Expanding the NMR toolbox for nucleosome studies

Ulric B. le Paige

ISBN 978-90-393-7224-1

Doctoral thesis
Expanding the NMR toolbox for nucleosome studies

Ulric B. le Paige
Utrecht university, NMR Spectroscopy Research Group,
Bijvoet Center for Biomolecular Research
Department of Chemistry, Faculty of Science
Utrecht University, the Netherlands
December 2019

Cover design : Ulric B. le Paige. PDBid : 1KX5
Layout design : Ulric B. le Paige & ProefschriftMaken

Expanding the NMR toolbox for nucleosome studies

Uitbreiding van de NMR toolbox voor studies aan nucleosomen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
maandag 9 december 2019 des middags te 12.45 uur

door

Ulric Bernard Agnès Marie Ghislain Roch le Paige

geboren op 14 april 1991 te Luik, België

Promotor:
Prof. dr. M. Baldus

Copromotor:
Dr. H. van Ingen

Table of contents

List of Abbreviations

2-ME – 2-mercaptoethanol

bp – base pair

CBB – Coomassie Brilliant Blue

CP – Cross Polarisation

CPA – Composite Particle Approach

Cryo-EM – Cryogenic Electron Microscopy

CSP – Chemical Shift Perturbation

Dm – Drosophila melanogaster

DNA – Deoxyribonucleic acid

DREAM – Dipolar Recoupling Enhanced by Amplitude Modulation

DTT – Dithiothreitol

E. coli – Escherichia coli

EMSA – Electrophoretic mobility shift assay

FRET – Förster Resonance Energy Transfer

FWHM – Full Width at Half Maximum

INEPT – Insensitive Nuclei Enhanced by Polarisation Transfer

IPTG – IsoPropyl β-D-1-ThioGalactopyranoside

LB – Lysogeny Broth

MAS – Magic Angle Spinning

Nle – Norleucine

NMR – Nuclear Magnetic Resonance

PAGE – PolyAcrylamide Gel Electrophoresis

PDBid – Protein Data Bank identifier

PTM – Post Translational Modification

RFDR – Radio Frequency Driven Recoupling

SAXS – Small Angle X-ray Scattering

SDS – Sodium Dodecyl Sulfate

SEM – Scanning Electron Microscopy

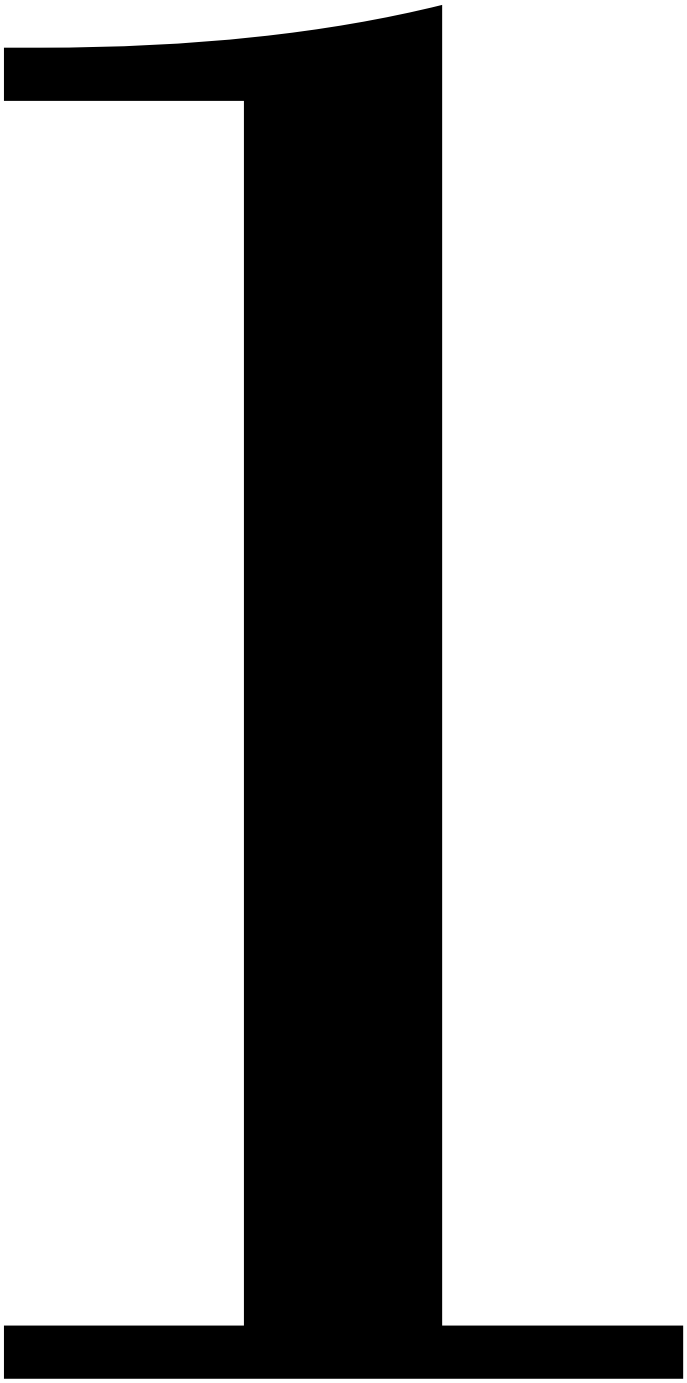
SHL – Super Helix Location

SS-NMR – Solid-State Nuclear Magnetic Resonance

SV-AUC – Sedimentation Velocity in Analytical UltraCentrifugation

TROSY – Transverse Relaxation Optimised SpectroscopY

TSA – Thermal Shift Assay.



Chapter 1

General Introduction

The nucleosome is a protein-DNA complex central to multiple biological functions, including gene regulation, DNA damage repair response, and epigenetic mechanisms. Thus, numerous diseases including cancer are related to a misfunction of processes that involve the nucleosome[1-3]. Thus, a mechanistic and structural description of the nucleosome and its interactions is required to properly understand their role in health and disease, both from a fundamental and therapeutic-oriented perspective. Since the first high-resolution structure of the nucleosome, extensive efforts from all fields of structural biology have allowed increasingly deep understanding of the molecular mechanisms of nucleosome-based biology[4]. Alongside the crucial contributions of crystallography and cryogenic electron microscopy (cryo-EM), NMR has allowed the unravelling of interactions between various chromatin factors and whole or parts of nucleosomes, and elucidation of their mechanistic aspects.

The nucleosome is the gateway, gatekeeper and key to the DNA

