleosome (dashed in Figure 3A, middle). The second possibility is that Met15 of H3, at the entrance of the LSD2 catalytic pocket, is connected to Lys36, located at the nucleosomal exit point that is closer to LSD2 amine oxidase (35-Å distance) (dashed in Figure 3A, left).

In class 3, the L-shaped LSD2 arches over the edge of the NCP and protrudes laterally from the nucleosomal disk (Figure 2A). The single contact point between the demethylase and the NCP involves a loop on the LSD2 linker region (residues 275–280) and the nucleosomal DNA at the super-helical location (SHL) ± 2 (Figure 3B). The distance covered by H3 from Met15 within the LSD2 active site to Lys36 in the NCP is ~43 Å, which is compatible with multiple extended conformations of the tail (dashed in Figure 3B, right).

From the inspection of the less well-defined class 4 map, it can be inferred that LSD2 sits tangentially to the NCP disk in proximity of the SHL ± 6 (Figure 2A). The putative contact areas on the NCP comprise a few nucleotides, together with residues from H2A (α2 and α3 helices) and H2B (loop α1–α2), close to the acidic patch (Figure 3C).

Despite their strikingly different binding configurations, the LSD2/NPAC-linker/NCP structures share a few insightful features. In all cases, the demethylase-NCP contacts are confined to a few amino acids and nucleotides and there is no extensive interacting surface between the two components. Furthermore, the H3 segment linking the LSD2 active site to the NCP core—from Met15 to Lys36—is always predicted to follow trajectories running along the DNA grooves. Consistently, these trajectories fall within the ensembles of nucleosomal H3-tail conformations



(legend on next page)

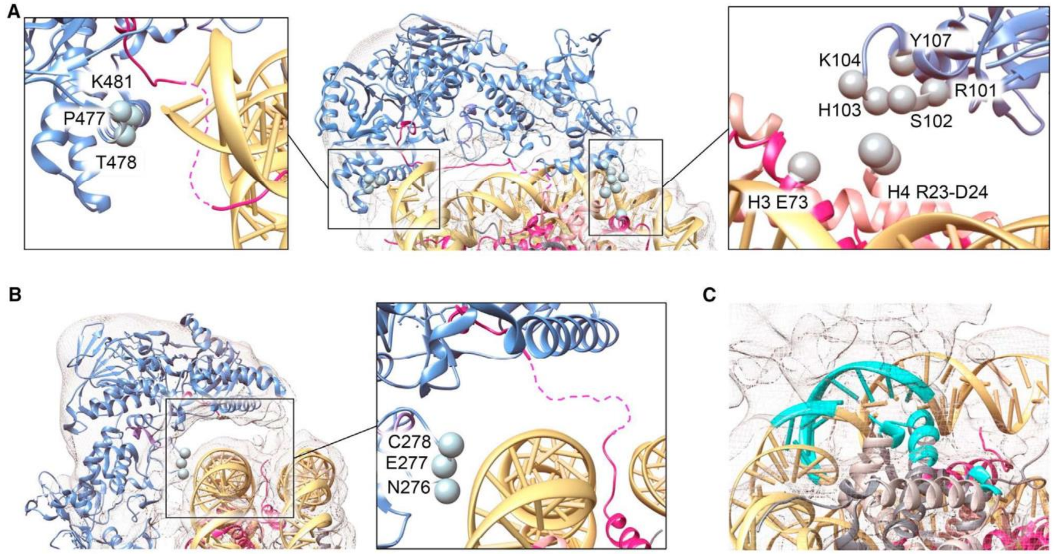


Figure 3. Close-Up Views of LSD2/NCP Interfaces

(A) Class 2 features two contact points between the LSD2 and the nucleosome, depicted in the close-up views.

(B) Class 3 displays a single contact point.

(C) Contact region (in cyan) of class 4 involves DNA super-helical location ± 6 and two short segments of H2A and H2B. Colors are the same as in Figure 2, with the NPAC linker in violet. The density maps are depicted as light gray mesh. Reference residues are labeled, and their Cα atoms are shown as spheres. Dashed lines highlight possible pathways of H3 tails connecting the NCP to the LSD2 catalytic site. See also Figures S3 and S4.

revealed by NMR and computational studies (Morrison et al., 2018). These findings demonstrate that the flexibility of the H3 tail enables multiple and catalytically active binding poses with limited contacts between the LSD2 and the nucleosomal main body.

The cryo-EM analysis revealed two other classes comprising isolated NCPs, though in different states. Class 1 contains roughly half of the particle images and corresponds to the isolated and fully assembled NCP at an overall resolution of 4.02 Å (Figures 2B, S3, and S4). Class 5 represents a hexasome, with its visible protruding DNA. The density map for this class shows that one H2A-H2B dimer is missing and, on the side of the nucleosome where DNA slipped away, the N-terminal α helix of H3 is unfolded (Figures 2C, S3, and S4). The histone-DNA particle featured by class 5 was described previously as a state of nucleosome dynamics and/or unfolding (Bilokapic et al., 2018). Therefore, the same cryo-EM grid allowed us to visualize a partly unfolded NCP, a fully folded particle, and three LSD2 complexes in different conformations.

LSD2 Engages the DNA Template via Electrostatic Interactions

Inspection of the LSD2 protein surface outlines several clusters of positively charged side chains, mostly localized on the zinc-finger domain and in contact with DNA, as shown by the cryo-EM three-dimensional structures (Figures 4A and 4B). To investigate the specificity and magnitude of these interactions, we carried out fluorescence polarization experiments with a 21-bp oligonucleotide. At low ionic strength, full-length LSD2 exhibited nanomolar affinity for DNA, whereas at 100 mM NaCl, the condition used for cryo-EM sample preparation and binding studies, the affinity became barely detectable (Figure 4C; Table S1). To gain more insights into DNA binding by LSD2, we mutated nine Arg and Lys residues that are part of the class 2 and class 3 LSD2/NCP interactions (Figures 4A and 4B), as well as two highly charged disordered segments (the N-terminal 30 residues and the conserved loop 241–258 at the rim of the H3-binding cleft) (Figure 1A). The mutations did not affect DNA binding, with the exception of the N-terminal Δ30 variant, which featured a

Figure 2. Single-Particle Cryo-EM Reveals that Multiple Conformations and Few Interactions Characterize the LSD2/NPAC-Linker/NCP Complex

The quality of the density maps can be best appreciated at the left, whereas model fitting can be best visualized at the right. H2A, H2B, H3, and H4 are shown in dark gray, light gray, purple, and pink, respectively. LSD2 (PDB: 4hsu) is in light blue. The NCP structure used for map fitting was obtained from PDB: 6esf. The DNA molecule is entirely visible in all maps. The first and the last three base pairs were modified to match our 601 sequence exactly. See also Figures S3 and S4.

(A) Overview of the three classes of the LSD2/NPAC-linker/NCP complex. For class 4, LSD2 is tentatively fitted at the left, simply as a reference.

(B) Class 1 represents the intact NCP, whereas class 5 is a partly unfolded NCP: a H2A-H2B dimer is missing, and the N-terminal helix of H3 is unstructured.

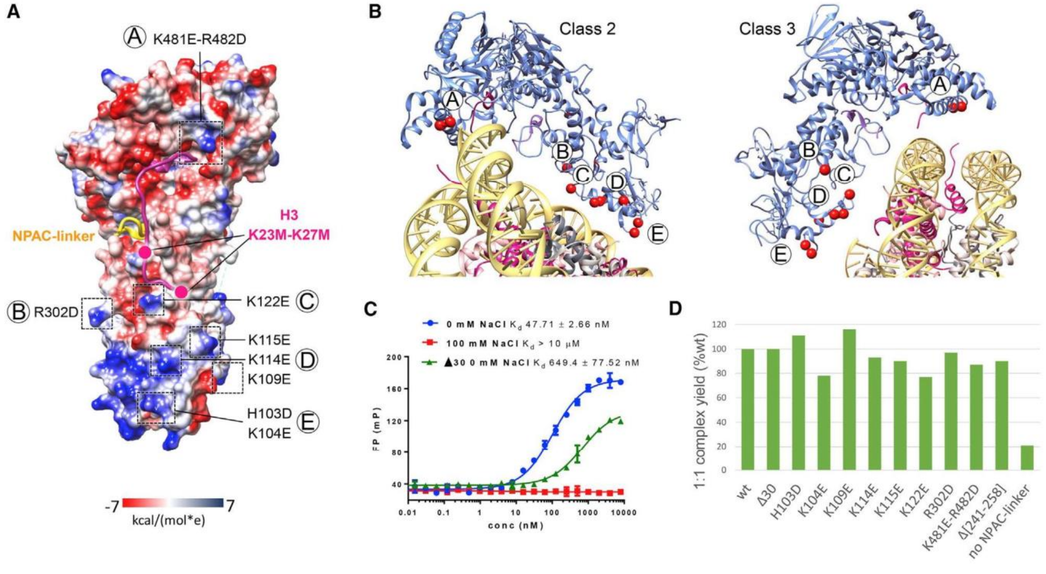


Figure 4. LSD2 Engages the DNA Template via Electrostatic Interactions

(A) The electrostatic surface of LSD2 exposed toward the nucleosome. Positively charged patches characterize the zinc-finger domain and part of the area surrounding the $H3_{1-16}$-binding pocket. The bar indicates the electrostatic potential in kcal/mol*e. Red represents negative electrostatic potential while blue represents positive charge potential.

(B) Close-up views of the LSD2-nucleosome interfaces in class 2 (middle) and class 3 (right), highlighting the LSD2 residues subjected to mutagenesis (red spheres at Cα atoms).

(C) Effect of ionic strength on the binding affinities of full-length and Δ30 LSD2 to DNA. Changes in the fluorescence polarization were measured in millipolarization (mP) units and plotted against LSD2 concentrations. Error bars refer to experiments carried out in triplicate. Data are represented as mean ± SEM. See also Table S1.

(D) Qualitative evaluation of LSD2 mutations on LSD2/NPAC-linker/NCP complex yield. The histogram shows the ratio of the LSD2/NPAC-linker/NCP peak to the absorbance of the free NCP peak (both recorded at 260 nm). The ratios are reported as percentages with reference to LSD2 wild type, which was given the 100% value.

10-fold decrease in affinity (Figure 4C; Table S1). We also found that all mutants bind NCP without drastic alteration in their efficiency with respect to wild-type protein (Figure 4D). These data corroborate the structural findings: LSD2 can weakly bind DNA through salt-sensitive electrostatic interactions. This association is likely brought about by various positively charged patches and does not depend on any specific cluster of residues and/or specific binding geometry.

The PWWP Domain of NPAC Is a H3 Reader and Strongly Binds DNA

In addition to the short LSD2-stimulating linker sequence, NPAC comprises several domains whose molecular functions remain poorly defined (Figure 1A). The NPAC N-terminal PWWP module is homologous to other known H3K36me3 binders (Qin and Min, 2014), in agreement with the NPAC localization at H3K36me3 chromatin loci (Vermeulen et al., 2010). However, NPAC has been shown to bind nucleosomes both with and without specific epigenetic marks (Fei et al., 2018; Sankaran et al., 2016). To determine whether the NPAC PWWP domain is a genuine histone reader, we used the SPOT-peptide array technology, employing a large library of epigenetically modified H3 and H4 peptides (Table S2). We observed a clear selectivity toward peptides containing residues 35–40 of H3, without an evident preference for specific epigenetic markers (Figure S5A). Consistent with this result, isothermal calorimetry performed with the unmodified $H3_{1-40}$ peptide yielded a K_d value of 30 μM (Figure S5B). The unfavorable reduction of conformational entropy upon binding is likely induced by the PWWP-guided partial folding of the 40-residue long peptide, in congruence with the previously characterized PWWP domain-histone peptide interactions (Eidahl et al., 2013; van Nuland et al., 2013). Bio-layer interferometry experiments on $H3_{28-48}$ peptides confirmed the affinity for the H3 C-terminal portion of the H3 tail, with no preference for the methylation states of K36 (Figure S5C). Next, we explored the capability of the NPAC PWWP domain to bind DNA. Fluorescence polarization experiments revealed a strong—primarily electrostatically driven—interaction, described by K_d = 0.68 μM at 100 mM NaCl (Figure S5D). Collectively, these observations highlight the PWWP of NPAC as a potent chromatin binder with very high DNA affinity when compared to other proteins of the same family (van Nuland et al., 2013; Weaver et al.,

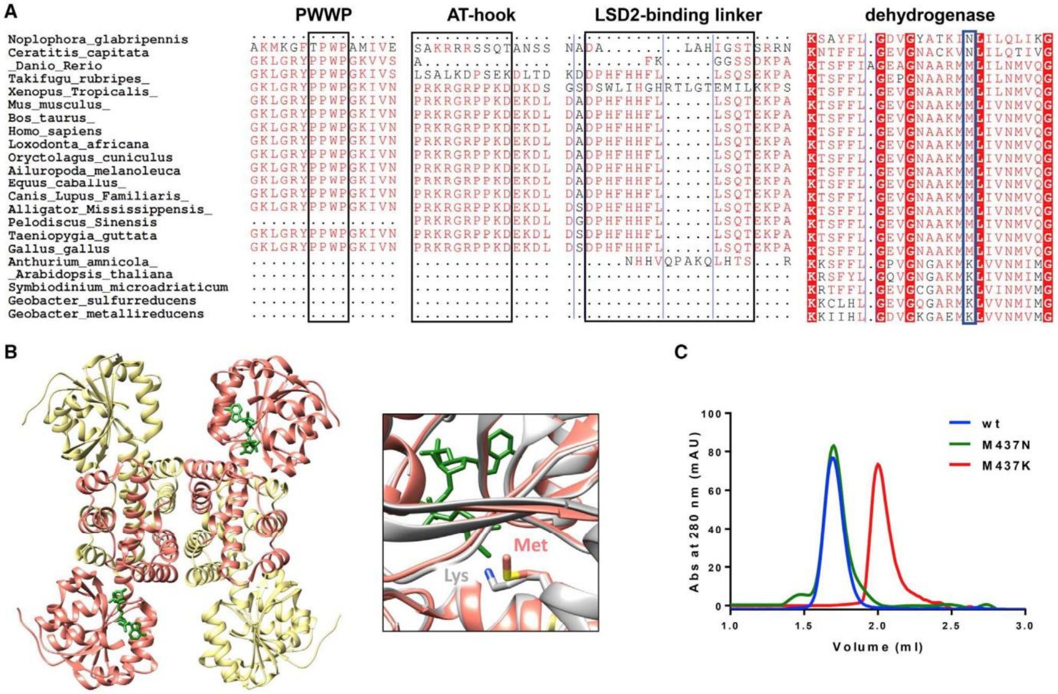


Figure 5. NPAC Dehydrogenase Domain Is Catalytically Impaired and Forms a Stable Tetramer

(A) NPAC homolog sequences are aligned with two representative members of the short-chain alcohol-dehydrogenase family from *Geobacter metallireducens* and *G. sulfurreducens*. The mutation of the catalytic lysine to methionine or asparagine is highlighted within a blue box.

(B) Crystal structure of the NPAC dehydrogenase domain (residues 261–553) shows a tetrameric assembly (PDB: 2uyy). The NADPH is visible in two subunits and is in green. See also Figures S6 and S7 and Table S3. The inset shows the comparison between the active-site regions of the NPAC dehydrogenase (in salmon) and those of the γ-hydroxybutyrate dehydrogenase from *G. sulfurreducens* (PDB: 3pdu; gray).

(C) Elution profiles of wild-type, M437N, and M437K NPAC dehydrogenase domains (5/150 column; Superdex 200, GE Healthcare).

2018; Wen et al., 2014). NPAC has been proposed to function as a nucleosome-destabilizing factor (Fei et al., 2018). Given its tight binding properties, the PWWP is likely responsible for this activity. In the course of our experiments, the nucleosomes tended to extensively precipitate upon mixing with the protein domain, corroborating this hypothesis.

NPAC Dehydrogenase Domain Is a Catalytically Inert, Tetramerization Module

The C-terminal domain of NPAC (residues 261–553) (Figures 1A and 5A) belongs to the highly conserved family of β-hydroxy acid dehydrogenases: cytosolic enzymes catalyzing the NAD^+- or $NADP^+$-dependent oxidation of specific β-hydroxy acid substrates (Njau et al., 2001; Zhang et al., 1999). Sequence homology correlates the dehydrogenase domain of NPAC to glyoxylate reductase. Despite featuring this fascinating topology for a chromatin protein, the function of the NPAC dehydrogenase domain remains unknown. As part of our investigation of the LSD2/NPAC system, we first confirmed that the protein tightly binds nicotinamide adenine dinucleotide phosphate (NADPH; Figure S6). Moreover, gel filtration and small-angle X-ray scattering (SAXS) firmly demonstrated that the protein is tetrameric in solution (Figures 5B and S7; Table S3) (PDB: 2uyy). Therefore, the NPAC dehydrogenase domain is a tetrameric NADH- or NADPH-binding protein that displays features typical of the β-hydroxy acid dehydrogenases.

We observed that the NPAC dehydrogenase domain hosts a methionine residue (Met437) in place of the highly conserved and catalytically essential lysine shared by all β-hydroxy acid dehydrogenase enzymes (Figures 5A and 5B). This finding strongly suggests that NPAC is catalytically inert, at least as a dehydrogenase enzyme. Consistently, a methionine or an asparagine at this position is a conserved feature among all NPAC orthologs exhibiting the LSD2-binding linker sequence (Figure 5A). Along these lines, we characterized two mutants: M437K and M437N. Surprisingly, the M437K protein proved to be monomeric but aggregation prone (Figure 5C) and far less stable than the wild type (10°C lower melting temperature). Conversely, the M437N mutant retained the tetrameric arrangement and, upon addition of NADPH, exhibited a thermostability value comparable to

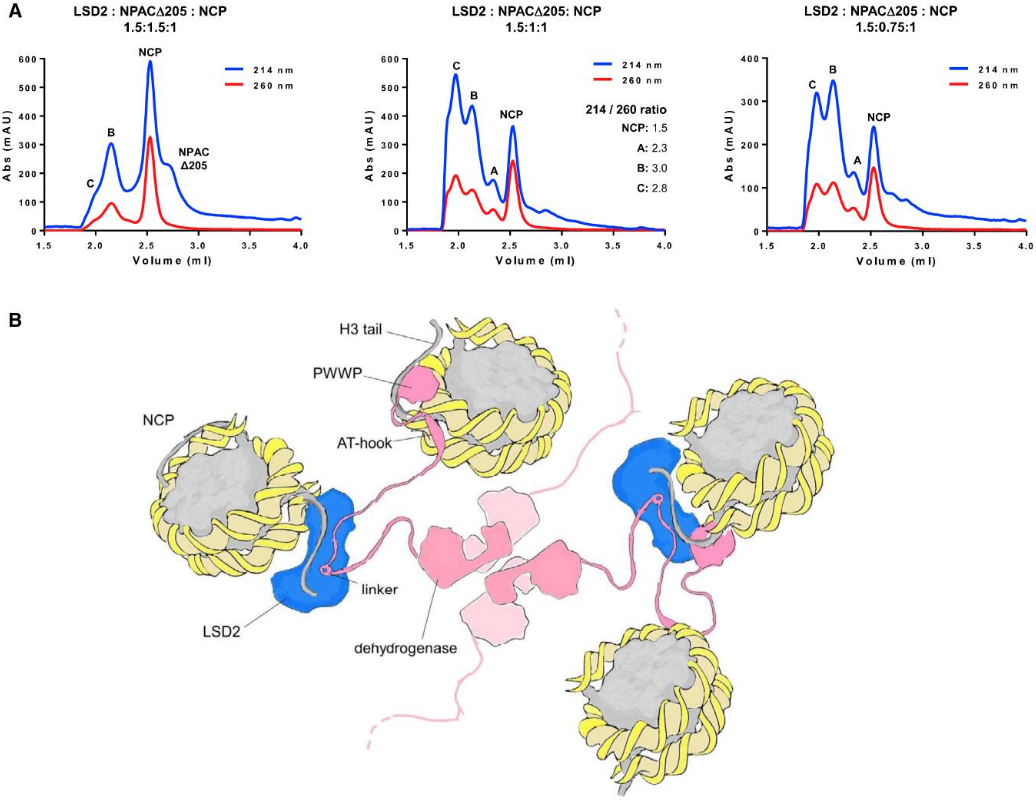


Figure 6. NPAC Tetramer Binds Multiple Copies of LSD2/Nucleosome

(A) NCP, LSD2, and NPACΔ205 (residues 206–553, containing the linker sequence and the dehydrogenase domain; see Figure 1A) were incubated at different molar ratios as shown above each panel (NPACΔ205 concentrations refer to the tetramer, molecular weight [MW] = 151 kDa) (Table S3). Decreasing relative amounts of NPACΔ205 with respect to nucleosome and LSD2 favors the formation of species at a higher molecular weight with greater DNA content (peak C), as calculated from the A_{214}/A_{260} ratio, which reflects the protein/DNA content. Peak A elutes at the same volume of the LSD2/NCP complex containing the NPAC linker only (V_e = 2.35 mL). This peak, occasionally present in the chromatograms, is a species that has lost an intact NPACΔ205 (likely by proteolysis). The experiments were performed on a WTC-030N5 column. See also Figure S7 and Table S3.

(B) Dehydrogenase domain of NPAC forms a stable tetramer decorated by flexible N-terminal arms, which comprise a PWWP domain, an AT-hook region predicted to bind DNA, and the LSD2-activating segment.

that of the wild-type protein. Combining these results, we concluded that in the course of evolution, NPAC lost essential catalytic elements but retained NADH or NADPH binding and consolidated a stable tetrameric assembly. Therefore, the dehydrogenase domain has all features of a stable oligomerization module.

LSD2/NPAC Is a Multimeric Complex

To broaden our understanding of LSD2/NPAC, we designed a human NPAC variant that comprised both the LSD2-binding linker and the C-terminal dehydrogenase domain (NPACΔ205) (Figure 1A). This protein behaved as predicted: it was tetrameric (SAXS data in Figure S7 and Table S3) and retained the ability to bind LSD2 and sustain nucleosome recognition (Figure 6A). These properties enabled the reconstitution of large complexes involving LSD2, tetrameric NPAC, and NCP. With increasing NCP and LSD2 relatively to the NPACΔ205 tetramer, a larger assembly (peak C in the chromatogram of Figure 6A) was favored over a minor yet conspicuous complex species (peak B), which instead prevailed at higher NPACΔ205 concentrations. The peak C complex featured higher DNA content than the peak B species, as indicated by the DNA-to-protein absorbance ratios (A_{214}/A_{260}). The large (~0.5 MDa) size of these complexes was confirmed by SAXS (Figure S7; Table S3). The interpretation of these data is inherently challenging due to the complexity of a system comprising multivalent binding species: the tetrameric NPAC, the dimeric histone core of the nucleosome, and the monomeric LSD2. Nevertheless, the chromatographic profiles, DNA content analysis, and SAXS data convincingly converge toward the idea that with increasing LSD2 and NCP concentrations, multimeric assemblies form, containing multiple copies of LSD2 and NCP per NPAC tetramer (Table S3).

CONCLUSIONS

The key finding of our work is that LSD2 and NPAC constitute a multimeric system with a rigid core and flexible units. The N-terminal portions of NPAC stretch out from the C-terminal stable tetrameric core, evolved from an ancestral dehydrogenase enzyme. Each of these flexible arms exposes two chromatin-binding modules plus a chromatin-modifying enzyme. The AT hook and the PWWP grasp the nucleosomal particles, possibly increasing the avidity and processivity of neighboring LSD2 enzymes (Figure 6B). The positively charged surface patches featured by LSD2, particularly by the zinc-finger domain, enable its weak association with nucleosomal DNA. Such short-lived encounter complexes drive the formation of the catalytically productive LSD2/NPAC/nucleosome interactions that rely on the NPAC-favored capture of the H3 tail by LSD2. This mode of nucleosome recognition, together with the multimeric architecture of the complex, seems exquisitely suited for processive nucleosome modification, in association with the rapidly advancing RNA polymerase on gene bodies.

It is instructive to compare these mechanistic features with those described for LSD1, the other human flavin-dependent histone demethylase. The overall architecture of LSD1/CoREST is designed to stably clamp the nucleosomes at gene promoters through tight interactions with both the H3 tail and the DNA (Baron and Vellore, 2012; Kim et al., 2015; Pilotto et al., 2015). This nucleosome-docking mode of operation clearly contrasts with the tail-based mechanism of the multimeric and dynamic LSD2/NPAC complex. LSD1 can thereby function to stably induce repression at gene promoters, whereas LSD2 can more dynamically operate to support active gene transcription by the RNA polymerase. Given these contrasting features, it is fascinating that both flavin-dependent systems rely on structurally and mechanistically different auxiliary factors—CoREST and NPAC—to overcome the same inhibitory histone tail-sequestering effect by nucleosomal DNA. Despite their different mechanistic strategies, multidomain architectures of the auxiliary proteins, and opposing biological outcomes, these multiprotein histone demethylase complexes are tailored to license their enzymatic activities only on their appropriate target regions of the chromatin.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Protein expression
- METHOD DETAILS
  - Plasmids cloning and mutagenesis
  - Proteins purification
  - Enzyme activity assays
  - Fluorescence polarization
  - Analytical size-exclusion chromatography
  - Thermal-shift assays
  - HPLC-MS analysis
  - Peptide interaction assay (SPOT blot analysis)
  - Isothermal Titration Calorimetry (ITC)
  - Biolayer Interferometry (BLI)
  - SAXS
  - Cryo-EM sample preparation and data acquisition
  - Image processing
  - Model building and structure analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Activity assays
  - Fluorescence polarization
  - Thermal-shift assays
  - Domain prediction and sequence alignments
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.03.061.

ACKNOWLEDGMENTS

This work was supported by AIRC (IG-11342), MIUR (Flagship Epigen project and PRIN 20152TE5PK), and the Center for Cancer Research at NCI, NIH, Bethesda, MD. The high-resolution datasets were collected at eBIC (Didcot, UK) under proposal em16082 and supported by iNEXT project H2020 (grant 653706) under proposal 2099. The cryo-EM platforms of the Grenoble Instruct Instruct-CERIC Center (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) are supported by FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB) and the Rhône-Alpes Region, the Fondation Recherche Medicale (FRM), the fonds FEDER, and the GIS-Infrastrutures en Biologie Sante et Agronomie (IBISA). We acknowledge eBIC of Diamond Light Source (Didcot, UK) for access to cryo-EM microscopes and all staff members for their essential support and expertise during data collections. This work used the computational resources of the NIH High-Performance Computing (HPC) Biowulf cluster (http://hpc.nih.gov). We acknowledge ESRF (Grenoble, France) for provision of synchrotron radiation facilities and its staff during SAXS data collection. We thank Federica Corana (Centro Grandi Strumenti, Pavia) for support in mass spectrometry, Sergio Valente (Sapienza, University of Rome) for the alkylating reagents, René Bærentsen (Aarhus) for help in mutagenesis, and Sara Sandin (Nanyang Technological University, Singapore) for help in sample optimization and grid preparation. S. Pilotto was recipient of EMBO and EPIGEN short-term fellowships and was awarded with a fellowship from Fondazione Buzzati-Traverso.

AUTHOR CONTRIBUTIONS

All listed authors performed experiments and analyzed data. C.M., B.M., and S. Pilotto performed all biochemical assays. S. Pilotto prepared the cryo-EM samples and collected all datasets. S.S., D.R., F.F., and G.S. supported the cryo-EM experiments. S. Pilotto and S.C. performed image processing. C.M., S. Picaud, and P.F. carried out the experiments with the PWWP domain. S.M. and S. Pilotto performed the experiments on the NPAC dehydrogenase domain. B.M. and F.F. collected and analyzed SEC-SAXS data. C.M., B.M., S. Pilotto, and S.C. prepared figures. All authors contributed to the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 14, 2019
Revised: February 28, 2019
Accepted: March 15, 2019
Published: April 9, 2019

REFERENCES

Barad, B.A., Echols, N., Wang, R.Y.R., Cheng, Y., DiMaio, F., Adams, P.D., and Fraser, J.S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. Nat. Methods *12*, 943–946.

Baron, R., and Vellore, N.A. (2012). LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition. Proc. Natl. Acad. Sci. USA *109*, 12509–12514.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., et al.; The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *45* (*D1*), D158–D169.

Bilokapic, S., Strauss, M., and Halic, M. (2018). Histone octamer rearranges to adapt to DNA unwrapping. Nat. Struct. Mol. Biol. *25*, 101–108.

Brennich, M.E., Round, A.R., and Hutin, S. (2017). Online Size-exclusion and Ion-exchange Chromatography on a SAXS Beamline. J. Vis. Exp. *119*, 1–9.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chen, F., Yang, H., Dong, Z., Fang, J., Wang, P., Zhu, T., Gong, W., Fang, R., Shi, Y.G., Li, Z., and Xu, Y. (2013). Structural insight into substrate recognition by histone demethylase LSD2/KDM1b. Cell Res. *23*, 306–309.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature *461*, 415–418.

Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A., et al. (2016). ELM 2016—data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res. *44* (*D1*), D294–D300.

Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S., et al. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Res. *41*, 3924–3936.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fang, R., Barbera, A.J., Xu, Y., Rutenberg, M., Leonor, T., Bi, Q., Lan, F., Mei, P., Yuan, G.C., Lian, C., et al. (2010). Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. Mol. Cell *39*, 222–233.

Fang, R., Chen, F., Dong, Z., Hu, D., Barbera, A.J., Clark, E.A., Fang, J., Yang, Y., Mei, P., Rutenberg, M., et al. (2013). LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. Mol. Cell *49*, 558–570.

Fei, J., Ishii, H., Hoeksema, M.A., Meitinger, F., Kassavetis, G.A., Glass, C.K., Ren, B., and Kadonaga, J.T. (2018). NDF, a nucleosome-destabilizing factor that facilitates transcription through nucleosomes. Genes Dev. *32*, 682–694.

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*, 214–231.

Forneris, F., Binda, C., Vanoni, M.A., Mattevi, A., and Battaglioli, E. (2005). Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. FEBS Lett. *579*, 2203–2207.

Förster, S., Apostol, L., and Bras, W. (2010). Scatter: Software for the analysis of nano-and mesoscale small-angle scattering. J. Appl. Crystallogr. *43*, 639–646.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., and Svergun, D.I. (2017). *ATSAS 2.8*: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. *50*, 1212–1225.

Fu, J., Yang, Z., Wei, J., Han, J., and Gu, J. (2006). Nuclear protein NP60 regulates p38 MAPK activity. J. Cell Sci. *119*, 115–123.

Gatchalian, J., Wang, X., Ikebe, J., Cox, K.L., Tencer, A.H., Zhang, Y., Burge, N.L., Di, L., Gibson, M.D., Musselman, C.A., et al. (2017). Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. Nat. Commun. *8*, 1489.

Grant, T., Rohou, A., and Grigorieff, N. (2018). *cis*TEM, user-friendly software for single-particle image processing. eLife *7*, e35383.

Karytinos, A., Forneris, F., Profumo, A., Ciossani, G., Battaglioli, E., Binda, C., and Mattevi, A. (2009). A novel mammalian flavin-dependent histone demethylase. J. Biol. Chem. *284*, 17775–17782.

Kim, S.A., Chatterjee, N., Jennings, M.J., Bartholomew, B., and Tan, S. (2015). Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex. Nucleic Acids Res. *43*, 4868–4880.

Kim, J.A., Kwon, M., and Kim, J. (2019). Allosteric Regulation of Chromatin-Modifying Enzymes. Biochemistry *58*, 15–23.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2004). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.

Lambert, J.P., Picaud, S., Fujisawa, T., Hou, H., Savitsky, P., Uusküla-Reimand, L., Gupta, G.D., Abdouni, H., Lin, Z.Y., Tucholska, M., et al. (2019). Interactome Rewiring Following Pharmacological Targeting of BET Bromodomains. Mol. Cell *73*, 621–638.e17.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432–435.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N., and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. *43* (*W1*), W580–W584.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Lu, A.W., Wang, L., Chen, L., Hui, S., Rabinowitz, J.D., and Rabinowitz, J.D. (2018). Extraction and quantitation of Nicotinamide Adenine Dinucleotide Redox Cofactors. Antioxid. Redox Signal. *28*, 167–179.

Luger, K., Rechsteiner, T.J., Flaus, A.J., Waye, M.M.Y., and Richmond, T.J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. J. Mol. Biol. *272*, 301–311.

Marabelli, C., Marrocco, B., and Mattevi, A. (2016). The growing structural and functional complexity of the LSD1/KDM1A histone demethylase. Curr. Opin. Struct. Biol. *41*, 135–144.

McGinty, R.K., and Tan, S. (2015). Nucleosome structure and function. Chem. Rev. *115*, 2255–2273.

Morrison, E.A., Bowerman, S., Sylvers, K.L., Wereszczynski, J., and Musselman, C.A. (2018). The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. eLife *7*, e31481.

Nagaoka, K., Hino, S., Sakamoto, A., Anan, K., Takase, R., Umehara, T., Yokoyama, S., Sasaki, Y., and Nakao, M. (2015). Lysine-specific demethylase 2 suppresses lipid influx and metabolism in hepatic cells. Mol. Cell. Biol. *35*, 1068–1080.

Njau, R.K., Herndon, C.A., and Hawes, J.W. (2001). New developments in our understanding of the β-hydroxyacid dehydrogenases. Chem. Biol. Interact. *130–132*, 785–791.

Panjkovich, A., and Svergun, D.I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. Bioinformatics *34*, 1944–1946.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Picaud, S., and Filippakopoulos, P. (2015). SPOTing Acetyl-Lysine Dependent Interactions. Microarrays *3*, 370–388.

Pilotto, S., Speranzini, V., Tortorici, M., Durand, D., Fish, A., Valente, S., Forneris, F., Mai, A., Sixma, T.K., Vachette, P., and Mattevi, A. (2015). Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation. Proc. Natl. Acad. Sci. USA *112*, 2752–2757.

Qin, S., and Min, J. (2014). Structure and function of the nucleosome-binding PWWP domain. Trends Biochem. Sci. *39*, 536–547.

Ricketts, M.D., Han, J., Szurgot, M., and Marmorstein, R. (2019). Molecular basis for chromatin assembly and modification by multiprotein complexes. Protein Sci. *28*, 329–343.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320–W324.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Sankaran, S.M., Wilkinson, A.W., Elias, J.E., and Gozani, O. (2016). A PWWP Domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated lys-36 of histone H3 and regulates NSD2 function at chromatin. J. Biol. Chem. *291*, 8465–8474.

Scheres, S.H. (2012). A Bayesian view on cryo-EM structure determination. J. Mol. Biol. *415*, 406–418.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Stützer, A., Liokatis, S., Kiesel, A., Schwarzer, D., Sprangers, R., Söding, J., Selenko, P., and Fischle, W. (2016). Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. Mol. Cell *61*, 247–259.

van Nuland, R., van Schaik, F.M.A., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M., and van Ingen, H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. Epigenetics Chromatin *6*, 12.

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.

Weaver, T.M., Morrison, E.A., and Musselman, C.A. (2018). Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. Molecules *23*, E2614.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., et al. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature *508*, 263–268.

Zhang, L., Chooback, L., and Cook, P.F. (1999). Lysine 183 is the general base in the 6-phosphogluconate dehydrogenase-catalyzed reaction. Biochemistry *38*, 11231–11238.

Zhang, Q., Qi, S., Xu, M., Yu, L., Tao, Y., Deng, Z., Wu, W., Li, J., Chen, Z., and Wong, J. (2013). Structure-function analysis reveals a novel mechanism for regulation of histone demethylase LSD2/AOF1/KDM1b. Cell Res. *23*, 225–241.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332.

Zhou, K., Gaullier, G., and Luger, K. (2019). Nucleosome structure and dynamics are coming of age. Nat. Struct. Mol. Biol. *26*, 3–13.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, e42166.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
P. Pastoris KM71-H	Invitrogen (ThermoFisher)	Cat#C18200
BL21(DE3)RPplus	Novagen (Merck)	EMD_BIO-69450
BL21(DE3)pLysS	Invitrogen	Cat#C606003
Chemicals, Peptides, and Recombinant Proteins		
Fusion High-Fidelity PCR Master Mix	Invitrogen (Thermo Scientific)	Cat#F531S
In-Fusion HD Cloning Kit	Takara (Clontech)	Cat#638920
1-methyl-1-(prop-2-ynyl)aziridinium chloride	Pilotto et al., 2015	https://www.pnas.org/content/pnas/suppl/2015/02/10/1419468112.DCSupplemental/pnas.1419468112.sapp.pdf
Methanol	Panreac AppliChem	A3493,1000PE
Sodium phosphate monobasic dihydrate	Sigma-Aldrich	CAS: 13472-35-0
Sodium Chloride	Carlo Erba	A194090050
Glycerol	Merck	CAS: 56-81-5
Imidazole	Panreac AppliChem	A1378,0250
Deoxyribonuclease I from bovine pancreas	Sigma-Aldrich	CAS: 9003-98-9
PMSF	Sigma-Aldrich	CAS: 329-98-6
Zirconia beads	BioSpec	Cat. No. 11079107
Miracloth filter paper	Merck	475855
Tris	Panreac AppliChem	A3992,0250
Tris(2-carboxyethyl)phosphine hydrochloride	Sigma-Aldrich	CAS: 51805-45-9
Isopropyl-β-D-thiogalactoside	Sigma-Aldrich	CAS:367-93-1
HEPES	Sigma-Aldrich	CAS: 7365-45-9
Dithiothreitol	Sigma-Aldrich	CAS: 3483-12-3
Peroxidase from horseradish	Sigma Aldrich	CAS:9003-99-0; Cat#P6782
Peptides used in binding and enzymatic assays	China Peptides	http://www.chinapeptides.com/english/index.aspx
Ampliflu Red	Sigma Aldrich	CAS:119171-73-2; Cat#90101
SYPRO Orange	Invitrogen	Cat#S5692
H3 peptide used in ITC	TUFTS	http://tucf.tufts.edu/home/peptide_synthesis
Critical Commercial Assays		
SPOT-blot peptide interaction assay	Picaud and Filippakopoulos, 2015	10.3390/microarrays4030370
Whatman Chromatography paper Grade 1CHR	GE Healthcare Life Sciences	Cat#3001-878
Horseradish peroxidase-conjugated anti-His antibody	Novagen (Merck)	Cat#71841
Pierce ECL Western Blotting Substrate	Thermo Fisher Scientific	Cat#32106
SuperStreptavidin (SSA) Dip and Read Biosensors for kinetic	Forte´Bio	Cat#18-0011
Deposited Data		
147 bp 601 nucleosome (class1)	This paper	EMDataResource: EMD-4704 PDB: 6R1T
LSD2/NPAC-linker (214-225)/nucleosome (class 2)	This paper	EMDataResource: EMD-4705 PDB: 6R1U
LSD2/NPAC-linker (214-225)/nucleosome (class 3)	This paper	EMDataResource: EMD-4710 PDB: 6R25

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LSD2/NPAC-linker (214-225)/nucleosome (class 4)	This paper	EMDataResource: EMD-4711
147 bp 601 nucleosome (class 5)	This paper	EMDataResource: EMD-4712
LSD2/NPAC-linker (214-225)/H3 (1-26)	Fang et al., 2013	PDB: 4HSU
cytokine-like nuclear factor NPAC dehydrogenase (261-553)	http://www.rcsb.org	PDB: 2UYY
LSD2 Δ30 (31-822)	This paper	SASBDB: SASDFU3
nucleosome	This paper	SASBDB: SASDFX3
NPAC dehydrogenase (261-553)	This paper	SASBDB: SASDFV3
NPAC Δ205 (206-553)	This paper	SASBDB: SASDFW3
LSD2 Δ30 (31-822)/NPAC Δ205 (206-553)/ nucleosome	This paper	SASBDB: SASDFY3
Oligonucleotides		
TAMRA labeled 21-nt DNA	Microsynth AG	https://www.microsynth.ch/standard-synthesis.html
Recombinant DNA		
LSD2 and NPAC full-length human genes	GeneArt	https://www.thermofisher.com/us/en/home/life-science/cloning/gene-synthesis/gene-strings-dna-fragments.html
pJ902Express vector	DNA 2.0 (now ATUM)	https://www.atum.bio/catalog/expression-vectors/yeast#CytoplasmicExpression2
NPAC 261-553 in pNIC28-Bsa4 vector	SGC	https://www.thesgc.org/
Software and Algorithms		
Clustal Omega (EMBL-EBI)	Li et al., 2015	https://www.ebi.ac.uk/Tools/msa/clustalo/
ESPript 3	Robert and Gouet, 2014	http://espript.ibcp.fr/ESPript/ESPript/
ELM	Dinkel et al., 2016	http://elm.eu.org/
Uniprot	Bateman et al., 2017	https://www.uniprot.org/
GraphPad Prismversion 6.00 for Windows	GraphPad	https://www.graphpad.com/
MicroCal Origin	Origin	https://www.originlab.com/
Forté Bio analysis software V.9.0.014	Forté Bio	www.fortebio.com
CHROMIX	Panjkovich and Svergun, 2018	https://www.embl-hamburg.de/biosaxs/chromixs.html
ATSAS package	Franke et al., 2017	https://www.embl-hamburg.de/biosaxs/software.html
Scatter	Förster et al., 2010	http://www.bioisis.net/tutorial/9
PRIMUS	Konarev et al., 2004	https://www.embl-hamburg.de/biosaxs/primus.html
cisTEM	Grant et al., 2018	https://cistem.org/
MotionCor2	Zheng et al., 2017	http://msg.ucsf.edu/em/software/motioncor2.html
Ctffind	Rohou and Grigorieff, 2015	http://grigoriefflab.janelia.org/ctf
RELION	Scheres, 2012	https://www3.mrc-lmb.cam.ac.uk/relion/index.php?title=Main_Page
Coot	Emsley and Cowtan, 2004	http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/
Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/
Other		
Carbon grids for cryo-EM	Quantifoil	https://www.emsdiasum.com/microscopy/products/grids/quantifoil.aspx

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Andrea Mattevi (andrea.mattevi@unipv.it).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Protein expression

All histone proteins were expressed in BL21(DE3) *pLysS E.coli* cells according to established protocols (Luger et al., 1997). LSD2 full-length proteins were expressed in *E. coli* BL21(DE3). Cells were grown in LB medium supplemented with 100 μg/ml ampicillin under shaking conditions till $O.D._{600nm}$ = 0.8. Induction was performed with 0.5 mM IPTG and 0.1 mM $ZnSO_4$ at 17°C for 18 hours. Cells were collected by centrifugation at 5000 rpm for 10 minutes and stored at −20°C. LSD2Δ30-expressing *P. pastoris* cells were grown in flasks at 30°C under shaking conditions for 72 hours and induced with methanol for 48 hours. Cells were then collected by centrifugation at 4000 rpm for 10 minutes and stored at −20°C. The PWWP domain (residues 1-105) was expressed in *E. coli* BL21(DE3). Cells were grown in 2xYT medium supplemented with 100 μg/ml kanamycin and 1% (w/v) glucose at 37°C until $O.D._{600}$ reached 1. Induction was performed with 0.1 mM IPTG for 16 h at 17°C. NPACΔ205 (linker + dehydrogenase) was expressed in *E. coli* BL21(DE3) RPplus, and cells were induced by autoinducing protocol. Cells were collected by centrifugation at 5000 rpm for 10 minutes and stored at −20°C. NPAC dehydrogenase proteins were expressed in *E. coli* BL21(DE3) RPplus. Cells were grown in TB medium at 37°C under shaking conditions till an $O.D._{600nm}$ of 0.8. Induction was performed with 0.5 mM IPTG at 17°C overnight. Cells were collected by centrifugation at 5000 rpm for 10 minutes and stored at −20°C.

METHOD DETAILS

Plasmids cloning and mutagenesis

All cloning and mutagenesis reactions were performed combining the Phusion High-Fidelity PCR Master Mix (Thermo Scientific) with the In-Fusion HD Cloning Kit (Clontech) protocols. DNA sequences were derived from the 601 Widom sequence (Lowary and Widom, 1998). A pMA vector containing 15 copies of the 147 bp 601 Widom sequence was purchased from GeneArt. DNAs were amplified in *E. coli* DH5α, purified and used for nucleosome reconstitution (Luger et al., 1997). The pET3 vectors (Ampicillin resistance) containing the sequences of the four *Xenopus laevis* histone proteins, were a kind gift from Dr. Toshiya Senda (Biomedicinal Information research Center, National Institute of Industrial Science and Technology, Tokyo). H3 K4C-K23M-K27M-C110A was obtained by two consecutive mutagenesis reactions starting from the H3 K4C-C110A already available in house. The sequence coding for human LSD2Δ30 (residues 31-822) was cloned from the full-length human gene (purchased from GeneArt) into a pJ902Express vector (DNA 2.0) with a C-terminal eGFP-His_8 tag fused to a cleavable Prescission-protease cleavage site. The vector was linearized with PmeI and inserted in the *P. pastoris* KM71-H strain (Invitrogen) genome through electroporation. The sequence coding for the full-length human LSD2 was cloned in a pGEX-6P-1 vector with a GST N-terminal tag fused to a cleavable Prescission-protease cleavage site. All LSD2 mutants were obtained from full length sequence (Figure 4). The gene for human NPAC was purchased from GeneArt and cloned in pET24a, bearing an N-terminal Flag-His_8-SUMO tag followed by Prescission protease recognition site. NPACΔ205 (residues 205-553) and NPAC-PWWP (residues 1-105) were obtained from the full-length construct. The pNIC28-Bsa4 vector carrying the N-terminal His_6 tagged NPAC dehydrogenase (residues 261-553) was a kind gift from Structural Genomics Consortium (SGC) of the University of Oxford (UK).