The human nucleus contains 2 meters of DNA, put end to end, containing nearly 24.000 genes, numerous additional regulatory sequences and structurally important regions[5-9]. To properly regulate which parts of the sequence should be accessed, to what degree and at what time, the cell has evolved a number of mechanisms that all converge on the nucleosomes that are formed along the DNA sequence. Being a tight protein-DNA complex, nucleosomes in first instance impede access to the DNA. Positioning nucleosomes within gene bodies or on promoter regions of repressed genes thus helps to repress unwanted transcription[10]. Moreover, the abilities of the cell to regulate the position of nucleosomes along the genome by chromatin remodelling allows to dynamically remove or install this barrier. Furthermore, by post-translational modifications of nucleosomes, specific genomic locations can be marked

to direct and regulate all DNA-templated activities such as genome maintenance and epigenetics[11].

Nucleosome and chromatin structure

The nucleosome is composed of two copies of four highly conserved proteins called histones, namely H2A, H2B, H3 and H4 (Figure 1A, B). The histones share a common three-helix fold in their C-terminus, with a central ~29 residues-long helix flanked with the two other helices orthogonally arranged. The N-terminus is composed of a mostly disordered region. The histones form heterodimeric complexes, where the two histone folds cross on the center of the main α-helix. Two H3-H4 dimers pair in a tetramer via the C-terminal helix of H3. Two H2A-H2B dimers further associate with the tetramer, forming an octamer. The octamer is wrapped in ~146 bp of DNA. Upon forming the octamer and nucleosome, parts of the N- and C-termini of histones H2A and H3 fold in short additional helices and all histones further stabilize the octamer complex with the formation of small di-stranded, two-three residue long interchain β-sheet. The whole complex harbors a pseudo two-fold symmetry with respect to a dyad axis (Figure 1A). In addition to the folded core, the N-terminal domains of the histones protrude out of the complex as disordered extensions. Overall positively charged, the histone octamer harbors a small acidic patch on the surface of the H2A-H2B dimer (Figure 1C)[12-15].

The histones bind and immobilize the DNA by establishing polar contacts with the phosphatidyl diester backbone of the DNA and a few non-polar contacts with the deoxyribose groups. Mostly occurring on the minor-groove-inwards regions of the DNA at super helix locations (SHL) of half-increments (Figure 1D), the polar interactions involve basic and hydroxyl sidechains, backbone amides and helix dipole moments from the histones to either directly bind the phosphate group of the DNA backbone or immobilize a water molecule to form a water-mediated bridge. Furthermore, an arginine sidechain enters the minor groove of the DNA at every turn to anchor the DNA to the histone core, further enforcing the heavily bent

conformation of the DNA[12,13,16,17]. The protruding disordered tails, also positively charged, transiently interact with the neighboring DNA[18-21].