Proteins purification

All purifications were carried out at 4°C, except for the histones, DNA sequences, and NPAC dehydrogenase performed at 20°C. Chromatographic steps were performed on ÄKTA systems (GE Healthcare).

Recombinant nucleosomes were prepared according to established protocols (Luger et al., 1997). H3 mutants K4C-C110A and K4C-K23M-K27M-C110A were alkylated with 1-methyl-1-(prop-2-ynyl)aziridinium chloride as previously described (Pilotto et al., 2015).

LSD2Δ30-expressing *P. pastoris* cells were re-suspended in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 25 mM imidazole, 2 μg/ml DNase, and 1 mM PMSF. Zirconia beads (BioSpec products) were used to lyse the cell in a bead-beater. Then beads were removed with a Miracloth filter paper (Merck) and the cell extract was centrifuged at 70,000 g for 30 min. The supernatant was loaded onto a His-Trap column (GE Healthcare). After a wash with 40 mM imidazole, the elution was obtained with 250 mM imidazole. The sample was supplemented with Prescission protease and dialyzed overnight in the same lysis buffer without imidazole. The day after a second His-Trap column step was used to purify the tag-free protein at 40 mM imidazole. The sample was then gel filtered through a Superdex 200 10/300 (GE Healthcare) equilibrated in 20 mM Tris pH 8 (25°C), 100 mM NaCl, 1 mM TCEP.

LSD2 expressing *E. coli* BL21(DE3) cell pellets were re-suspended in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 25 mM imidazole, 2% (w/v) lysozyme, and 1 mM PMSF. Cell suspension was sonicated and centrifuged at 70,000 g for 30 min. The supernatant was loaded onto a GST-Trap column (GE Healthcare) and the tagged-protein eluted with 50 mM GSH. His-Prescission protease was added (1 mg/15 mL of sample). The sample was dialyzed overnight in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl and

5% glycerol. After dialysis, the sample was loaded onto a second GST-Trap to remove the GST-tag. The unbound fractions were then loaded onto a His-Trap column to remove the His-tagged protease. A final size-exclusion chromatography step was performed on a Superdex 200 10/300 equilibrated in 20 mM Tris pH 8 (25°C), 200 mM NaCl, 1 mM TCEP.

The PWWP domain (residues 1-105) expressing cells were resuspended in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 1 mM PMSF and sonicated. The suspension was centrifuged at 56,000 g for 1 h and the cell extract was loaded onto a His-Trap column. The resin was washed first with 50 mM NaH_2PO_4 pH 8.0, 1 M KCl, 5% (v/v) glycerol and then with 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 50 mM imidazole. Elution was carried out in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 250 mM imidazole. His-tagged Prescission protease was added and the sample was dialyzed overnight in resuspension buffer. The sample was then passed a second time through a His-Trap column, and the flow-through was collected. Finally, a Superdex 75 10/300 chromatographic step was carried out in 20 mM PIPES pH 6.5, 200 mM NaCl.

NPAC dehydrogenase cell pellet was resuspended in 50 mM HEPES pH 7.5, 500 mM NaCl, 40 mM imidazole, 5% (v/v) glycerol, supplemented with 1 mM PMSF. The suspension was lysed by sonication and centrifuged at 70,000 g for 35 min. Cell extract was loaded onto a 5 mL HisTrap column, washed and eluted in 50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole. The sample step was supplemented with 6xHis-TEV protease and dialyzed overnight, against 2 L of 50 mM HEPES pH 7.5, 500 mM NaCl, 5% (v/v) glycerol. A second His-Trap column was used to remove the His-tag and TEV protease. The flow-through was loaded onto a Superdex 200 16/60 (GE Healthcare) in 50 mM HEPES pH 7.5, 500 mM NaCl, 5% (v/v) glycerol.

NPACΔ205 (linker + dehydrogenase) cell pellet was resuspended in 50 mM HEPES pH 7.5, 5% (v/v) glycerol, 500 mM NaCl, 40 mM imidazole, 1 mM PMSF before cell disruption by sonication and successive centrifugation at 70,000 g for 40 min. The supernatant was loaded onto a HisTrap column and the protein was eluted by adding 250 mM imidazole. Incubation with His-tagged SUMO protease was carried out overnight in dialysis against 50 mM HEPES pH 7.5, 5% (v/v) glycerol, 500 mM NaCl. Both tag and protease were removed through a His-Trap column. The flow through was loaded on a Superdex 200 10/300 in 50 mM HEPES pH 7.5, 5% (v/v) glycerol, 500 mM NaCl, 0.5 mM dithiothreitol.

Enzyme activity assays

Peptides were purchased from ChinaPeptides. Activity measurements were performed with peroxidase-coupled assays on a Clariostar plate reader (BMG Labtech). The reactions were carried out in 50 mM HEPES pH 8.5, 0.1 mM Amplex Red, 0.3 mM horseradish peroxidase, 0.3 μM LSD2Δ30, with and without 3 μM NPAC-linker at different ionic strengths (0 mM NaCl, 100 mM NaCl, 150 mM NaCl, 200 mM NaCl). H3K4me1 and H3K4me2 H3 N-terminal peptides of different lengths (1-21, 1-30, and 1-40), and H3K4me2 K23M K27M 1-40 peptide were serially diluted. Measured fluorescence signal reflects the enzymatic conversion of Amplex Red to resorufin, as described previously (Pilotto et al., 2015). Initial velocity values were fitted to the Michaelis-Menten equation using GraphPad Prism version 6.00 for Windows, GraphPad Software, San Diego California USA, (https://www.graphpad.com/). Inhibition assays were performed at 100 mM NaCl using $H3_{1\text{-}40}$ K4me2 substrate and 100 μM putative peptide inhibitor: $H2A_{1\text{-}19}$, $H2A_{117\text{-}129}$, $H2B_{4\text{-}19}$, $H3_{16\text{-}40}$, $H3_{21\text{-}40}$, $H4_{1\text{-}30}$.

Fluorescence polarization

We measured the change in polarization of a TAMRA-labeled 21-bp DNA (5′-AGTCGCCAGGAACCAGTGTCA-3′) through a Clariostar plate reader (BMG Labtech), as previously described (Pilotto et al., 2015). Experiments were carried out in 15 mM Tris pH 8, 0.01% (v/v) Tween 20 with 0-150 mM NaCl and 5 nM DNA. Peptide binding was measured as described (Pilotto et al., 2015).

Analytical size-exclusion chromatography

10 μM (final concentration) semisynthetic NCPs were incubated with 20 μM LSD2Δ30 and 100 μM NPAC-linker (1:2:10 molar ratio) for an hour on ice in 20 mM Tris pH 7.5, 100 mM NaCl. The mixture was loaded onto silica gel columns WTC-030N5 or WTC-030S5 (Wyatt Technology) equilibrated in 15 mM HEPES pH 7.3 (25°C), 200 mM NaCl. The elution profile was recorded at 260, 280, and 400 nm.

LSD2Δ30 was incubated with NPACΔ205 and semisynthetic nucleosomes for one hour in 20 mM Tris pH 7.5, 1 mM EDTA, 1 mM dithiothreitol, 200 mM NaCl. The mixture was loaded onto silica gel columns WTC-030N5 or WTC-030S5 (Wyatt Technology) equilibrated in 15 mM HEPES pH 7.3 (25°C), 200 mM NaCl. The elution profile was recorded at 214, 260, and 280 nm.

Analytical-SEC experiments of NPAC dehydrogenase were performed with Superdex 200 5/150 (Ge Healthcare) on ÄKTAmicro system (Ge Healthcare) in buffer 50 mM HEPES pH 7.5, 100 mM NaCl, 5% (v/v) glycerol. The elution profile was recorded at 214, 260, and 280 nm.

Thermal-shift assays

Thermofluor (Biorad) experiments were carried out in duplicates in 20 μL wells containing 25 μM NPAC dehydrogenase and SYPRO Orange (Invitrogen) 5X in 50 mM HEPES pH 7.5, 500 mM NaCl, 5% (v/v) glycerol without or with 50 μM NADH or NADPH.

HPLC-MS analysis

To avoid contamination by HEPES and reduce salt content, NPAC dehydrogenase protein was first passed through a Hi-Trap Desalting column (GE Healthcare) at 20°C in 200 mM NaCl, 25 mM Tris pH 7.5. Following a published protocol (Lu et al., 2018), resulting average micrographs were processed in the framework of RELION 2.1 and 3.0 (Scheres, 2012) as specified. The integrated images were used for CTF estimation with CTFFIND4 (Rohou and Grigorieff, 2015) followed by automated particle picking using a Gaussian blob as reference within the RELION 2.1 workflow. A total of 490,558 particles were extracted from 1906 selected micrographs with a box size of 300 × 300 pixels. The extracted particles were subjected to multiple rounds of 2D classifications and low-population or poorly-defined classes were discarded to remove junk or inconsistent particles. In parallel, an initial ab-initio density map was separately produced by processing the same dataset using cisTEM (Grant et al., 2018). This map was used as a reference, low-pass filtered to 40 Å, for subsequent image processing steps carried out in RELION (Zivanov et al., 2018). The 400,624 particles remaining after 2D classification were subjected to multiple rounds of 3D classifications in RELION 2.1 to yield five distinct classes with resolutions ranging from ~4 – ~8 Å. 3D refinements followed by local resolution filtering were performed using RELION 3.0 (Zivanov et al., 2018) to produce the final maps.

Model building and structure analysis

An initial model for the density map corresponding to the nucleosome core particle (class 1) was generated using a rigid body fit of a crystal structure (PDB: 6ESF) in UCSF Chimera (Pettersen et al., 2004). The fitted model was further improved by manual rebuilding in COOT (Emsley and Cowtan, 2004). For the LSD2-nucleosome complex maps (classes 2 and 3), the crystal structure corresponding to the nucleosome core particle (PDB: 6ESF) was fitted followed by rigid-body docking of the crystal structure of LSD2/NPAC-linker (PDB: 4HSU) into the additional density using UCSF Chimera (Pettersen et al., 2004). Like the nucleosome core particle, the combined models of the LSD2-nucleosome complex were also subjected to manual rebuilding in COOT (Emsley and Cowtan, 2004). The final models were validated using Molprobity (Chen et al., 2010) and EMRinger (Barad et al., 2015).

QUANTIFICATION AND STATISTICAL ANALYSIS

Activity assays

Initial velocity values were fitted to the Michaelis-Menten equation using GraphPad Prism version 6.00 for Windows, GraphPad Software, San Diego California USA, (https://www.graphpad.com/).

Fluorescence polarization

Experiments were carried out in triplicates. K_d values were determined fitting the data to the Hill equation with GraphPad Prism version 6.00 for Windows, GraphPad Software, San Diego California USA, (https://www.graphpad.com/).

Thermal-shift assays

Thermofluor (Biorad) experiments were carried out in triplicates. All data were analyzed using Microsoft Excel.

Domain prediction and sequence alignments

Alignments were performed with Clustal Omega tool of EMBL-EBI web platform (Li et al., 2015), and ESPrit 3 (Robert and Gouet, 2014). Sequence-based predictions were performed with ELM online software (Dinkel et al., 2016). Homologous sequences for both LSD2 and NPAC were retrieved from Uniprot database (Bateman et al., 2017). Twenty representative sequences of GLYR1/NPAC were selected among mammals (*Ailuropoda melanoleuca, Bos taurus, Canis lupus familiaris, Equus caballus, Homo sapiens, Loxodonta africana, Mus musculus, Oryctolagus cuniculus*), amphibians (*Xenopus tropicalis*), reptiles (*Alligator mississippensis, Pelodiscus sinensis*), birds (*Gallus gallus, Taeniopygia guttata*), fishes (*Danio rerio, Takifugu rubripes*), insects

Biolayer Interferometry (BLI)

Experiments were performed on an Octet RED384 system (FortéBio) at 25°C in 25 mM HEPES, pH 7.5, 100 mM NaCl and 0.5 mM TCEP and 0.01% TWEEN-20 using the FortéBio data acquisition software V.9.0.049. Biotinylated peptides (Alta Biosciences, Histone Set 1-H3 library, F9: $H3_{28\text{-}48}$ and G1: $H3_{28\text{-}48}$ K36me3) were first immobilized onto Super Streptavidin biosensors (SuperStreptavidin (SSA) Dip and Read Biosensors for kinetic #18-0011, FortéBio), pre-equilibrated in the BLI buffer then quenched in a solution of 5 μM Biotin: baseline equilibration 60 s, peptide loading for 240 s, quenching for 60 s, 1000 x rpm shake speed, at 25°C. The immobilized peptides were subsequently used in association and dissociation measurements performed within a time window of 600 s (base line equilibration 120 s, association for 240 s, dissociation for 240 s, 1000 x rpm shake speed, at 25°C). PWWP concentrations (0.082, 0.24, 0.74, 2.22, 6.66, 20, and 60 μM) were used in order to determine binding constants. Interference patterns from peptide-coated biosensors without protein were used as controls. After referencing corrections, the subtracted binding interference data were analyzed using the FortéBio analysis software (FortéBio data analysis software V.9.0.014) provided with the instrument following the manufacturer's protocols.

SAXS

SEC-SAXS experiments were performed at BM29 beamline of ESRF (Grenoble, France). All samples were analyzed using Nexera High Pressure Liquid Chromatography (HPLC; Shimadzu) system connected online to SAXS sample capillary (Brennich et al., 2017). About 0.2-0.3 mg of sample were injected, at the flow rate of 0.25 mL/min, onto a WTC-030N5 column (Wyatt Technology) pre-equilibrated with 15 mM HEPES pH 7.3, 200 mM NaCl at 20°C. Scattering data were recorded using per sec frame rate on Pilatus 1 M detector located at a fixed distance of 2.87 m from the sample, allowing a global q range of 0.03–4.5 nm with a wavelength of 0.01 nm. Data analysis of collected frames, including blank subtraction and averaging, was carried out using Chromix (Panjkovich and Svergun, 2018). Radii of gyration (Rg), molar mass estimates and distance distribution functions P(r) were computed using both the ATSAS package (Franke et al., 2017; Konarev et al., 2004) and Scatter (Förster et al., 2010), yielding the same values within the experimental errors (Table S3).

(*Ceratitis capitate, Noplophora glabripennis*), plants (*Anthurium aminicola, Arabidopsis thaliana*), and chromalveolata (*Symbiodinium microadriaticum*).

DATA AND SOFTWARE AVAILABILITY

The cryo-EM maps and atomic coordinates have been deposited in the Electron Microscopy Data Bank and Protein Data Bank under accession codes EMDataResource: EMD-4704 PDB: 6R1T (class 1), EMDataResource: EMD-4705 PDB: 6R1U (class 2), EMDataResource: EMD-4710 PDB: 6R25 (class 3), EMDataResource: EMD-4711 (class 4), and EMDataResource: EMD-4712 (class 5). SEC-SAXS experimental data and ab-initio model have been deposited in Small Angle Scattering Biological Data Bank with accession codes SASBDB: ASDFU3 (LSD2 Δ30 (31-822)), SASBDB: SASDFX3 (nucleosome), SASBDB: SASDFV3 (NPAC dehydrogenase (261-553)), SASBDB: SASDFW3 (NPAC Δ205 (206-553)), SASBDB: SASDFY3 (LSD2 (31-822)/NPAC Δ205 (206-553)/ nucleosome).

Cell Reports, Volume *27*

Supplemental Information

A Tail-Based Mechanism Drives Nucleosome Demethylation by the LSD2/NPAC Multimeric Complex

Chiara Marabelli, Biagina Marrocco, Simona Pilotto, Sagar Chittori, Sarah Picaud, Sara Marchese, Giuseppe Ciossani, Federico Forneris, Panagis Filippakopoulos, Guy Schoehn, Daniela Rhodes, Sriram Subramaniam, and Andrea Mattevi

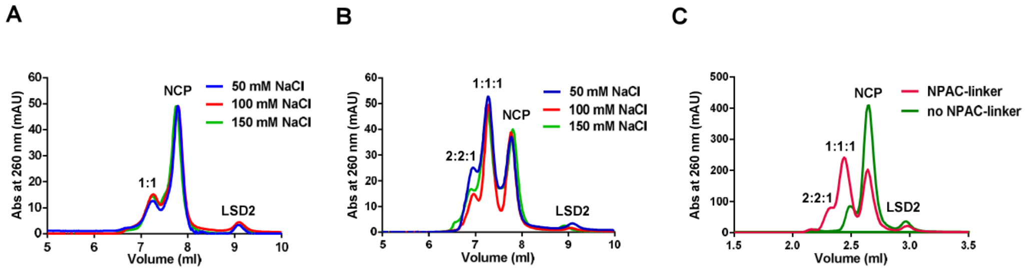


Figure S1 related to Figure 1. Formation of the complex between LSD2 and the nucleosome core particle is dependent on NPAC-linker and independent on ionic strength. A-B) The ionic strength does not affect the efficiency of LSD2/NPAC-linker/NCP complex formation, either in absence (**A**) or presence (**B**) of the NPAC-linker (WTC-030S5 column). **C)** Elution profiles of a semi-synthetic NCP carrying a double mutation on H3 (K23M K27M), incubated with LSD2 with and without NPAC-linker (WTC-030N5 column). Protein concentrations as in the main Figure 1C.

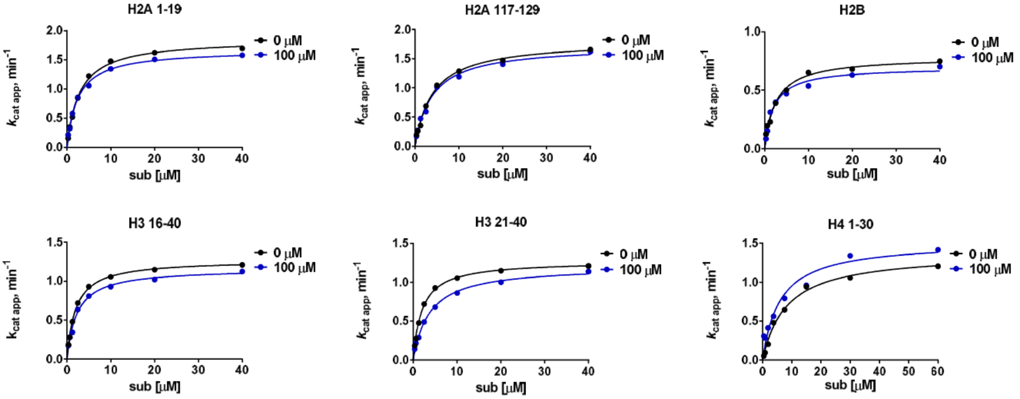


Figure S2 related to Table 1. Histone tails do not interfere with H3-K4 demethylase activity of LSD2. Apparent k_{cat} values measured in absence and presence (100 μM) of the histone tail peptides listed above each plot. Very low or no inhibition was detected. The data were fit to the Michaelis-Menten curve. The experiments were performed in 50 mM HEPES pH 8.5, 100 mM NaCl, 0.3 μM LSD2, 3 μM NPAC-linker, 0 - 40 μM H3K4me2$_{1-40}$ peptide substrate.

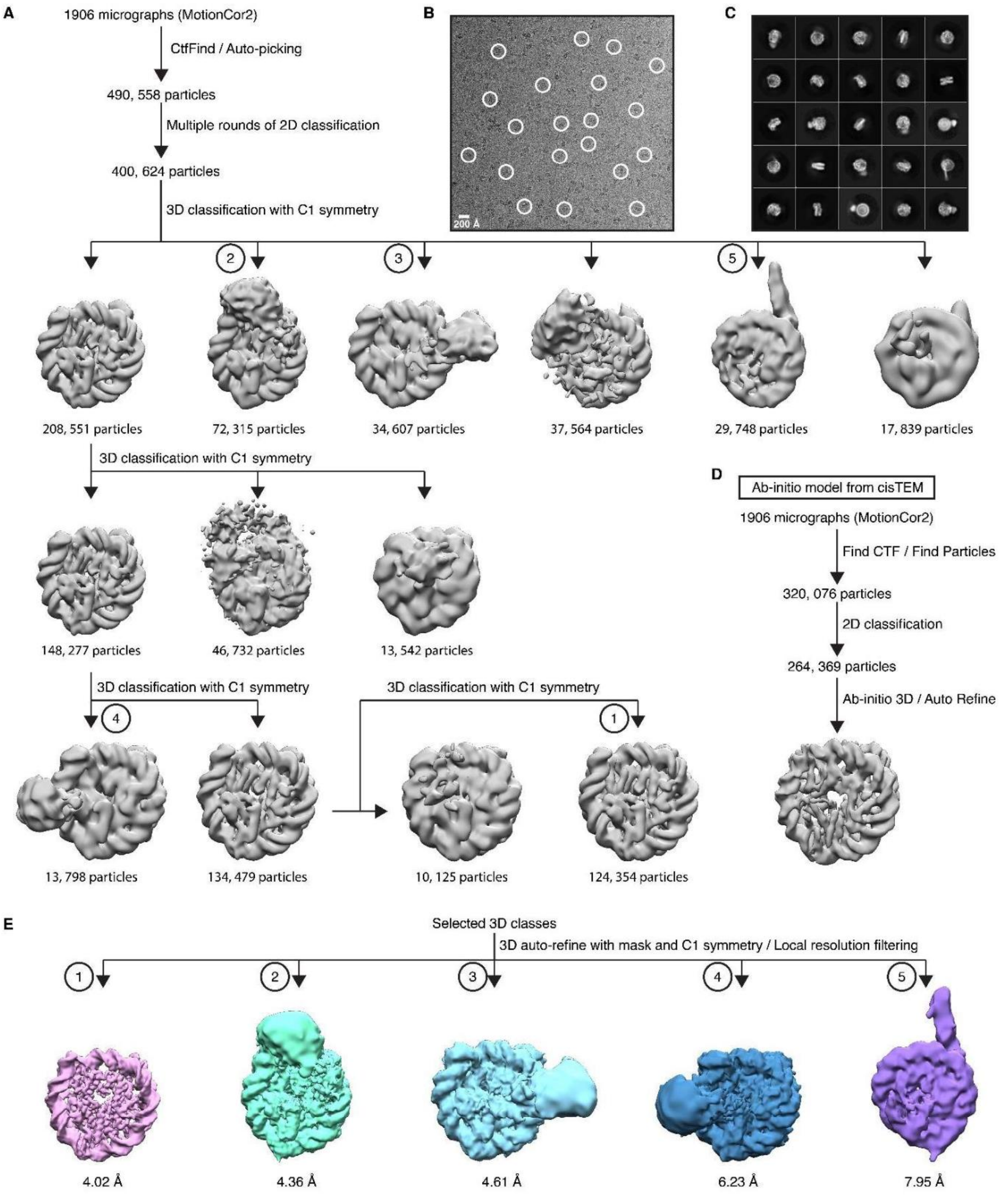


Figure S3 related to Figure 2. Processing of cryo-EM micrographs. A) Processing work-flow. **B)** A micrograph used for data processing showing the sample embedded in ice (scale bar: 200 Å). Examples of representative particles are outlined by white circles. **C)** Representative 2D class averages. **D)** *Ab-initio* determination of NCP density map using cisTEM**. E)** The five final classes used for structural analysis.

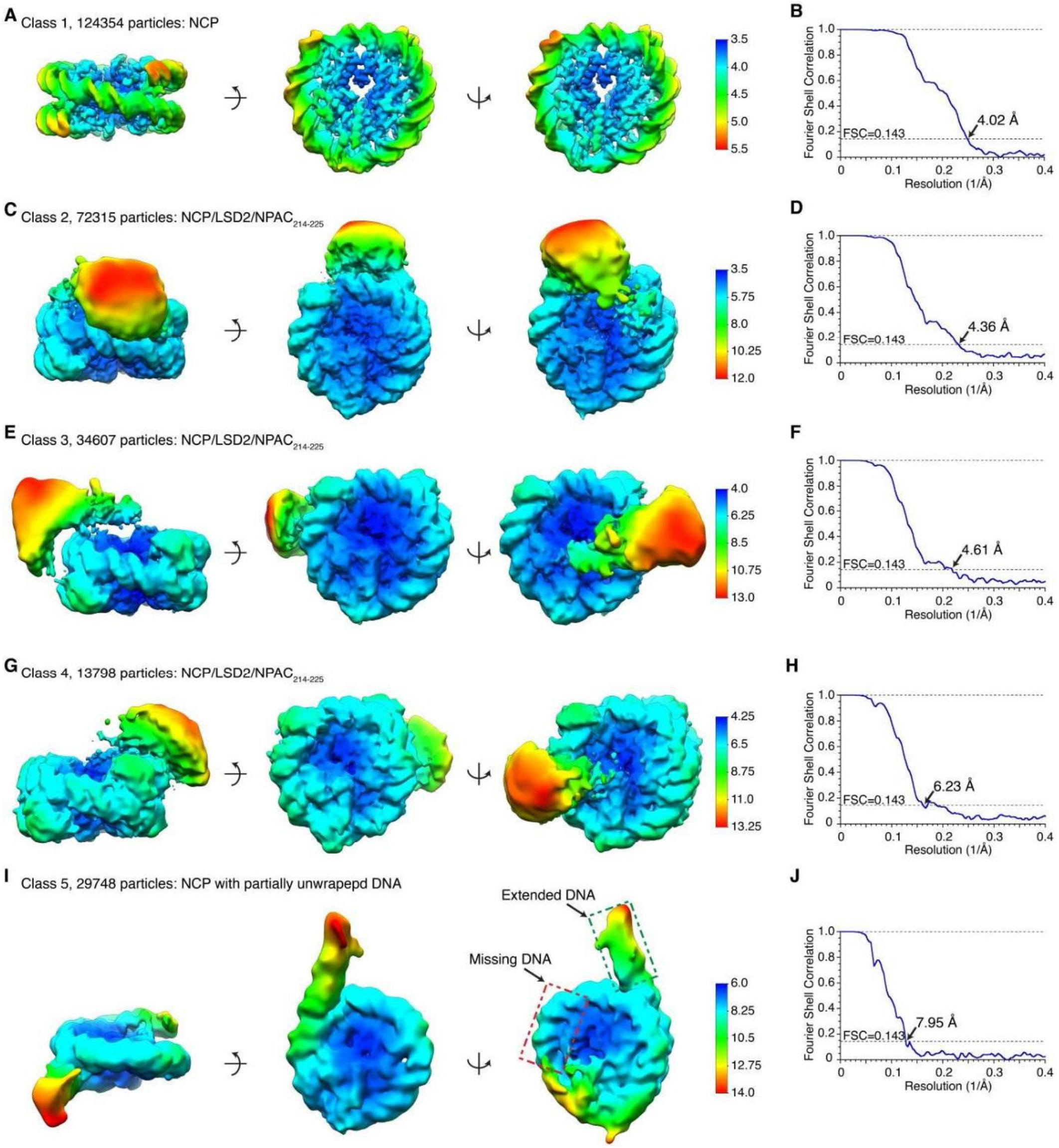


Figure S4 related to Figure 2. The five classes revealed by cryo-EM data processing. The electron density map of each class is shown in three orientations using a colour gradient for the resolution. The corresponding Fourier-Shell-Correlation (FSC) curve is reported on the right. Class 1 is the isolated nucleosome **(A-B),** whereas classes 2, 3, and 4 are the LSD2/NPAC-linker/NCP complexes, in which LSD2 adopts different binding modes **(C-D, E-F, G-H)**. Class 5 is an unfolded NCP: DNA is sliding away from the core, now composed by an H3-H4 tetramer and only one H2A-H2B dimer (**I-J**). The map for class 1 is depicted at a higher counter-level for illustration of high-resolution features of the histone core.

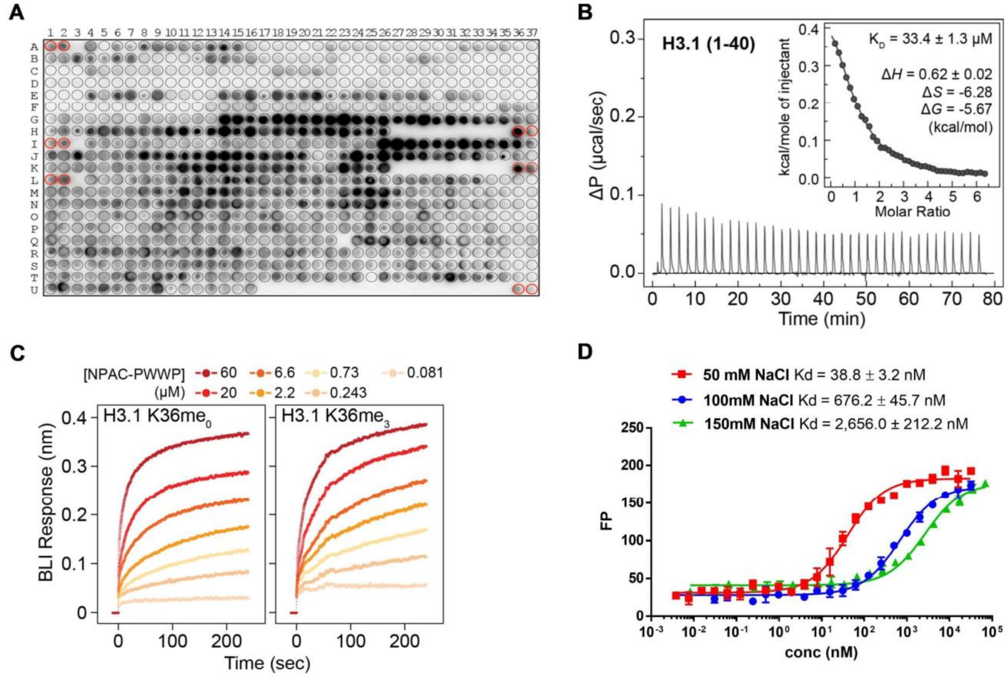


Figure S5 related to Figure 5. NPAC-PWWP is a chromatin-binding module. A) SPOT-array technology probes histone-peptide binding to the NPAC-PWWP domain (residues 1-105). Cellulose membrane was spotted with an array of 20-aminoacid long peptides covering the sequences of the H3.1, H3.3, and H4 N-terminal tails with single and multiple epigenetic modifications (Table S2). Control peptides are shown in red circles. Tightest binding was consistently observed with peptides covering residues 26-45 of H3. **B)** Isothermal calorimetry experiment for H3-tail binding to PWWP. The experiment was carried out using the PWWP protein inside the sample cell and the naked histone H3 peptide (residues 1-40) in the syringe. **C)** Bio-Layer Interferometry shows that Lys36 methylation state does not affect association of $H3_{28\text{-}48}$ to NPAC-PWWP. **D)** Fluorescence polarization binding assay using a 21 bp DNA sequence labelled with TAMRA fluorescent dye.

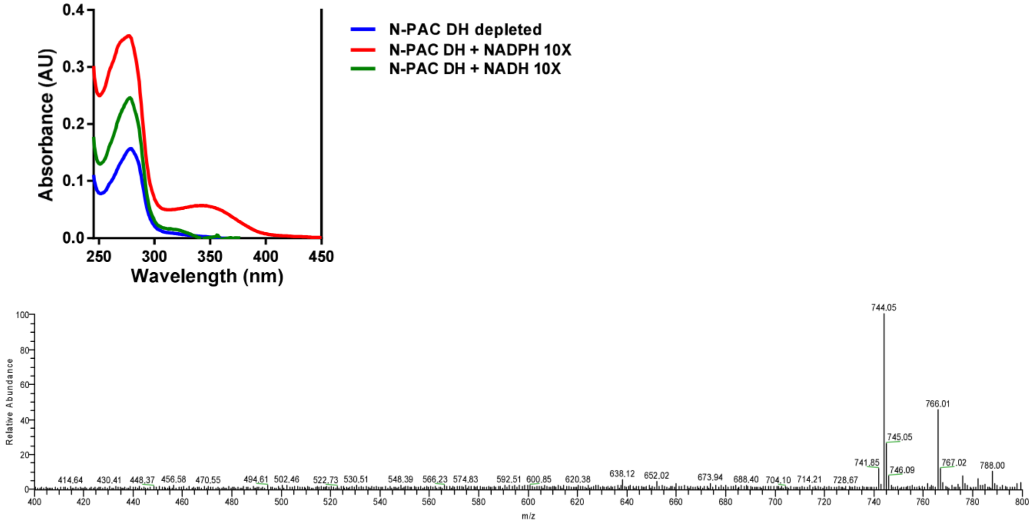


Figure S6 related to Figure 5. NPAC dehydrogenase domain binds NADPH. Top: NPAC dehydrogenase domain (residues 261-553) was dialysed against a solution consisting of 4 M KBr, 100 mM K_2HPO_4 pH 6.5, and activated charcoal to remove the natural cofactor (blue curve). The dialyzed sample was then desalted and split into two fractions which were incubated with NADH (green curve) and NADPH (red curve), respectively. Each sample was then loaded on a desalting column. The Uv/Vis absorbance spectrum confirms that NPAC dehydrogenase retains NADPH but not NADH as shown from the absorbance peak of the reduced nicotinamide at 348 nm. Bottom: cofactor identification by mass-spec. The cofactor was extracted through denaturation of the recombinant NPAC dehydrogenase domain, purified, and identified by HPLC coupled with a mass spectrometer. The spectrum is fully consistent with NADPH, with no contamination by $NADP^+$ or NAD^+ (742 and 663 Da, respectively).

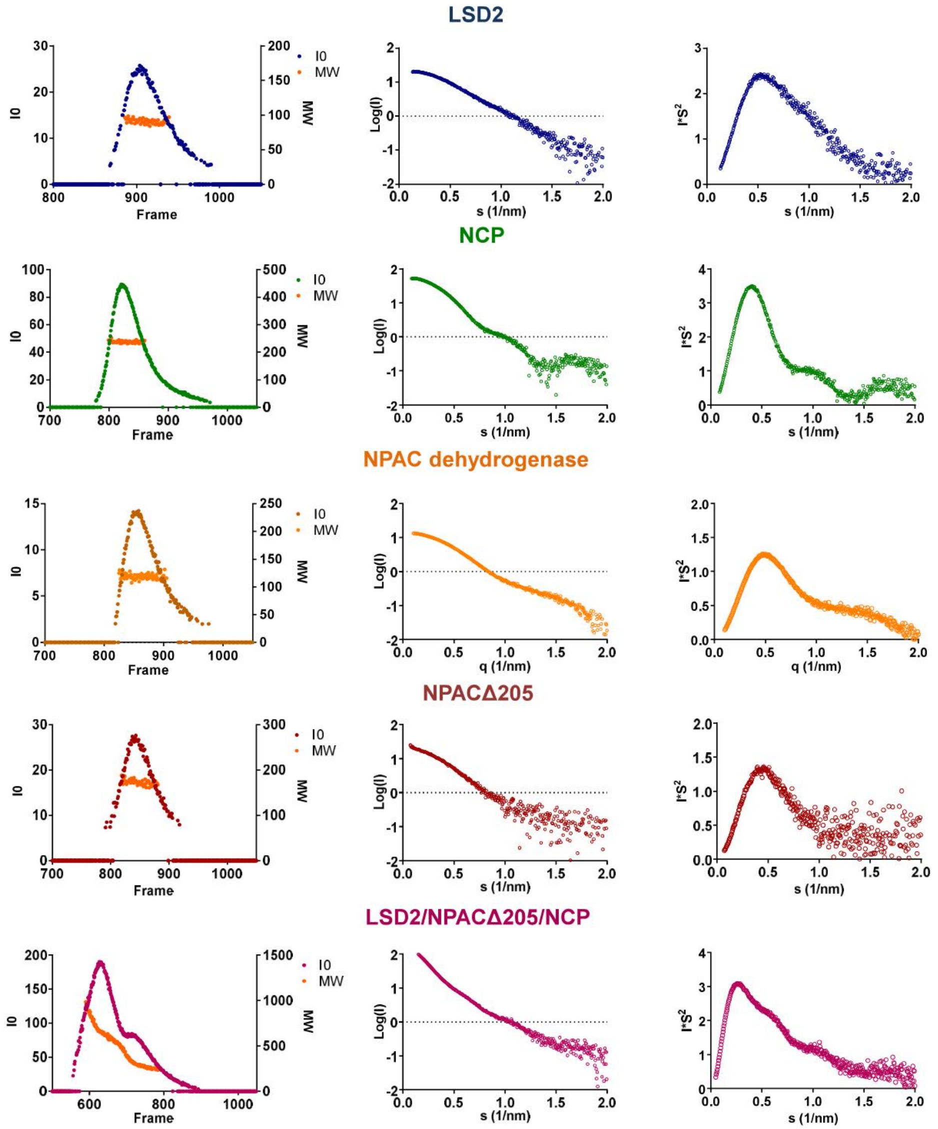


Figure S7 related to Figure 6. SAXS analysis of LSD2, nucleosomes, NPAC dehydrogenase, NPACΔ205 and LSD2/NPACΔ205/NCP complex. For each sample: the left panel shows the intensity chromatogram with the distribution of molecular weight (orange) along the elution peak, the central panel shows the average scattering plot, the right panel shows the Kratky plot. The average scattering curve and the Kratky plot for the LSD2/NPACΔ205/NCP refer to the region B-C of the chromatogram shown in the main Figure 6A, corresponding to the highest peak of the intensity chromatogram. The structural parameters derived from Guinier and P(r) analyses are reported in Table S3.

Table S1 related to Figure 4. DNA binding to LSD2[a]

	K_d (nM)	
	No NaCl	100 mM NaCl
LSD2	47.71 ± 2.66	> 10^4
LSD2+NPAC-linker[b]	30.71 ± 2.43	> 10^4
LSD2 K104E	114.5 ± 11.2	> 10^4
LSD2 K114E	52.12 ± 7.37	> 10^4
LSD2 K115E	36.02 ± 8.62	> 10^4
LSD2 K122E	41.58 ± 4.89	> 10^4
LSD2 R302D	41.89 ± 4.27	> 10^4
LSD2 K481E R482D	110.3 ± 19.92	> 10^4
LSD2 Δ241-258	26.49 ± 3.05	> 10^4
LSD2 Δ30[c]	649.4 ± 77.52	No binding
LSD2 Δ30+NPAC-linker	1284 ± 267.2	No binding

Values are means of three experiments ± standard error mean.

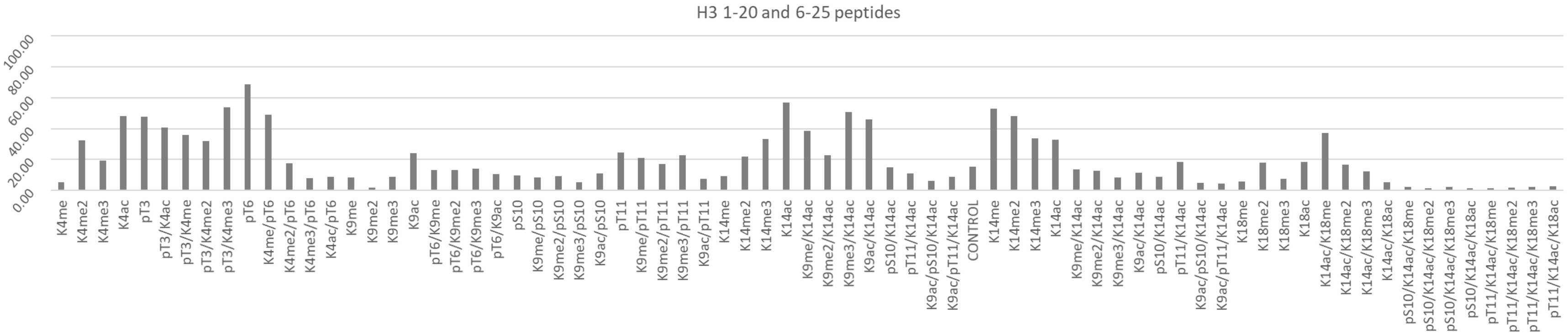
[a] K_d values were measured using double stranded DNA with the sequence 5'-AGTCGCCAGGAACCAGTGTCA-3' labelled with TAMRA at 5'-end.

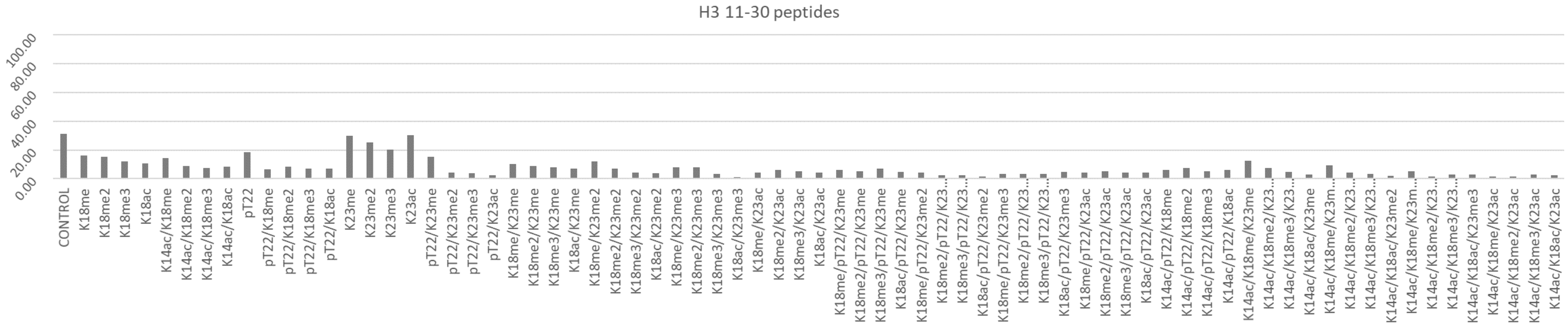
[b] NPAC-linker comprises residues 214-225 (Figure 1).

[c] LSD2 Δ30 lacks the first 30 residues (ATPRGRTKKKASFDHSPDSLPLRSSGRQ).

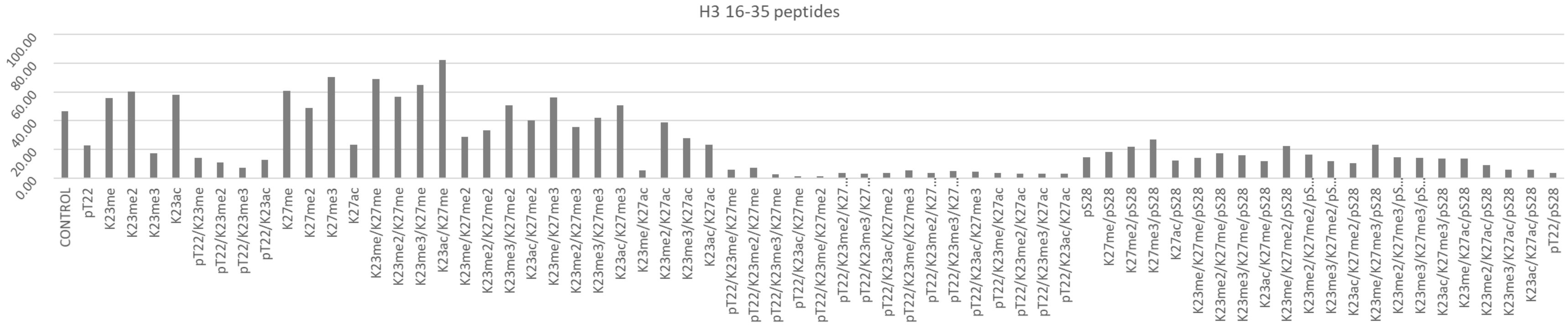
Table S2: quantification of NPAC PWWP$_{1-105}$ binding to histone modified peptides

This table does not correspond to the published one. For the aim of this thesis, I choose not to show the numerical values (published by Marabelli *et al*., 2019), but to represent them on a histogram table in order to facilitate visual inspection of the results obtained from the SPOT binding assay.

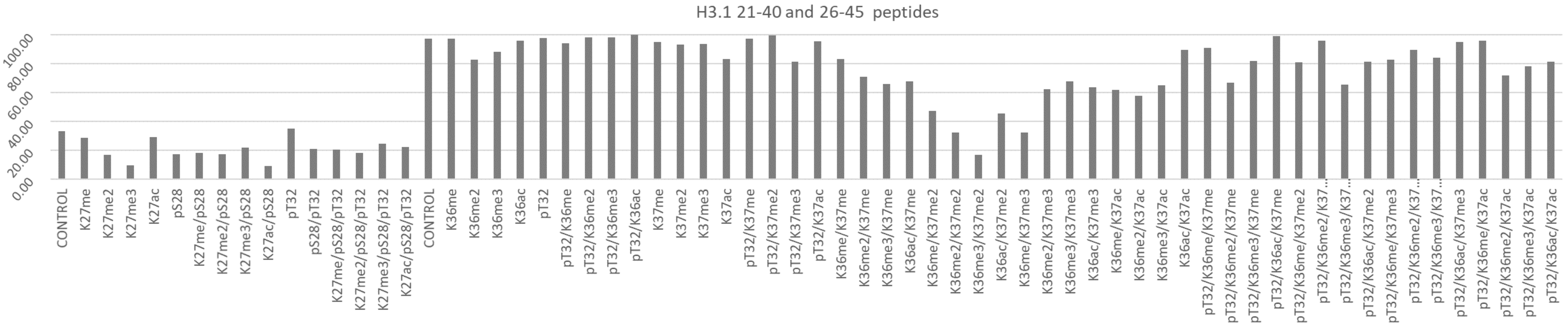


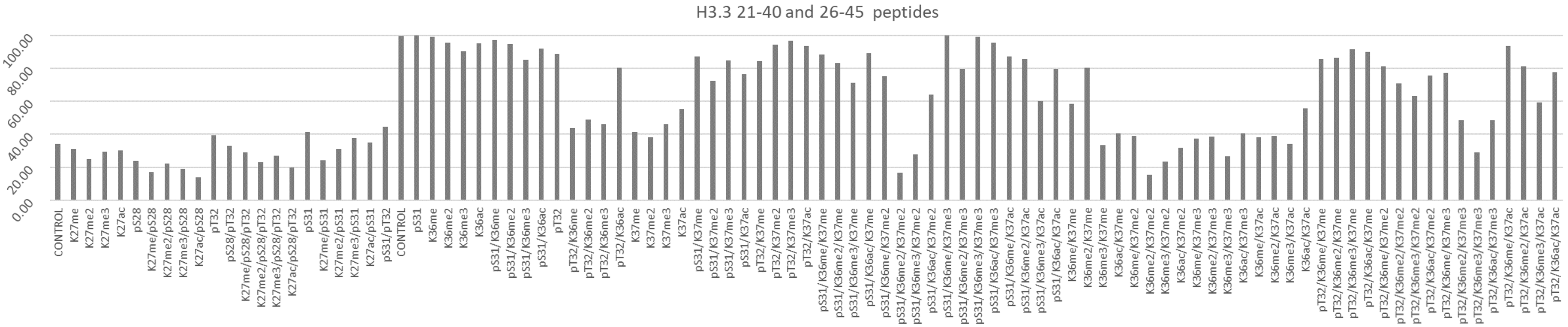


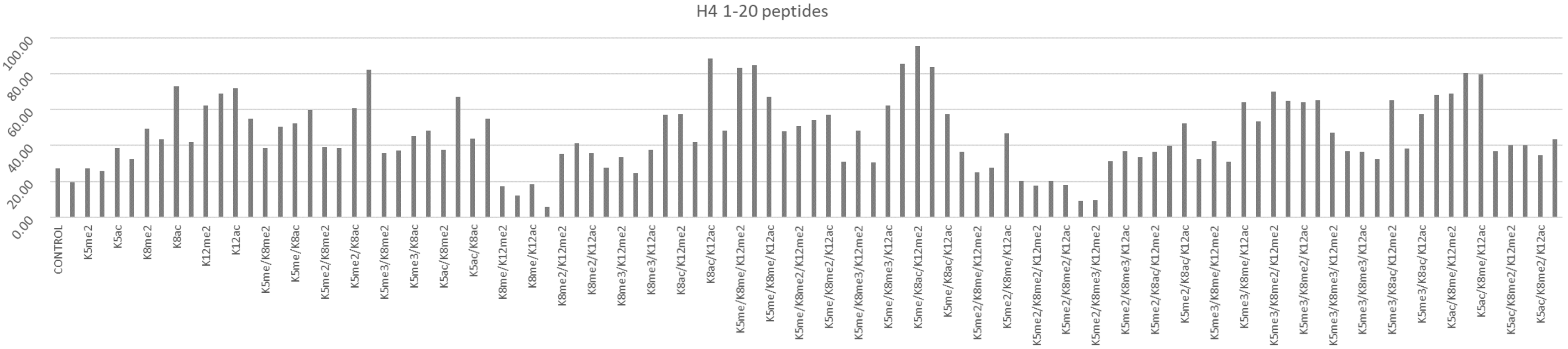
H3 11-30 peptides

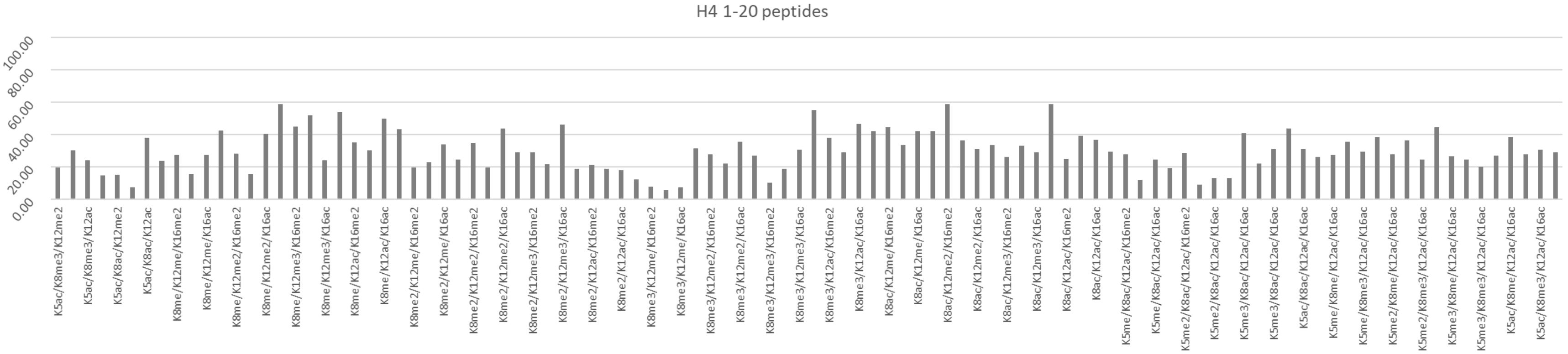


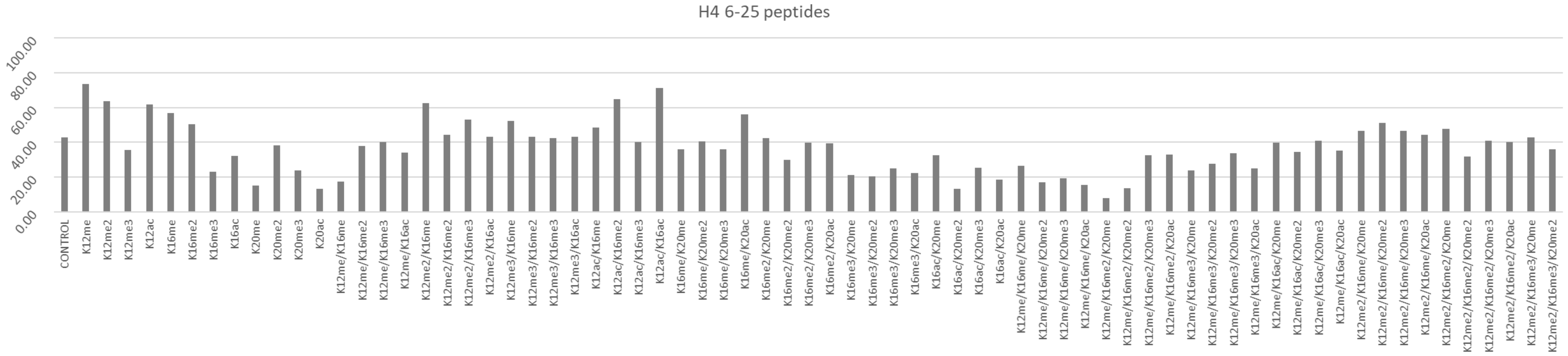
H3 16-35 peptides


H3.1 21-40 and 26-45 peptides
100.00
80.00
60.00
40.00
20.00
0.00
CONTROL
K27me
K27me2
K27me3
K27ac
pS28
K27me/pS28
K27me2/pS28
K27me3/pS28
K27ac/pS28
pT32
pS28/pT32
K27me/pS28/pT32
K27me2/pS28/pT32
K27me3/pS28/pT32
K27ac/pS28/pT32
CONTROL
K36me
K36me2
K36me3
K36ac
pT32
pT32/K36me
pT32/K36me2
pT32/K36me3
pT32/K36ac
K37me
K37me2
K37me3
K37ac
pT32/K37me
pT32/K37me2
pT32/K37me3
pT32/K37ac
K36me/K37me
K36me2/K37me
K36me3/K37me
K36ac/K37me
K36me/K37me2
K36me2/K37me2
K36me3/K37me2
K36ac/K37me2
K36me/K37me3
K36me2/K37me3
K36me3/K37me3
K36ac/K37me3
K36me/K37ac
K36me2/K37ac
K36me3/K37ac
K36ac/K37ac
pT32/K36me/K37me
pT32/K36me2/K37me
pT32/K36me3/K37me
pT32/K36ac/K37me
pT32/K36me/K37me2
pT32/K36me2/K37...
pT32/K36me3/K37...
pT32/K36ac/K37me2
pT32/K36me/K37me3
pT32/K36me2/K37...
pT32/K36me3/K37...
pT32/K36ac/K37me3
pT32/K36me/K37ac
pT32/K36me2/K37ac
pT32/K36me3/K37ac
pT32/K36ac/K37ac


H3.3 21-40 and 26-45 peptides
100.00
80.00
60.00
40.00
20.00
0.00
CONTROL
K27me
K27me2
K27me3
K27ac
pS28
K27me/pS28
K27me2/pS28
K27me3/pS28
K27ac/pS28
pT32
pS28/pT32
K27me/pS28/pT32
K27me2/pS28/pT32
K27me3/pS28/pT32
K27ac/pS28/pT32
pS31
K27me/pS31
K27me2/pS31
K27me3/pS31
K27ac/pS31
pS31/pT32
CONTROL
pS31
K36me
K36me2
K36me3
K36ac
pS31/K36me
pS31/K36me2
pS31/K36me3
pS31/K36ac
pT32
pT32/K36me
pT32/K36me2
pT32/K36me3
pT32/K36ac
K37me
K37me2
K37me3
K37ac
pS31/K37me
pS31/K37me2
pS31/K37me3
pS31/K37ac
pT32/K37me
pT32/K37me2
pT32/K37me3
pT32/K37ac
pS31/K36me/K37me
pS31/K36me2/K37me
pS31/K36me3/K37me
pS31/K36ac/K37me
pS31/K36me/K37me2
pS31/K36me2/K37me2
pS31/K36me3/K37me2
pS31/K36ac/K37me2
pS31/K36me/K37me3
pS31/K36me2/K37me3
pS31/K36me3/K37me3
pS31/K36ac/K37me3
pS31/K36me/K37ac
pS31/K36me2/K37ac
pS31/K36me3/K37ac
pS31/K36ac/K37ac
K36me/K37me
K36me2/K37me
K36me3/K37me
K36ac/K37me
K36me/K37me2
K36me2/K37me2
K36me3/K37me2
K36ac/K37me2
K36me/K37me3
K36me2/K37me3
K36me3/K37me3
K36ac/K37me3
K36me/K37ac
K36me2/K37ac
K36me3/K37ac
K36ac/K37ac
pT32/K36me/K37me
pT32/K36me2/K37me
pT32/K36me3/K37me
pT32/K36ac/K37me
pT32/K36me/K37me2
pT32/K36me2/K37me2
pT32/K36me3/K37me2
pT32/K36ac/K37me2
pT32/K36me/K37me3
pT32/K36me2/K37me3
pT32/K36me3/K37me3
pT32/K36ac/K37me3
pT32/K36me/K37ac
pT32/K36me2/K37ac
pT32/K36me3/K37ac
pT32/K36ac/K37ac


H4 1-20 peptides
100.00
80.00
60.00
40.00
20.00
0.00
CONTROL
K5me2
K5ac
K8me2
K8ac
K12me2
K12ac
K5me/K8me2
K5me/K8ac
K5me2/K8me2
K5me2/K8ac
K5me3/K8me2
K5me3/K8ac
K5ac/K8me2
K5ac/K8ac
K8me/K12me2
K8me/K12ac
K8me2/K12me2
K8me2/K12ac
K8me3/K12me2
K8me3/K12ac
K8ac/K12me2
K8ac/K12ac
K5me/K8me/K12me2
K5me/K8me/K12ac
K5me/K8me2/K12me2
K5me/K8me2/K12ac
K5me/K8me3/K12me2
K5me/K8me3/K12ac
K5me/K8ac/K12me2
K5me/K8ac/K12ac
K5me2/K8me/K12me2
K5me2/K8me/K12ac
K5me2/K8me2/K12me2
K5me2/K8me2/K12ac
K5me2/K8me3/K12me2
K5me2/K8me3/K12ac
K5me2/K8ac/K12me2
K5me2/K8ac/K12ac
K5me3/K8me/K12me2
K5me3/K8me/K12ac
K5me3/K8me2/K12me2
K5me3/K8me2/K12ac
K5me3/K8me3/K12me2
K5me3/K8me3/K12ac
K5me3/K8ac/K12me2
K5me3/K8ac/K12ac
K5ac/K8me/K12me2
K5ac/K8me/K12ac
K5ac/K8me2/K12me2
K5ac/K8me2/K12ac


H4 1-20 peptides
100.00
80.00
60.00
40.00
20.00
0.00
K5ac/K8me3/K12me2
K5ac/K8me3/K12ac
K5ac/K8ac/K12me2
K5ac/K8ac/K12ac
K8me/K12me/K16me2
K8me/K12me/K16ac
K8me/K12me2/K16me2
K8me/K12me2/K16ac
K8me/K12me3/K16me2
K8me/K12me3/K16ac
K8me/K12ac/K16me2
K8me/K12ac/K16ac
K8me2/K12me/K16me2
K8me2/K12me/K16ac
K8me2/K12me2/K16me2
K8me2/K12me2/K16ac
K8me2/K12me3/K16me2
K8me2/K12me3/K16ac
K8me2/K12ac/K16me2
K8me2/K12ac/K16ac
K8me3/K12me/K16me2
K8me3/K12me/K16ac
K8me3/K12me2/K16me2
K8me3/K12me2/K16ac
K8me3/K12me3/K16me2
K8me3/K12me3/K16ac
K8me3/K12ac/K16me2
K8me3/K12ac/K16ac
K8ac/K12me/K16me2
K8ac/K12me/K16ac
K8ac/K12me2/K16me2
K8ac/K12me2/K16ac
K8ac/K12me3/K16me2
K8ac/K12me3/K16ac
K8ac/K12ac/K16me2
K8ac/K12ac/K16ac
K5me/K8ac/K12ac/K16me2
K5me/K8ac/K12ac/K16ac
K5me2/K8ac/K12ac/K16me2
K5me2/K8ac/K12ac/K16ac
K5me3/K8ac/K12ac/K16ac
K5me3/K8ac/K12ac/K16ac
K5ac/K8ac/K12ac/K16ac
K5me/K8me/K12ac/K16ac
K5me/K8me3/K12ac/K16ac
K5me2/K8me/K12ac/K16ac
K5me2/K8me3/K12ac/K16ac
K5me3/K8me/K12ac/K16ac
K5me3/K8me3/K12ac/K16ac
K5ac/K8me/K12ac/K16ac
K5ac/K8me3/K12ac/K16ac


H4 6-25 peptides
100.00
80.00
60.00
40.00
20.00
0.00
CONTROL
K12me
K12me2
K12me3
K12ac
K16me
K16me2
K16me3
K16ac
K20me
K20me2
K20me3
K20ac
K12me/K16me
K12me/K16me2
K12me/K16me3
K12me/K16ac
K12me2/K16me
K12me2/K16me2
K12me2/K16me3
K12me2/K16ac
K12me3/K16me
K12me3/K16me2
K12me3/K16me3
K12me3/K16ac
K12ac/K16me
K12ac/K16me2
K12ac/K16me3
K12ac/K16ac
K16me/K20me
K16me/K20me2
K16me/K20me3
K16me/K20ac
K16me2/K20me
K16me2/K20me2
K16me2/K20me3
K16me2/K20ac
K16me3/K20me
K16me3/K20me2
K16me3/K20me3
K16me3/K20ac
K16ac/K20me
K16ac/K20me2
K16ac/K20me3
K16ac/K20ac
K12me/K16me/K20me
K12me/K16me/K20me2
K12me/K16me/K20me3
K12me/K16me/K20ac
K12me/K16me2/K20me
K12me/K16me2/K20me2
K12me/K16me2/K20me3
K12me/K16me2/K20ac
K12me/K16me3/K20me
K12me/K16me3/K20me2
K12me/K16me3/K20me3
K12me/K16me3/K20ac
K12me/K16ac/K20me
K12me/K16ac/K20me2
K12me/K16ac/K20me3
K12me/K16ac/K20ac
K12me2/K16me/K20me
K12me2/K16me/K20me2
K12me2/K16me/K20me3
K12me2/K16me/K20ac
K12me2/K16me2/K20me
K12me2/K16me2/K20me2
K12me2/K16me2/K20me3
K12me2/K16me2/K20ac
K12me2/K16me3/K20me
K12me2/K16me3/K20me2

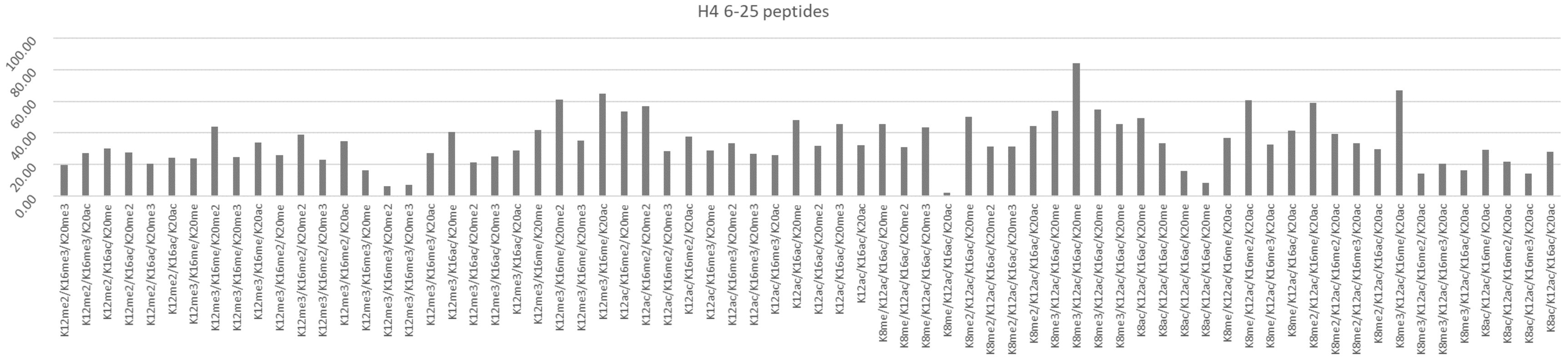

H4 6-25 peptides
100.00
80.00
60.00
40.00
20.00
0.00
K12me2/K16me3/K20me3
K12me2/K16me3/K20ac
K12me2/K16ac/K20me
K12me2/K16ac/K20me2
K12me2/K16ac/K20me3
K12me2/K16ac/K20ac
K12me3/K16me/K20me
K12me3/K16me/K20me2
K12me3/K16me/K20me3
K12me3/K16me/K20ac
K12me3/K16me2/K20me
K12me3/K16me2/K20me2
K12me3/K16me2/K20me3
K12me3/K16me2/K20ac
K12me3/K16me3/K20me
K12me3/K16me3/K20me2
K12me3/K16me3/K20me3
K12me3/K16me3/K20ac
K12me3/K16ac/K20me
K12me3/K16ac/K20me2
K12me3/K16ac/K20me3
K12me3/K16ac/K20ac
K12me3/K16me/K20me
K12me3/K16me/K20me2
K12me3/K16me/K20me3
K12me3/K16me/K20ac
K12ac/K16me2/K20me
K12ac/K16me2/K20me2
K12ac/K16me2/K20me3
K12ac/K16me2/K20ac
K12ac/K16me3/K20me
K12ac/K16me3/K20me2
K12ac/K16me3/K20me3
K12ac/K16me3/K20ac
K12ac/K16ac/K20me
K12ac/K16ac/K20me2
K12ac/K16ac/K20me3
K12ac/K16ac/K20ac
K8me/K12ac/K16ac/K20me
K8me/K12ac/K16ac/K20me2
K8me/K12ac/K16ac/K20me3
K8me/K12ac/K16ac/K20ac
K8me2/K12ac/K16ac/K20me
K8me2/K12ac/K16ac/K20me2
K8me2/K12ac/K16ac/K20me3
K8me2/K12ac/K16ac/K20ac
K8me3/K12ac/K16ac/K20me
K8me3/K12ac/K16ac/K20me
K8me3/K12ac/K16ac/K20me
K8me3/K12ac/K16ac/K20me
K8ac/K12ac/K16ac/K20me
K8ac/K12ac/K16ac/K20me
K8ac/K12ac/K16ac/K20me
K8ac/K12ac/K16ac/K20me
K8me/K12ac/K16me/K20ac
K8me/K12ac/K16me2/K20ac
K8me/K12ac/K16me3/K20ac
K8me/K12ac/K16ac/K20ac
K8me2/K12ac/K16me/K20ac
K8me2/K12ac/K16me2/K20ac
K8me2/K12ac/K16me3/K20ac
K8me2/K12ac/K16ac/K20ac
K8me3/K12ac/K16me/K20ac
K8me3/K12ac/K16me2/K20ac
K8me3/K12ac/K16me3/K20ac
K8me3/K12ac/K16ac/K20ac
K8ac/K12ac/K16me/K20ac
K8ac/K12ac/K16me2/K20ac
K8ac/K12ac/K16me3/K20ac
K8ac/K12ac/K16ac/K20ac

Table S3 related to Figures 5-6. Structural parameters calculated from SEC-SAXS data[a]

	LSD2	**NPAC dehydrogenase**	**NPAC Δ205**	**NCP**	**LSD2/NPACΔ205/ NCP[b]**
P(r) analysis					
I (0)	27.01	14.07	24.00	90.55	188.92
Rg (nm)	3.40	3.54	4.01	4.27	8.03
Dmax (nm)	10.59	10.86	11.36	12.80	27.30
Vp (nm^3)	123.82	168.68	240.06	353.22	1,090
Guinier analysis					
I (0)	26.89 ± 0.07	14.07 ± 0.01	24.20 ± 0.07	91.23 ± 0.10	188.92 ± 0.35
Rg (nm)	3.36 ± 0.12	3.54 ± 0.01	4.05 ± 0.21	4.35 ± 0.01	8.03 ± 0.07
MW estimate					
Credibility Interval (kDa)	86.95 – 95.80	111.25 – 134.30	151.45 – 176.60	194.96 – 264.20	455.20 – 614.45
Calculated MW (kDa)	89.88	4 x 31.59	4 x 37.54	198.66	438,70[c] - 727,24[d]

[a] I(0): Scattering intensity at zero angle; Rg: radius of gyration; D_{max}: maximum distance in the scattering particle; Vp: porod volume. NPACΔ205 and NPAC dehydrogenase contain residues 206-553 and 261-553, respectively. LSD2 includes residues 31-883. Scattering curves are shown in Figure S7.

[b] Peaks B-C of the chromatogram shown in Figure 6A and Figure S7-bottom left.

[c] Calculated molecular weight of a complex containing one nucleosome, one LSD2, and a tetramer of NPAC Δ205.

[d] Calculated molecular weight of a complex containing two nucleosomes, two LSD2s, and a tetramer of NPAC Δ205.

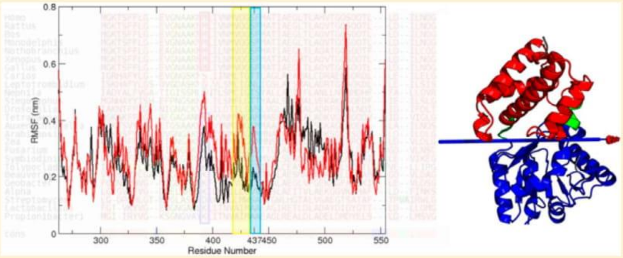
JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING
Cite This: J. Chem. Inf. Model. XXXX, XXX, XXX−XXX
Article
pubs.acs.org/jcim

Impact of Mutations on NPAC Structural Dynamics: Mechanistic Insights from MD Simulations

Marco Montefiori,[†] Simona Pilotto,[‡] Chiara Marabelli,[‡] Elisabetta Moroni,[†] Mariarosaria Ferraro,[†] Stefano A. Serapian,[§] Andrea Mattevi,[*,‡] and Giorgio Colombo[*,†,§]

[†]ICRM-CNR, Via Mario Bianco 9, 20131 Milano, Italy
[‡]Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9, 27100 Pavia, Italy
[§]University of Pavia, Department of Chemistry, V.le Taramelli 12, 27100 Pavia, Italy

Supporting Information



ABSTRACT: NPAC is a cytokine-like nuclear factor involved in chromatin modification and regulation of gene expression. In humans, the C-terminal domain of NPAC has the conserved structure of the β-hydroxyacid dehydrogenases (β-HAD) protein superfamily, which forms a stable tetrameric core scaffold for demethylase enzymes and organizes multiple sites for chromatin interactions. In spite of the close structural resemblance to other β-HAD family members, the human NPAC dehydrogenase domain lacks a highly conserved catalytic lysine, substituted by a methionine. The reintroduction of the catalytic lysine by M437 K mutation results in a significant decrease of stability of the tetramer. Here, we have computationally investigated the molecular determinants of the functional differences between methionine and lysine-containing NPAC proteins. We find that the single mutation can determine strong consequences in terms of dynamics, stability, and ultimately ability to assemble in supramolecular complexes: the higher stability and lower flexibility of the methionine variant structurally preorganizes the monomer for tetramerization, whereas lysine increases flexibility and favors conformations that, while catalytically active, are not optimal for tetrameric assembly. We combine structure-dynamics analysis to an evolutionary study of NPAC sequences, showing that the methionine mutation occurs in a specifically flexible region of the lysine-containing protein, flanked by two domains that concentrate most of the stabilizing interactions. In our model, such separation of stability nuclei and flexible regions appears to favor the functional innovability of the protein.

INTRODUCTION

NPAC, also known by the name of glycoxylate reductase homologue 1 (glyr1) or nuclear protein 60 (NP60), is a cytokine-like nuclear factor involved in chromatin modification and regulation of gene expression.[1−3] It has been demonstrated that NPAC directly interacts with LSD2/KDM1B and modulates its histone demethylase activity on mono- and dimethyl-Lys4 of histone H3 (H3K4me1 and H3K4me2).[2−4] NPAC is also involved in the p38 MAPK signaling pathway through the phosphorylation (and activation) of p38α and, last, the positive regulation of activating transcription factor 2 (ATF2).[1−3] A recent study demonstrated a correlation between a reduced NPAC expression and hepatocellular carcinoma.[3]

NPAC is a multidomain protein[5] (see domain organization scheme in Figure 1A). The N-terminal PWWP domain is connected to the C-terminal dehydrogenase domain by a long stretch of 170 residues.[5] This stretch contains two different protein−protein interaction motifs: an AT-hook module which likely binds DNA and p38α and a linker region consisting of 12 amino acids that tightly bind and activate LSD2.[1,2,5] The C-terminal domain has the conserved structure of the β-hydroxyacid dehydrogenases (β-HAD) protein superfamily. β-HAD members catalyze the NAD$^+$ or NADP$^+$-dependent oxidation of β-hydroxy acids to their corresponding semi-aldehydes and share the same overall folding topology and three-dimensional structure.[6] Similar to the β-HAD enzymes, the dehydrogenase domain of NPAC can be divided into two different regions, a Rossman fold and a six-α-helical bundle, connected by a long α-helix (Figure 1B).[6,7] The structural and

Received: July 18, 2019
Published: August 13, 2019

DOI: 10.1021/acs.jcim.9b00588
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

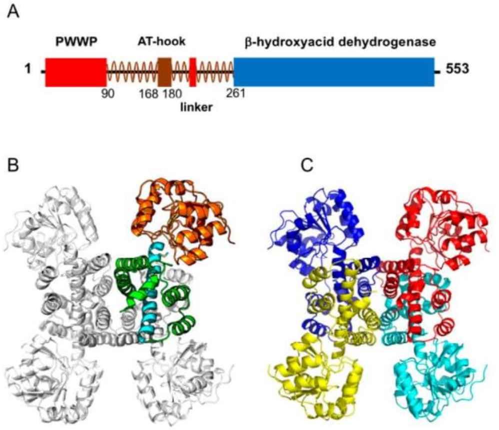


Figure 1. Domain organization and structure of the human NPAC tetramer dehydrogenase domain (PDB: 2uyy). (A) Simplified scheme depicting the domain organization of NPAC. (B) Subdivision of monomer A in the different regions, the Rossman-fold globular domain in orange, the bridging α9-helix in cyan, and the α-helical bundle in green. The mutated Met437 is represented in stick. (C) Subdivision of the tetramer in the four monomers (chain A, red; chain B, cyan; chain C, blue; chain D, yellow).

biochemical analysis clearly demonstrated that the NPAC dehydrogenase domain forms a homotetrameric assembly (Figure 1C). This feature underlies the assembly of the LSD2-NPAC. The two proteins form a chromatin-modifying multienzyme complex where the NPAC dehydrogenase forms a stable tetrameric core anchoring the flexible arms formed by the NPAC N-terminal segments, each arm bearing a demethylase enzyme (LSD2) and multiple sites (PWWP and AT-hook domains) for chromatin interactions. This architecture ensures the efficient and processive modification of the chromatin to favor transcription elongation.[5]

LSD2-NPAC has indeed been predominantly found in actively RNAPolII-transcribed gene bodies.[8]

β-HAD family members are characterized by conserved sequence and structural features,[9] and NPAC dehydrogenase domain falls within the same trend. Yet, the human NPAC lacks an extremely conserved lysine (Met437) which is essential for the dehydrogenase activity. Intriguingly, the same inactivating lysine-to-methionine mutation is conserved in other orthologs, and it seems to have appeared during evolution at the same time as the enzyme acquired a chromatin-related function.[10] The effect of the catalytic lysine was investigated by Marabelli et al. through the mutation of methionine back to lysine (M437 K).[5] Interestingly, this variation resulted in a significant decrease of stability of the tetramer, which promptly disassembled.[5] These results led to the intriguing hypothesis that a progenitor dehydrogenase enzyme has evolved to the extant, catalytically inactive, NPAC dehydrogenase whose main novel function is the formation of an oligomerization scaffold at the heart of a chromatin-modifying multienzyme complex.

Here, we set up a comparative framework to investigate the common and differential traits of internal dynamics and energetics of Methionine- and Lysine- containing NPAC proteins that determine the stability and motions of the isolated monomers. The working hypothesis is that differences in the conformational dynamic and stability properties of the monomers reverberate in different functional properties evidenced by the observed different tendencies to tetramerize.

RESULTS AND DISCUSSION

The 3D structure of human NPAC (PDB code 2uyy) was used as starting point for extensive atomistic MD simulations and their analyses. The technical details are described in the Materials and Methods section. Specifically, the wild-type and the M437 K mutant NPAC proteins were simulated for 1.5 μs, made up by three independent replicas of 500 ns each, and all of the analyses were carried out on the combinations of the equilibrated parts of the respective trajectories (see the Materials and Methods). Models of tetramers were also built, simulated, and analyzed via similar procedures. Specifically, rigid domains and their principal motions were extracted from the MD trajectories. We further coupled this analysis to the investigation of the stability and plasticity properties of the different sequences through newly developed methods of internal dynamics and energy. Finally, we carried out evolutionary analyses of NPAC sequences to trace the initial onset of the Lys-to-Met variation, in the light of possible functional requirements. The overall aim was to characterize distinctive elements that underlie the tendency to form (or oppose) stable and functional tetramers and define how evolution may have shaped sequences of homologous proteins to carry out such different tasks via small sequence variations.

Dynamic Domain Analysis: The M437 K Mutation Creates a New Molecular Hinge between the Two NPAC Domains. First, we asked whether the M437 K mutation perturbs the global and/or internal dynamics of the respective human NPAC monomers. To address this question, we first ran structural cluster analysis of trajectories from the wild-type and mutant MD simulations. Interestingly, while a major dominant cluster emerged for the wild-type protein, indicative of the sampling of overall similar structures, M437 K appeared to explore a wider range of conformational arrangements. The representative structures of the two main clusters of the mutant were significantly different from each other and from the starting conformation observed in the wild-type crystal. Table 1 shows the difference in root-mean-square deviation (RMSD) values (calculated on the Cα trace) between the two most populated clusters in the wild type and M437 K mutant simulations. While the two most populated clusters in the wild-type protein have fairly similar RMSD values (0.174 nm), indicative of a substantially stable structure, in the mutant the difference is significantly larger (0.6 nm). This is indicative of the fact that the wild type fluctuates close to the initial native structure whereas

Table 1. RMSD Values (nm) Calculated on the Cα, between the Two Most Representative Clusters in the Wild Type and M437 K Mutant Simulations

	cluster 1 wild-type	cluster 2 wild-type	cluster 1 M437 K	cluster 2 M437 K
cluster 1 wild-type		0.174	0.506	0.706
cluster 2 wild-type			0.503	0.670
cluster 1 M437 K				0.600
cluster 2 M437 K				

the mutant explores different conformational states, each characterized by a larger variability in the structural ensemble. The RMSD evolution of the single trajectories are reported in Figure S1, showing the different behavior of the two systems and the variations in the single trajectories.

Intrigued by the observation that restoration of the putative catalytic Lys may facilitate the exploration of different structural ensembles, we set out to better characterize the large-scale motions of the N-terminal Rossman-fold domain (residues 262−437) and of the C-terminal helix bundle domain (residues 438−552), forming the dehydrogenase scaffold (Figure 1). To this end, we analyzed the representative structures of the two most populated clusters from each protein with the DynDom approach.[11] DynDom defines protein domains, hinge axes, and amino acid residues involved in the hinge motions. The algorithm works by carrying out a best fit of the two conformations and defining possible rotation vectors. On the basis of the theorem of Chasles, which states that the general displacement of a rigid body is a screw motion, interdomain screw axes are determined. The location of the screw axis indicates the hinge location and kind of motion allowed by the interdomain connections.[11] DynDom analysis on the first two clusters of the wild-type monomer did not reveal any dynamic domain movement, confirming a relatively stable structure. In contrast, the run with the two representatives of the two principal clusters from monomeric M437 K pointed to the presence of a molecular hinge (residues 436−438) splitting the protein in two uncoupled domains. Interestingly, the molecular hinge is located in the sequence stretch hosting the mutation. The DynDom analysis also highlighted a twisting axis, which crosses the long α-helix in between the two domains, as shown in Figure 2. The analysis further showed that the two domains lean toward each other. This leads to the reduction of the physical separation between the two domains and to the formation of new intramonomer contacts. Detailed inspection of the M437 K ensemble revealed that most of the contacts between the two domains are established by residues that in the wild-type tetramer are part of the intermolecular contact region (residue 438 and residue 552 of the proximal subunit). To further investigate the dynamic differences between the wild-type and the M437 K mutant we further compared the structural representative of the most populated clusters of the wild-type and M437 K conformers. Also in this case, DynDom analysis returned two domains, with the same subdivision as described above. The hinge was again identified in the residues 437−438 (Figure 2). In summary, MD simulations indicate that the M437 K mutation distorts the tetramer-competent conformation of the monomer.

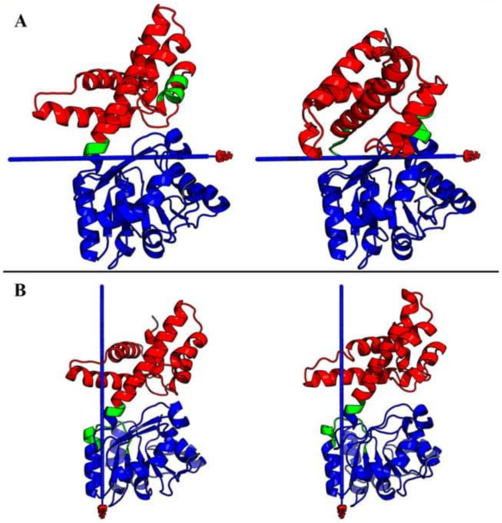


Figure 2. Preferential axes of rotation of the two domains identified via DynDom analyses. (A) the twisting between the two most populated conformations in the mutant. (B) The comparison between the ideal axis of rotation of the mutant (right part of the subfigure) and that of the wild-type protein.

M437 K Mutation Induces Differential Internal Dynamics in the Monomer: Insights from H-Bonding, Cavity Fluctuation, and Pair Distance Fluctuation Analyses. To further investigate the effect of the mutation, we characterized the persistency of native hydrogen bonds and inner cavities upon mutation, both in the monomeric and tetrameric forms.

Consistent with the observed conformational change that makes the two domains of the mutant bend toward each other and establish new interactions, the occupancy of most of the hydrogen bond donors increased in the mutant simulation, with the exception of two H-bonds (S332s-C330m and N473s-L469m). As shown in Figure 3, most of the newly formed interactions of the mutated monomer are localized in the region of the globular domain that would face the tetramerization domain.

The global effect of the M437 K mutation was additionally demonstrated by the inspection of the cavities and pockets in the conformational ensembles of the two mutants. Transient cavity formation has been linked to the possibility of defining pathways that interconnect different dynamic regions in a protein. In our case, the appearance of persistent cavities can be associated with the structural instability of a particular subregion of the protein. To shed light on these aspects, we applied a novel algorithm that examines dynamic pocket formation and cross-talk (see the Materials and Methods and ref 12). As reported in Figure 4, only a small cavity of the wild-type protein, between the globular and the tetramerization domain, consistently showed stable and conserved networks. In contrast, the mutant featured four large, persistent and interconnected cavities. Solvent could aptly penetrate such cavities favoring unfolding processes. Indeed, experimental observations indicate that monomeric M347 K mutant has a tendency to aggregate.[5] These findings confirmed that the M437 K mutation strongly affects protein conformation and dynamics.

We next analyzed internal dynamic changes induced by the M437 K mutation. The underlying hypothesis here is that specific protein regions can respond to it through the modulation of their coordination patterns with proximal and/or distal regions. The latter aspect was captured through the analysis of the distance fluctuation map of the two protein structures (Figure 5). This analysis reports on the internal rigidity within the protein by evaluating the fluctuation of distances for amino acid pairs.[13−15] Each matrix entry reports on the fluctuation of the inter-residue distance in the corresponding residue pairs and matrix regions. Relatively low values identify protein subdomains that move together (in coordination) while undergoing structural fluctuations. Here, we first computed the

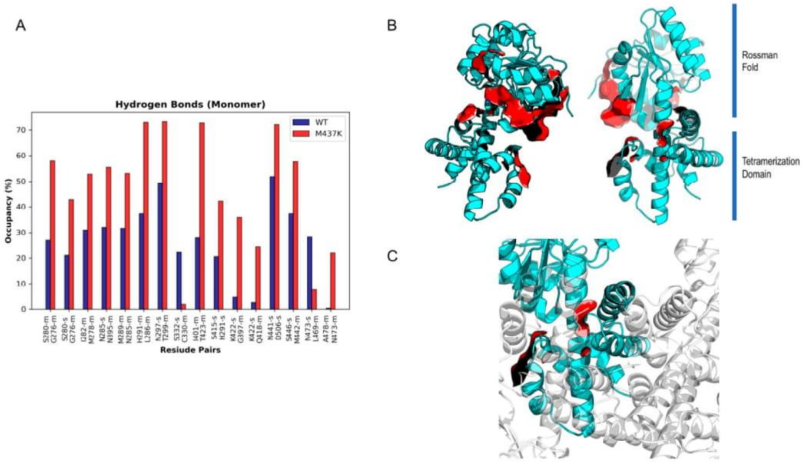


Figure 3. (A) Comparison of the occupancy of the hydrogen bonds in the wild-type (blue) and M437 K mutant (red) monomer. The reported hydrogen bonds differ by more than 20% in their occupancies. The residues are represented with the one-letter code, and their number followed by an "m" for main-chain interactions and "s" for side-chain interactions. (B) The regions in the mutant monomer where new intramolecular hydrogen bonds are established are highlighted as a red surface. (C) The regions in the tetramerization domain that are involved in the formation of new intramolecular hydrogen bonds shown in the context of the tetramer.

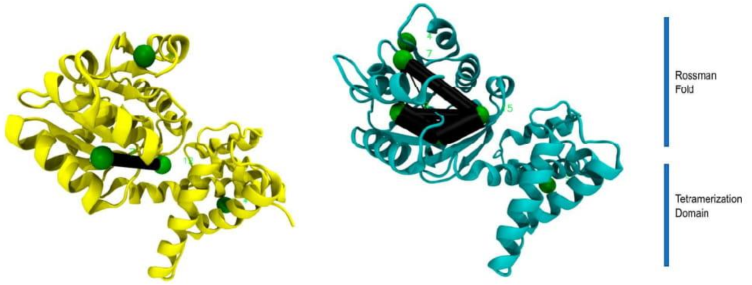


Figure 4. Cavity network for the monomer wild-type (yellow, left) and mutant M437 K (cyan, right).

pairwise amino acid fluctuations from the MD trajectories for the wild-type and M437 K mutant. Overall, the matrices for both wild-type and M437 K mutant turned out to be similar, with mostly overlapping but not identical patterns of small and large fluctuations of inter-residue distances. The finer differences in the global dynamics of the two proteins were best revealed by subtracting the distance matrix of the mutant from the matrix calculated from the simulation of the wild-type protein (Figure 5A). No significant differences could be appreciated in the N-terminal region. By contrast, the C-terminal region of the protein showed diverging dynamical properties with the wild type featuring higher mobility for residues 467−487 and the M437 K mutant being more flexible in the segment 235−280 (Figure 5B). The substructures on the monomer that become more flexible upon mutation from M to K are depicted in blue in the 3D structure in Figure 5B. Critically, residues 235−280 are at the heart of the tetramerization interface: the M437 K mutation is thereby confirmed to affect the dynamical and structural properties of those domain regions directly involved in oligomerization.

Energetic Impact of the M437 K Mutation on the Main Determinants of Structural stabilization. To investigate the wild-type and mutant proteins at a finer level of detail and shed light on whether specific networks of interactions among residues are differentially determined by the sequence, we applied the Energy Decomposition Method to the representative structures sampled for the two proteins.[16−19] The aim was to identify the intramolecular interaction networks of the wild-type structural ensemble that are modulated or perturbed by sequence variation. More specifically, the energy decomposition method detects relevant energetic couplings between residues through a principal component decomposition of the non-bonded interaction energy matrix for the whole protein. An effective mean coupling energy between all the residue-pairs maps the principal energetic interactions in the native three-dimensional state of the protein and the eigenvector associated with the lowest eigenvalue is used to identify the most stabilizing amino acids. In particular, considering its squared components as the weights of the corresponding residues in the structural stabilization, "hot spots" can be defined as those residues with a weight higher than a threshold *t*. This threshold is chosen equal

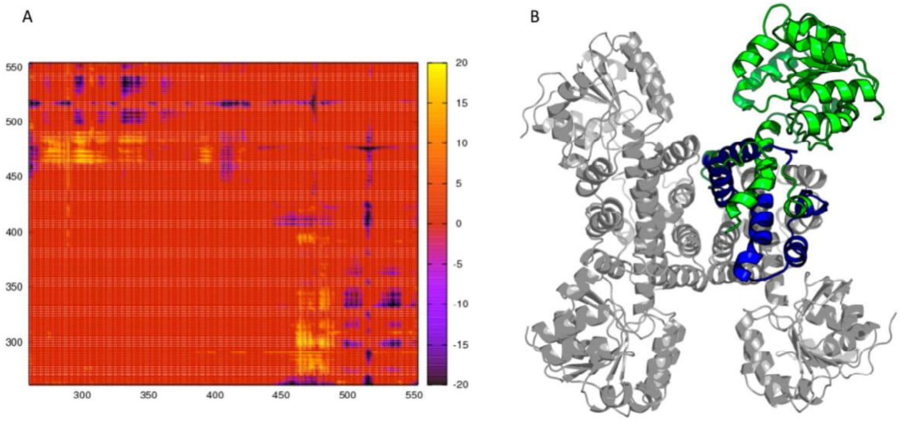


Figure 5. (A) Distance fluctuation matrix difference for the monomer (wild-type - M437 K Mutant). The values in blue are the ones for which the fluctuations are larger in the M437 K mutant compared to the wild-type, and the yellow shades are the opposite. (B) The computationally predicted flexible helices of the M437 K mutant are mapped on the three-dimensional structure of the tetramer and depicted in blue.

to the squared component of a normalized "flat eigenvector" (a normalized vector whose components provide the same contribution for each site), whereby each residue contributes equally to the structural stability and, therefore, the threshold *t* is equal to *1/N*, where *N* is the number of the eigenvector components. Upon projection of the results of the analysis on the structures, it is possible to visualize networks of principal stabilizing amino acid interactions and to explain the long-range effects of mutation-induced perturbations on protein conformational properties. In this context, the Energy Decomposition Method detects the most relevant residue−residue interactions that implicitly define the clusters which are necessary to define the three-dimensional fold of the different domains (or subunits). Indeed, this method previously proved to efficiently account for several properties, such as the effects of mutations in terms of stability variation and the modulation of local and global conformational dynamics.[16−19]

The profile of the first eigenvector (the one associated with the lowest eigenvalue), which reports on the contribution of each single residue to the stabilization of the structures, shows interesting differences between the wild-type and M437 K mutant. In particular, the profile changes in the region around the mutation site (Figure 6).

In the mutant, a significant quenching of the intermolecular interactions can be observed, determined by a clear reduction of the contribution to the stabilization of region 420−435 with respect to the wild-type, paralleled by an increase of the contribution of the residues in the N- and C-terminal domains, indicating a polarization of the nuclei of stabilizing interactions. In contrast, the wild-type protein is characterized by a more diffuse network of pair interactions that extends throughout the structure as shown in Figure 7. Collectively, these data confirm that the energetic uncoupling of the two domains in the mutant lowers/negatively affects the preorganization of the monomer structure in the necessary ensemble for tetramer assembly.

In this model, higher conformational freedom imparted by the mutation is detrimental to the stabilization of the tetramer. This aspect is further reflected in a second parameter shown to be associated with protein stability, namely the gap between the first eigenvalue (which recapitulates most of the stabilization energy) and all the others (which also include repulsive and nonstabilizing interactions, see also the Materials and Methods section).[16−19] We characterized this gap by calculating the difference between the first (most negative) and second eigenvalue, divided by the average spacing between all of the others:

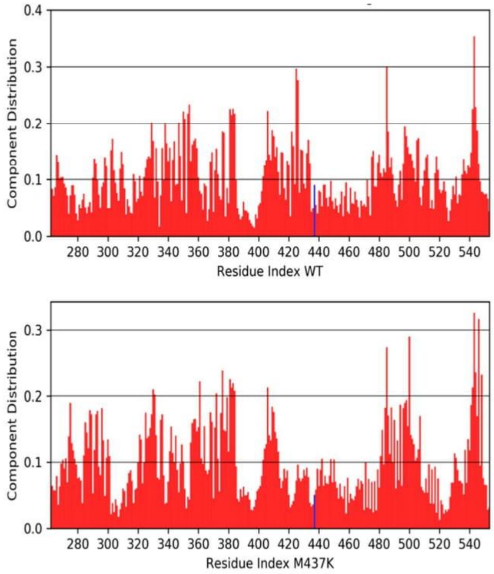


Figure 6. Distribution of interacting energy against the residue position for the wild-type (upper panel) and the M437 K (lower panel) monomers. The position of the mutation (residue 437) is labeled in blue.

$$\frac{|\lambda_1 - \lambda_2|}{\Delta\tilde{\lambda}} \tag{1}$$

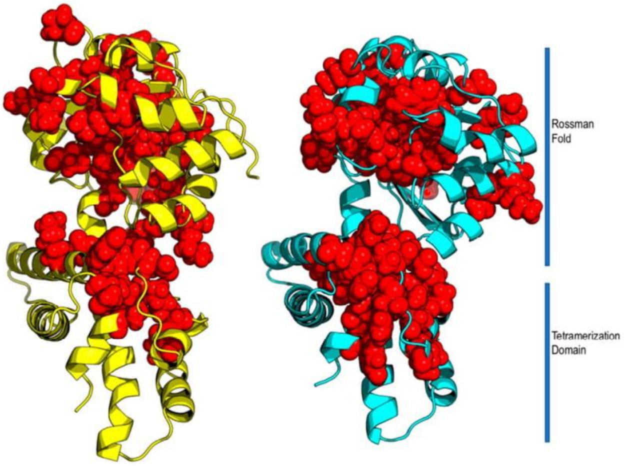


Figure 7. Representation of component distribution larger than 0.15 in both the wild-type (yellow, left) and M437 K monomers (cyan, right). The residues with component distribution larger than 0.15 are represented as van der Waals spheres in red.

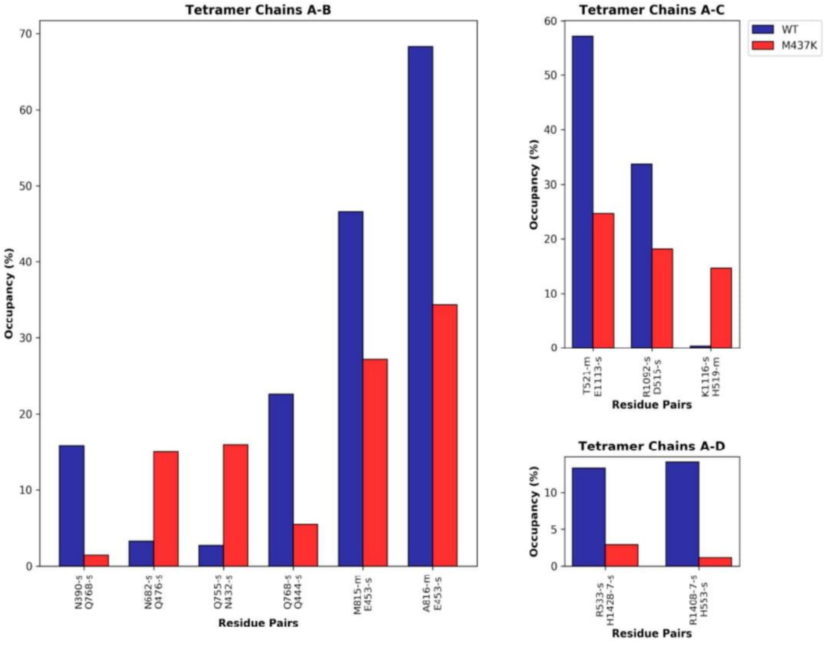
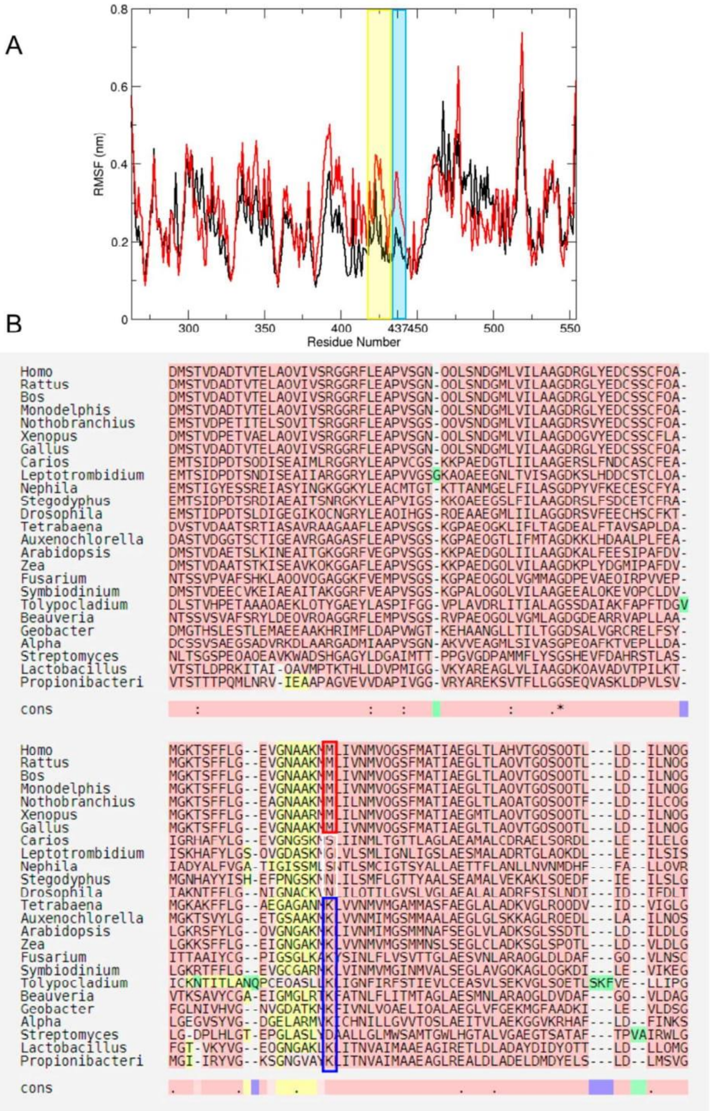


Figure 8. Comparison of the occupancy of the intersubunit hydrogen bonds with chain A as the reference. The reported hydrogen bonds reported exhibit occupancy differences larger than 10%. The residues are represented with the one-letter code and their residue number followed by "m" for main-chain interactions and "s" for side-chain interactions.

The large difference between the average values calculated for wild-type and (68.9) and M437 K (59.1) proteins corroborate the notion that the wild-type dehydrogenase is more stable than the M437 K mutant.

It must be noted here that other energy calculation methods could be used to estimate the differential stability in the monomers and even in the tetramers, in a more quantitative way. These methods include Free Energy Perturbation[20] or MMPBSA[21] approaches. However, by testing the latter on our trajectories, we were unable to obtain significant differences between the wild-type and the M437 K mutant trajectories, both for the monomers and for the tetramer. We suggest that the



```
Homo              DMSTVDADTVTELAOVIVSRGGRFLEAPVSGN-OOLSNDGMLVILAAGDRGLYEDCSSCFOA-
Rattus            DMSTVDADTVTELAOVIVSRGGRFLEAPVSGN-OOLSNDGMLVILAAGDRGLYEDCSSCFOA-
Bos               DMSTVDADTVTELAOVIVSRGGRFLEAPVSGN-OOLSNDGMLVILAAGDRGLYEDCSSCFOA-
Monodelphis       DMSTVDADTVTELAOVIVSRGGRFLEAPVSGN-OOLSNDGMLVILAAGDRGLYEDCSSCFOA-
Nothobranchius    EMSTVDPETITELSOVITSRGGRFLEAPVSGS-OOVSNDGMLVILAAGDRSVYEDCSSCFOA-
Xenopus           DMSTVDPETVAELAOVIVSRGGRFLEAPVSGN-OOLSNDGMLVILAAGDOGVYEDCSSCFLA-
Gallus            DMSTVDADTVTELAOVIVSRGGRFLEAPVSGN-OOLSNDGMLVILAAGDRGLYEDCSSCFOA-
Carios            EMTSIDPDTSODISEAIMLRGGRYLEAPVCGS-KKPAEDGTLIILAAGERSLFNDCASCFEA-
Leptotrombidium   EMTSIDPDTSNDISEAIIARGGRYLEAPVVGSGKAOAEEGNLTVISAGDKSLHDDCSTCLOA-
Nephila           EMSTIGYESSREIASYINGKGGKYLEACMTGT-KTTANMGELFILASGDPYVFKECESCFYA-
Stegodyphus       EMTSIDPDTSRDIAEAITSNRGKYLEAPVIGS-KKOAEEGSLFILAAGDRSLFSDCETCFRA-
Drosophila        EMSTIDPDTSLDIGEGIKOCNGRYLEAOIHGS-ROEAAEGMLIILAGGDRSVFEECHSCFKT-
Tetrabaena        DVSTVDAATSRTIASAVRAAGAAFLEAPVSGS-KGPAEOGKLIFLTAGDEALFTAVSAPLDA-
Auxenochlorella   DASTVDGGTSCTIGEAVRGAGASFLEAPVSGS-KGPAEOGTLIFMTAGDKKLHDAALPLFEA-
Arabidopsis       DMSTVDAETSLKINEAITGKGGRFVEGPVSGS-KKPAEDGOLIILAAGDKALFEESIPAFDV-
Zea               DMSTVDAATSTKISEAVKOKGGAFLEAPVSGS-KKPAEDGOLVILAAGDKPLYDGMIPAFDV-
Fusarium          NTSSVPVAFSHKLAOOVOGAGGKFVEMPVSGS-KGPAEOGOLVGMMAGDPEVAEOIRPVVEP-
Symbiodinium      DMSTVDEECVKEIAEAITAKGGRFVEAPVSGS-KGPALOGOLVILAAGEEALOKEVOPCLDV-
Tolypocladium     DLSTVHPETAAAOAEKLOTYGAEYLASPIFGG-VPLAVDRLITIALAGSSDAIAKFAPFTDGV
Beauveria         NTSSVSVAFSRYLDEOVROAGGRFLEMPVSGS-RVPAEOGOLVGMLAGDGDEARRVAPLLAA-
Geobacter         DMGTHSLESTLEMAEEAAKHRIMFLDAPVWGT-KEHAANGLLTILTGGDSALVGRCRELFSY-
Alpha             DCSSVSAEGSADVRKDLAARGADMIAAPVSGN-AKVVEAGMLSIVASGPEOAFKTVEPLLDA-
Streptomyces      NLTSGSPEOAOEAVKWADSHGAGYLDGAIMTT-PPGVGDPAMMFLYSGSHEVFDAHRSTLAS-
Lactobacillus     VTSTLDPRKITAI-OAVMPTKTHLLDVPMIGG-VKYAREAGLVLIAAGDKOAVADVTPILKT-
Propionibacteri   VTSTTTPQMLNRV-IEAAPAGVEVVDAPIVGG-VRYAREKSVTFLLGGSEQVASKLDPVLSV-

cons                 :                     :   :             :    .*

Homo              MGKTSFFLG--EVGNAAKNMLIVNMVOGSFMATIAEGLTLAHVTGOSOOTL---LD--ILNOG
Rattus            MGKTSFFLG--EVGNAAKNMLIVNMVOGSFMATIAEGLTLAOVTGOSOOTL---LD--ILNOG
Bos               MGKTSFFLG--EVGNAAKNMLIVNMVOGSFMATIAEGLTLAOVTGOSOOTL---LD--ILNOG
Monodelphis       MGKTSFFLG--EVGNAAKNMLIVNMVOGSFMATIAEGLTLAOVTGOSOOTL---LD--ILNOG
Nothobranchius    MGKTSFFLG--EAGNAAKNMLILNMVOGSFMATIAEGLTLAOATGOSOOTF---LD--ILCOG
Xenopus           MGKTSFFLG--EVGNAARNMLILNMVOGSFMATIAEGMTLAOVTGOSOOTL---LD--ILNOG
Gallus            MGKTSFFLG--EVGNAAKNMLIVNMVOGSFMATIAEGLTLAOVTGOSOOTL---LD--ILNOG
Carios            IGRHAFYLG--EVGNGSKNSLIINMLTGTTLAGLAEAMALCDRAELSORDL---LE--ILELG
Leptotrombidium   ISKHAFYLGS-OVGDASKNGLVLSMLIGNLIGSLAESMALADRTGLAOKDL---LE--ILSIS
Nephila           IADYALFVGA-TIGISSMLSNTLSMCIGTSYALLAETTFLANLLNVNMDHF---FA--LLOVR
Stegodyphus       MGNHAYYISH-EFPNGSKNNLILSMFLGTTYAALSEAMALVEKAKLSOEDF---IE--ILSLG
Drosophila        IAKNTFFLG--NIGNACKVNLILOTILGVSLVGLAEALALADRFSISLNDI---ID--IFDLT
Tetrabaena        MGKAKFFLG--AEGAGANNKLVVNMVMGAMMASFAEGLALADKVGLROODV---ID--VIGLG
Auxenochlorella   MGKTSVYLG--ETGSAAKNKLVVNMIMGSMMAALAEGLGLSKKAGLROEDL---LA--ILNOS
Arabidopsis       LGKRSFYLG--OVGNGAKNKLIVNMIMGSMMNAFSEGLVLADKSGLSSDTL---LD--ILDLG
Zea               LGKKSFFLG--EIGNGAKNKLVVNMVMGSMMNSLSEGLCLADKSGLSPOTL---LD--VLDLG
Fusarium          ITTAAIYCG--PIGSGLKAKYSINLFLVSVTTGLAESVNLARAOGLDLDAF---GO--VLNSC
Symbiodinium      LGKRTFFLG--EVGCGARNKLVINMVMGINMVALSEGLAVGOKAGLOGKDI---LE--VIKEG
Tolypocladium     ICKNTITLANQPCEOASLLKLIGNFIRFSTIEVLCEASVLSEKVGLSOETLSKFVE--LLIPG
Beauveria         VTKSAVYCGA-EIGMGLRTKFATNLFLITMTAGLAESMNLARAOGLDVDAF---GO--VLDAG
Geobacter         FGLNIVHVG--NVGDATKNKFIVNLVOAELIOALAEGLVFGEKMGFAADKI---LE--VLDSG
Alpha             LGEGVSYVG--DGELARMVKICHNILLGVVTOSLAEITVLAEKGGVKRHAF---LD--FINKS
Streptomyces      LG-DPLHLGT-EPGLASLYDAALLGLMWSAMTGWLHGTALVGAEGTSATAF---TPVAIRWLG
Lactobacillus     FGT-VKYVG--EOGNGAKLKLITNVAIMAAEAGIRETLDLADAYDIDYOTT---LD--LLOMG
Propionibacteri   MGI-IRYVG--KSGNGVAYKLITNVAIMAAEAGLREALDLADELDMDYELS---LD--LMSVG

cons              .       .       .                    .   .                  .
```

Figure 9. (A) The differential flexibilities of the wild-type protein (black line) compared to the M437 K mutant (red line), calculated on the equilibrated parts of respective MD trajectories. The blue shading indicates the location of the mutation, while the yellow box indicates the region where significant sequence variability (and flexibility) can be found. (B) The analysis of sequences on the basis of their evolutionary relationships: the transition from K to M in higher organisms is highlighted. The boxes and yellow background correspond to the shaded areas in subfigure A.

ability of our Energy Decomposition Method to capture such differences should be related to the fact that the algorithm specifically focuses on the stabilization part of the internal interaction energy.

Effects of the Mutations on the Tetramer. We ran the simulations of the wild-type and mutated tetramer to identify possible differences in the hydrogen bonding network and in the energetic profile of the protein. We did not expect to observe dramatic conformational changes in the tetramer simulations: such changes would expectedly occur on time scales which far exceed the simulation times that are currently feasible. However, the resulting models may give information on how the M437 K mutation can disrupt or decrease the stability of an assembly, if formed at all. Despite these caveats, hydrogen bond analysis of

DOI: 10.1021/acs.jcim.9b00588
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

the tetramer models shows interesting differences between the two sequences, in particular with regards to interchain hydrogen-bonds. The number and persistency of interchain H-bonds formed by the wild-type sequence appears to be consistently higher than that of the mutant (Figure 8, only chain A is exemplified for simplicity). This result indicates that, once organized in the structure intended for supramolecular assembly, the wild-type protein is able to productively establish a higher number of intermolecular interactions with other homologous monomers. This feature mirrors the ability of the wild-type sequence to populate the native three-dimensional structure, stabilized by a diffuse network of pair interactions and in general by a lower flexibility in the tetramerization domain. By contrast, the mutation destabilizes the conformation of the monomer: the internal flexibility and less diffuse network of stabilizing interactions facilitate the sampling of alternative conformations that are not optimal for stabilizing a supramolecular assembly.

Evolutionary Analysis Suggests That the K to M Mutation Imparts NPAC a Different Function in Animals. Energy decomposition analysis indicates that the M437 K mutation occurs at a loosely packed region confined between two separated and ordered scaffolds where most of the stabilizing interactions (and therefore stabilization energy) concentrate. While in the wild-type, the region around Met437 shares stabilization determinants with the two N- and C-terminal domains, the M437 K mutant appears to be highly polarized with the nuclei of interactions stabilizing the well-separated N- and C-terminal domains. It was previously proposed that fold polarity can be linked to functional innovation in different protein families[22,23] and that folds associated with multiple functions such as, e.g., the TIM barrel, show significant polarity. In these systems, the regions containing functionally oriented residues are typically plastic. In this context, polarity reverberates in the possibility to introduce mutations, particularly in flexible regions, that may have limited effects on the stability of the monomeric protein, while favoring the acquisition of new functions. Interestingly, in the Lys437-containing protein, the amino acid undergoing mutation is located in a stretch that is significantly more flexible than the one of the Met437-containing counterpart, as shown by the root-mean-square fluctuation (RMSF) analysis (see Figure 9A, whereby the boxed areas represents the location of residue 437 and its immediate surroundings).

We thus asked whether the observed polarization between stabilization nuclei and plastic regions in the Lys437-containing protein compared to the Met437-containing protein could be reconnected with evolutionary traits indicative of the emergence of a new function from an initial ancestor. To proceed along this route, we first analyzed all the sequences corresponding to the NPAC domain present in the Uniprot database, and started by aligning sequences as divergent as possible (initially we considered for instance 2 mammalian, 2 nonmammalian animals, 2 bacterial, 2 plants, 2 monocellular eukaryotes, 2 archea, etc.). We then enlarged the data set by adding 1 sequence from each of the aforementioned organisms in an iterative fashion to obtain a supervised alignment representative of different organisms with different levels of complexity. The sequence alignment was then fed to the MEGA (Molecular Evolutionary Genetics Analysis, https://megasoftware.net/) software to build the corresponding evolutionary tree (Figure S2).

The evolutionary tree information was next used to compare the representative sequences of branches that originate from a certain common node. The end result of this procedure shows that the protein evolved into a noncatalytic form in higher organisms (Figure 9B). Such form also has a fundamentally different function, dedicated to a structural scaffolding role, rather than to performing enzymatic transformations of substrates.

Overall, based on these data, it is tempting to hypothesize that the concentration of most of the stabilizing interactions at the N- and C- terminal domains in the Lys437-protein favored innovability of the functional properties of NPAC. It can be expected that if mutations primarily perturb (active site) residues that are weakly energetically coupled (or even uncoupled) to the stability scaffolds and are thus significantly plastic, they may not have a disruptive effect on the general fold and favor evolvability. In other words, the low coupling between active-site/stability-core can favor functional residues to vary and thus reverberate in the acquisition of new functions. In this framework, the fold of NPAC appears to be evolvable. Furthermore, it is interesting to note that on the one hand the evolution from Lys to Met is specific for pluricellular eukaryotes (fungi such as Tolypocladium, Beauveria, Fusarium and Streptomices feature a charged residue), and on the other hand that once Lys has evolved to Met in higher organisms, the region connecting the two domains becomes energetically important for stabilization and coupled to the N- and C-scaffolds, with a significant decrease in flexibility. This suggests that further mutations would be energetically destabilizing and that once the protein reached this stage it may not further evolve.

Consistent with this picture, from the conformational dynamics point of view, the transition from the charged Lys437 in lower organisms to the neutral and hydrophobic Met437 determines an overall rigidification of the protein that aptly preorganizes it for tetramerization. On the other hand, the K-variant is characterized by higher internal flexibility, which may be associated with the necessity for the protein to dynamically adapt to substrates and cofactors to bind them efficiently for subsequent enzymatic transformations.

■ CONCLUSIONS

In this paper, we have carried out an extensive comparative analysis to shed light on the molecular determinants of the functional differences between M- and K-containing versions of the NPAC proteins. We have shown how a single mutation in a particular region of a protein can determine strong consequences in terms of dynamics, stability and ability to assemble in supramolecular complexes. In our model, the stability and lower flexibility of the Met-containing protein structurally preorganizes the monomer for tetramerization, whereas the mutation of Met to Lys determines an overall different dynamic behavior, increases flexibility and favors those conformations that are not optimal for tetrameric assembly. In our model, this is the framework to rationalize the emergence of the Met437 version of the protein, whereby the new function shifts from the catalytic one to the scaffolding and tetramerization one. In conclusion, our work elucidates the common and differential molecular determinants that underlie the relationships between NPAC sequence, structure, dynamics and assembly, while providing a novel perspective in the analysis of the evolutionary factors that determine functional innovability in proteins.

DOI: 10.1021/acs.jcim.9b00588
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

MATERIAL AND METHODS

System Preparation and MD Simulations. Simulations were started from crystal structure with Protein Data Bank code 2uyy.

Missing atoms and side chains where added using Maestro (v. 2015−4) protein preparation wizard. Methionine to Lysine mutation was also performed using Maestro.[24]

The simulations were run for both WT and mutated (M437 K) monomer (subunit A) and the tetramer for a total of 4 different systems. All system where minimized and thermalized in 6 subsequent steps ($T = 50$, 100, 150, 200, 250, and 300 K, respectively).

All MD simulations where run in 3 independent replicas of 500 ns each (1.5 μs in total per system), using a Langevin thermostat, integration step of 0.002 ps and periodic boundary conditions.

Specifically, MD simulations were performed using Amber16 pmemd.CUDA with the all atom ff14SB force field under periodic boundary conditions.[25]

The triclinic simulative box, filled with TIP3P water molecules[26] and rendered electroneutral by addition of 3 Na+ (WT monomer), 2 Na+ (M437 K mutant monomer), 12 Na+ (tetramer WT),and 8 Na+ (tetramer M437 K mutant) counterions consists of a final number of atoms of about 38 300 (monomer WT and M437 K mutant), 121 000 (tetramer WT, tetramer M437 K mutant) particles for each system. The systems were then subjected to a round of minimization of 10 000 steps of steepest descend followed by 10 000 steps of conjugate gradient. Relaxation of water molecules and thermalization in NPT environment were carried out for 1.2 ns at 1 fs time-step. In particular, six runs of 200 ps each were carried out increasing the temperature of 50 K at each step, starting from 50 to 300 K. The systems were then simulated with a 2 fs time-step in periodic boundary conditions in the NVT ensemble,[27] using a cutoff of 8 Å for the evaluation of short-range nonbonded interactions and the Particle Mesh Ewald method for the long-range electrostatic interactions.[28] The temperature was kept constant at 300 K with Langevin thermostat. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm.[29] The atomic positions were saved every 10 ps for analysis.

The equilibrated parts of the trajectories were used for subsequent analyses. Equilibration of the trajectories was checked by monitoring the equilibration of the RMSD with respect to the initial structure and of the internal protein energy. The equilibrated parts of each trajectory for each mutant (monomer or tetramer) were next combined into a metatrajectory, which was subsequently used for all the reported characterizations.

Classical structural analyses were carried out with the tools in the Gromacs 4.5.5 package or with code written in-house.

Distance Fluctuation. To characterize the impact of the mutation on the internal dynamics of NPAC, we made use of the previously introduced distance fluctuation (DF) analysis.[13,14] For each MD trajectory in different bound states, we computed on the combined meta trajectory made from the equilibrated parts of the different repeats (time interval 50−500 ns), the matrix of distance fluctuations, in which each element of the matrix corresponds to the DF parameter:

$$DF_{ij} = \left\langle (d_{ij} - \left\langle d_{ij} \right\rangle)^2 \right\rangle \tag{2}$$

where d_{ij} is the (time-dependent) distance of the centers of mass of the side chains of amino acids i and j and the brackets indicate the time-average over the trajectory. This parameter is invariant under translations and rotations of the molecules and, unlike the covariance matrix, does not depend on the choice of a particular protein reference structure. The DF matrix, can be used to assess the intrinsic flexibility of proteins, and how it changes upon mutation.

The DF was calculated for all pairs of residues during the trajectories. This parameter characterizes residues that move in a highly coordinated fashion, and it is actually able to reflect the presence of specific coordination patterns and quasi-rigid domains motion in the protein of interest. In particular, pairs of amino acids belonging to the same quasi-rigid domain are associated with small distance fluctuations and vice versa.

Difference Matrices. To identify relevant differences in the internal dynamics of different mutants, we calculated the difference matrix by subtracting the values of the DF-matrix of the mutant from the DF-matrix of the WT.

If the value was larger than the mean value for the matrix plus 2 times the standard deviation, or if it was lower than the mean value of the matrix minus 2 times the standard deviation, the value was accepted and kept, otherwise it was set to zero. A negative value means larger flexibility for the mutant. A positive value indicates lower flexibility for the mutant.

Clustering. To cluster the simulation we used Biki (v. 1.3.5, www.bikitech.com), with the standard clustering algorithm. For clustering Biki uses the k-medoid algorithm, which can be briefly described as follows: suppose that n conformations visited during the trajectory can be grouped into $k = 5$ ($k < n$) clusters, where k is predefined. The optimal k value for defining meaningfully different conformations is set on the basis of different tests with different values of k. The RMSD between structures was used here as a (dis)similarity distance measure.[30] In the case of WT, one dominant cluster is identified, with a global conformation similar to the initial crystal structure. In the case of the mutant, the first two clusters share about 50 and 30% of the visited conformations.

Pocket Tracking. The protein pocket network was calculated using Biki (v. 1.3.5), using a skip of 100 frames for the monomer and 500 for the tetramer. The pocket tracker algorithm (Pocketron) creates a similarity matrix that compares each pocket of each frame to the other stored ones.[12] It is also able to identify events of merging and splitting of different pockets using the Jaccard index.

Energy Decomposition. The energy decomposition matrix was calculated using the Energy Decomposition Algorithm previously developed by us,[16−18,31−34] to study the energetic (stability) signature associated with the folded state of globular proteins. This approach proved able to identify the residues that contribute the most to determine the stability of a certain protein 3D organization.

The Energy Decomposition Method is based on the calculation of the interaction matrix Mij, which is determined by evaluating average, inter-residue, nonbonded (van der Waals and electrostatics) interaction energies between residue pairs, calculated over all structures visited during an MD trajectory. For a protein of N residues, this calculation yields an $N \times N$ matrix of pair couplings m_{ij} such that the total average non bonded energy of the protein is given by the sum over the matrix entries. We showed that, after diagonalizing the matrix M_{ij}, one can approximate pair couplings using the first eigenvalue lambda and eigenvector w:

DOI: 10.1021/acs.jcim.9b00588
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

$$E_{MD} = \sum_{i,j=1}^{N} m_{ij} = \sum_{i,j=1}^{N} \sum_{k=1}^{N} \lambda_k w_i^k w_j^k \approx \sum_{i,j=1}^{N} \lambda_1 w_i^1 w_j^1 \tag{3}$$

The eigenvector profile reports on the single residue contributions to the essential stabilization energy.

In previous papers, it was demonstrated that the separation between the first eigenvalue λ_1 and the remaining ones reverberates the presence of a well-defined folding nucleus that is generally associated with a stable fold. In other words, if we define $\Delta\lambda = \lambda_2 - \lambda_1$ and λ as the average separation between all remaining eigenvalues, the higher the value of the ratio $\Delta\lambda/\lambda$, the stronger the residues of the nucleus are mutually interacting.

Dynamic Domains. The presence of dynamic domains and molecular hinges was investigated using the Web server DynDom (http://dyndom.cmp.uea.ac.uk/dyndom/main.jsp).

DynDom defines protein domains, hinge axes and amino acid residues involved in the hinge motions. The algorithm works by carrying out a best fit of the two conformations and defining possible rotation vectors. On the basis of the theorem of Chasles, which states that the general displacement of a rigid body is a screw motion, interdomain screw axes are determined and location of the interdomain screw axis indicates the hinge location and kind of motion allowed by the interdomain connections.[11]

The structures used to identify dynamics domains and molecular hinges where the two most populated of each cluster in the different simulations.

Hydrogen Bonds. Hydrogen bonds were calculated using VMD tool. Standard settings (donor−acceptor distance: 3 Å, angle cutoff: 20°) were used.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00588.

> Time dependent evolution of RMSD in the different simulated systems; Evolutionary tree of the NPAC domain sequence, obtained from the alignment described in the main text (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: andrea.mattevi@unipv.it.
*E-mail: g.colombo@unipv.it.

ORCID

Mariarosaria Ferraro: 0000-0003-3982-953X
Andrea Mattevi: 0000-0002-9523-7128
Giorgio Colombo: 0000-0002-1318-668X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Professor Italo Oliviero Carugo, Department of Chemistry of the University of Pavia, for precious advice on sequence analysis and for fruitful discussions. This work was supported by PRIN 2015 (Prot. 20152TE5PK) (A.M.) and by PRIN 2015 (Prot. 2015JTL4HL)

■ REFERENCES

(1) Fu, J.; Yang, Z.; Wei, J.; Han, J.; Gu, J. Nuclear Protein Np60 Regulates P38 Mapk Activity. *J. Cell Sci.* **2006**, *119*, 115−123.

(2) Fang, R.; Chen, F.; Dong, Z.; Hu, D.; Barbera, A. J.; Clark, E. A.; Fang, J.; Yang, Y.; Mei, P.; Rutenberg, M.; Li, Z.; Zhang, Y.; Xu, Y.; Yang, H.; Wang, P.; Simon, M. D.; Zhou, Q.; Li, J.; Marynick, M. P.; Li, X.; Lu, H.; Kaiser, U. B.; Kingston, R. E.; Xu, Y.; Shi, Y. G. Lsd2/Kdm1b and Its Cofactor Npac/Glyr1 Endow a Structural and Molecular Model for Regulation of H3k4 Demethylation. *Mol. Cell* **2013**, *49*, 558−570.

(3) Zhang, J.; Sun, N.; Guo, W.; Wu, X.; Yang, X.; Jin, H.; Zhang, Y.; Wu, X.; Zhang, F.; Hu, L.; Hu, H.; Gao, Y. Identification of Npac as a Novel Biomarker and Regulator for Hepatocellular Carcinoma. *J. Cell. Biochem.* **2019**, *120*, 8228−8237.

(4) Fang, R.; Barbera, A. J.; Xu, Y.; Rutenberg, M.; Leonor, T.; Bi, Q.; Lan, F.; Mei, P.; Yuan, G.-C.; Lian, C.; Peng, J.; Cheng, D.; Sui, G.; Kaiser, U. B.; Shi, Y.; Shi, Y. G. Human Lsd2/Kdm1b/Aof1 Regulates Gene Transcription by Modulating Intragenic H3k4me2Methylation. *Mol. Cell* **2010**, *39*, 222−233.

(5) Marabelli, C.; Marrocco, B.; Pilotto, S.; Chittori, S.; Picaud, S.; Marchese, S.; Ciossani, G.; Forneris, F.; Filippakopoulos, P.; Schoehn, G.; Rhodes, D.; Subramaniam, S.; Mattevi, A. A Tail-Based Mechanism Drives Nucleosome Demethylation by the Lsd2/Npac Multimeric Complex. *Cell Rep.* **2019**, *27*, 387−399.

(6) Zhang, Y.; Zheng, Y.; Quin, L.; Whang, S.; Buchko, G. W.; Garavito, R. M. Structural Characterization of a B-Hydroxyacid Dehydrogenase from GeobacterSulfurreducens and GeobacterMetallireducens with Succinic Semialdehyde Reductase Activity. *Biochimie* **2014**, *104*, 61−69.

(7) Rao, S. T.; Rossmann, M. G. Comparison of Super-Secondary Structures in Proteins. *J. Mol. Biol.* **1973**, *76*, 241−256.

(8) Fei, J.; Ishii, H.; Hoeksema, M. A.; Meitinger, F.; Kassavetis, G. A.; Glass, C. K.; Ren, B.; Kadonaga, J. T. Ndf, a Nucleosome-Destabilizing Factor That Facilitates Transcription through Nucleosomes. *Genes Dev.* **2018**, *32*, 682−694.

(9) Sellés Vidal, L.; Kelly, C. L.; Mordaka, P. M.; Heap, J. T. Review of Nad(P)H-Dependent Oxidoreductases: Properties, Engineering and Application. *Biochim. Biophys. Acta, Proteins Proteomics* **2018**, *1866*, 327−347.

(10) Njau, R. K.; Herndon, C. A.; Hawes, J. W. New Developments in Our Understanding of the B-Hydroxyacid Dehydrogenases. *Chem.-Biol. Interact.* **2001**, *130−132*, 785−791.

(11) Poornam, G.; Matsumoto, A.; Ishida, H.; Hayward, S. A Method for the Analysis of Domain Movements in Large Biomolecular Complexes. *Proteins: Struct., Funct., Genet.* **2009**, *76*, 201−221.

(12) La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **2017**, *3*, 946−960.

(13) Morra, G.; Potestio, R.; Micheletti, C.; Colombo, G. Corresponding Functional Dynamics across the Hsp90 Chaperone Family: Insights from a Multiscale Analysis of Md Simulations. *PLoS Comput. Biol.* **2012**, *8*, No. e1002433.

(14) Moroni, E.; Agard, D. A.; Colombo, G. The Structural Asymmetry of Mitochondrial Hsp90 (Trap1) Determines Fine Tuning of Functional Dynamics. *J. Chem. Theory Comput.* **2018**, *14*, 1033−1044.

(15) Paladino, A.; Marchetti, F.; Ponzoni, L.; Colombo, G. The Interplay between Structural Stability and Plasticity Determines Mutation Profiles and Chaperone Dependence in Protein Kinases. *J. Chem. Theory Comput.* **2018**, *14*, 1059−1070.

(16) Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G. Understanding the Determinants of Stability and Folding of Small Globular Proteins from Their Energetics. *Protein Sci.* **2004**, *13*, 113−124.

(17) Morra, G.; Colombo, G. Relationship between Energy Distribution and Fold Stability: Insights from Molecular Dynamics Simulations of Native and Mutant Proteins. *Proteins: Struct., Funct., Genet.* **2008**, *72*, 660−672.

(18) Genoni, A.; Morra, G.; Colombo, G. Identification of Domains in Protein Structures from the Analysis of Intramolecular Interactions. *J. Phys. Chem. B* **2012**, *116*, 3331−3343.

(19) Morra, G.; Meli, M.; Colombo, G. How the Ligand-Induced Reorganization of Protein Internal Energies Is Coupled to Conformational Events. *J. Chem. Theory Comput.* **2018**, *14*, 5992−6001.

(20) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064−4092.

(21) Lee, M. R.; Duan, Y.; Kollman, P. A. Use of Mm-Pb/Sa in Estimating the Free Energies of Proteins: Applications to Native, Intermediates and Unfolded Villin Headpiece. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 309−316.

(22) Dellus-Gur, E.; Toth-Petroczy, A.; Elias, M.; Tawfik, D. S. What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-Offs. *J. Mol. Biol.* **2013**, *425*, 2609−2621.

(23) Toth-Petroczy, A.; Tawfik, D. S. The Robustness and Innovability of Protein Folds. *Curr. Opin. Struct. Biol.* **2014**, *26*, 131−138.

(24) Maestro *Schrödinger Release 2019−2: Maestro*; Schrödinger, LLC: New York, 2019.

(25) Case, D. A.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P. L. C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber*; University of California: San Francisco, CA, 2017.

(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(27) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(28) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(29) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the Shake and Rattle Algorithms for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(30) Park, H.-S.; Jun, C.-H. A Simple and Fast Algorithm for K-Medoids Clustering. *Expert Systems with Applications* **2009**, *36*, 3336−3341.

(31) Ragona, L.; Colombo, G.; Catalano, M.; Molinari, H. Determinants of Protein Stability and Folding: Comparative Analysis of Beta-Lactoglobulins and Liver Basic Fatty Acid Binding Protein. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 366−376.

(32) Colacino, S.; Tiana, G.; Broglia, R. A.; Colombo, G. The Determinants of Stability in the Human Prion Protein: Insights into the Folding and Misfolding from the Analysis of the Change in the Stabilization Energy Distribution in Different Condition. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 698−707.

(33) Colacino, S.; Tiana, G.; Colombo, G. Similar Folds with Different Stabilization Mechanisms: The Cases of Prion and Doppel Proteins. *BMC Struct. Biol.* **2006**, *6*, 17.

(34) Morra, G.; Baragli, C.; Colombo, G. Selecting Sequences That Fold into a Defined 3d Structure: A New Approach for Protein Design Based on Molecular Dynamics and Energetics. *Biophys. Chem.* **2010**, *146*, 76−84.

Supporting Information

The Impact of Mutations on NPAC Structural Dynamics: Mechanistic Insights from MD Simulations

Marco Montefiori,[1] Simona Pilotto,[2] Chiara Marabelli,[2] Elisabetta Moroni,[1] Mariarosaria Ferraro,[1] Stefano A. Serapian,[3] Andrea Mattevi,[2,*] Giorgio Colombo[1,3,*]

1) ICRM-CNR, Via Mario Bianco 9, 20131 Milano, Italy.
2) Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9, 27100 Pavia, Italy
3) University of Pavia, Department of Chemistry, V.le Taramelli 12, 27100 Pavia, Italy.

Corresponding authors

*E-mail: andrea.mattevi@unipv.it; g.colombo@unipv.it

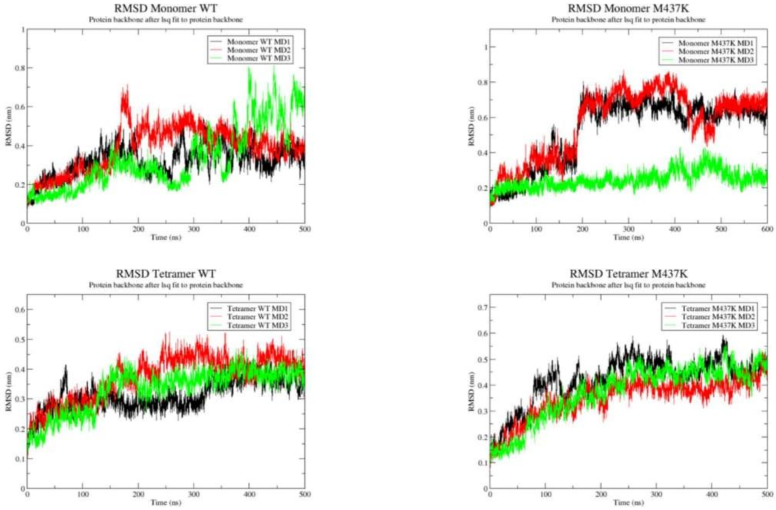


Figure S1. Time evolution of backbone RMSD from the original crystal structure. All residues are considered in the calculation.

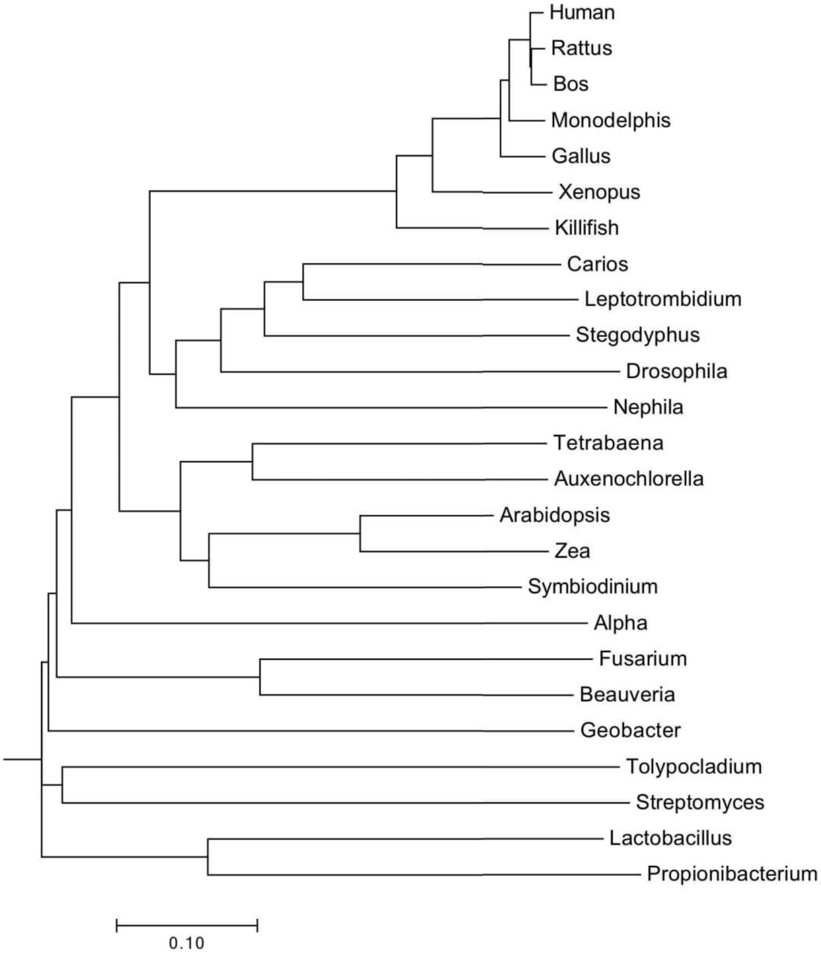


Figure S2. The evolutionary tree of the NPAC sequences obtained via the alignment procedure described in the main text.

Research outputs

1. Scientific publications

- Pilotto S, Speranzini V, Marabelli C, Rusconi F, Toffolo E, Grillo B, Battaglioli E, Mattevi A. **LSD1/KDM1A mutations associated to a newly described form of intellectual disability impair demethylase activity and binding to transcription factors**. *Hum Mol Genet. 2016;25(12):2578-2587.*

- Marabelli, C., Marrocco, B., Mattevi, A. **The growing structural and functional complexity of the LSD1/KDM1A histone demethylase**. *Curr. Opin. Struct. Biol. 2016, 41: 135-144.*

- Speranzini V, Ciossani G, Marabelli C, Mattevi A. **Probing the interaction of the p53 C-terminal domain to the histone demethylase LSD1.** *Arch Biochem Biophys. 2017,(17)30403-4.*

- Marabelli C, Marrocco B, Pilotto S, Chittori S, Picaud S, Marchese S, Ciossani G, Forneris F, Filippakopoulos P, Schoehn G, Rhodes D, Subramaniam S, Mattevi A. **A tail-based mechanism drives nucleosome demethylation by the LSD2/NPAC multimeric complex.** *Cell Reports 2019, 27: 1-13.*

- Montefiori M, Moroni E, Pilotto S, Carugo O, Marabelli C, Mattevi A, Colombo G. **The impact of mutations on NPAC structural dynamics: mechanistic insights from MD simulations.** *Journal of Chemical Information and Modeling, 2019, in press.*

2. Conference posters

- Simona Pilotto, Chiara Marabelli, Biagina Marrocco, Andrea Mattevi. **Nucleosome recognition by flavin-dependent histone demethylases 1 and 2.** *6th Murnau Conference on Structural Bology. September 14th-17th, 2016. Murnau, Germany.*

- Simona Pilotto, Biagina Marrocco, Chiara Marabelli, Sara Marchese, Sara Sandin, Daniela Rhodes and Andrea Mattevi. **LSD2/N-PAC reveal new insights in nucleosome recognition.** *The Gordon Research Conference on Chromatin Structure and Function. July 22-27th, 2018. Grand Summit Hotel at Sunday River, Newry, ME, US.*

- Chiara Marabelli, Biagina Marrocco, Simona Pilotto, Sagar Chittori, Sarah Picaud, Sara Marchese, Giuseppe Ciossani, Federico Forneris, Panagis Filippakopoulos, Guy Schoehn, Daniela Rhodes,Sriram Subramaniam, Andrea Mattevi. **A tail-based mechanism drives nucleosome demethylation by LSD2/NPAC.** *The 2019 SIBBM "Frontiers in Molecular Biology" Seminar, June 11-13th, 2019. Opificio Golinelli, Bologna, Italy.*

3. Oral presentations

- **Molecular mechanisms of the chromatin and drug design.** First Joint meeting Pavia-Pisa. February 11-12th, 2019. Collegio Volta, Pavia

- **Nucleosome recognitino by histone demethylases.** *Resonance in Biology meeting. September 13-14th, 2019. Volta college, Pavia, Italy.*

- **Nucleosome recognitino by histone demethylases.** *SIB 60th meeting. September 18-20th, 2019. Lecce, Italy.*