Nucleosomes are additionally able to interact with one another[12,22-24]. Via the interaction of the H4 tail from one nucleosome with the acidic patch of a neighbor, two nucleosomes can stack (Figure 1E). When considering isolated nucleosomes or nucleosomes separated by a sufficiently long length of DNA within one strand, the stacks can end up forming a structure akin to a column[25-28]. When linked by a shorter length of DNA, this 'linker' DNA induces torsional constraints which strongly influences the position of the next nucleosome. Thus, for homogeneously distributed nucleosomes, regular structures can be formed. With a linker length in increments of a complete turn of DNA, corresponding to 10 base pairs (bp), nucleosomes will be positioned parallel and opposite from one another. Then, a zig-zag structure can be formed in which nucleosomes form stacks of four nucleosomes paired two by two. Combined, such stacks can form 30 nm-diameter chromatin fibers[29,30]. Long thought to be the higher-order structure of chromatin compaction, the physiological relevance of such extended 30 nm structure is increasingly debated[22,23,31,32].

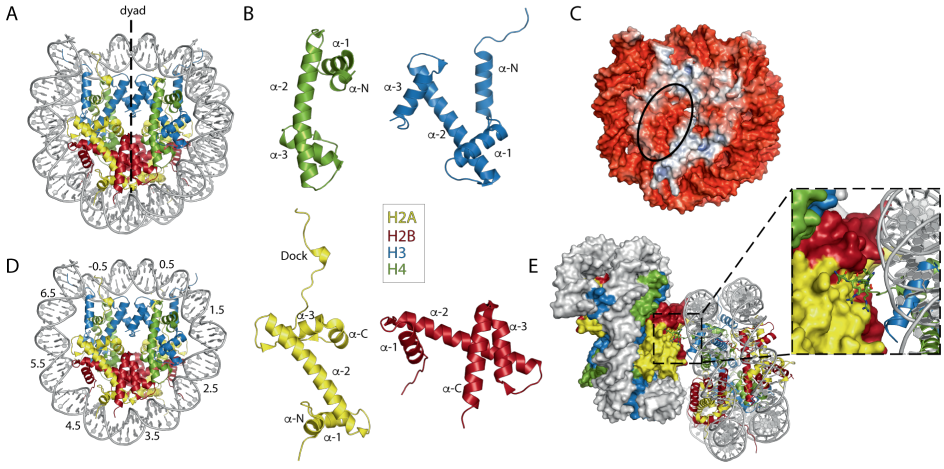


Figure 1: Nucleosome structure. (A) Ribbon representation of the crystal structure of the nucleosome (PDBid 1AOI)[12], showing a two-fold pseudosymmetry axis (dyad). (B) The four histones and their characteristic fold composed of α-1 to α-3. Additional helices on the

N- and C-termini are highlighted. (C) Surface representation of the nucleosome. Electrostatic potential coloring of the surface, calculated by APBS[33], with the acidic patch highlighted by an oval. (D) The histones bind the DNA on minor groove-inwards positions. Numbering of the superhelix locations (SHL) is symmetrical with the dyad. (E) Crystal contacts in PDBid 1AOI involve the H4 tail contacting the acidic patch of a neighboring nucleosome.

The nucleosome as an interaction platform

A multitude of proteins and protein complexes interact with nucleosomes[34]. They typically make use of one to several possible interfaces, namely the histone core, the histone tails and the DNA[4]. In order to regulate these interactions, the nucleosome can be modified, thereby altering one or several interfaces, to favor or hinder specific interactions. The strategies thereof can be classified in three processes.

First, the canonical histones can be replaced by a histone variant through the action of specific chaperones and remodellers. Depending on the variant, various extents of structural changes can be introduced in a specific nucleosome[35]. The presence of certain variants are correlated with specific regions of the genome[36-38] or with genome maintenance[39,40] and embryologic development[41].

Secondly, the nucleosome can be shifted along the DNA, or disassembled and reassembled, thereby changing the exposure of certain DNA sequences. This effect is mediated by nucleosome remodellers and chaperones[42,43]. This can be important to allow for gene transcription, genome replication, and putatively for transcription factors to access their specific sequences in nucleosomal context.

Last but not least, the sidechains of numerous amino acids can be modified by the covalent incorporation of site-specific post translational modifications (PTM). Numerous positions and types of modifications are possible, targeting glycine, lysines, arginines, serines, threonines, tyrosines, cysteine, glutamine and glutamic acids[44], occurring mostly on the tails albeit the histone core is also a significant target. Each modification involves specific writer, reader and eventually eraser proteins that install, bind or remove these modifications, respectively. Additionally, several PTMs are

involved in cross-talk mechanisms in which they cooperatively or antagonistically influence specific interactions with proteins or modify histone tail accessibility[20,45-47].

Methylated lysines: importance of context for structural studies.

One of the most abundant modifications is Lysine methylation which frequently occurs on the flexible histone tails. This flexibility allowed the use of histone tail peptides as cheap, modular and effective substrates for the structural study of tail-binding proteins or protein domains[48-51]. However, for lysines located near or on the nucleosome core, a peptide substrate becomes inefficient as elements from the whole nucleosome become part of the interface[52]. For instance, the methylation of H3 on K79, located on the nucleosome core surface, by the methyltransferases Dot1 depends on the presence of the nucleosomal DNA and the H4 tail[53]. Additionally, the Dot1 enzymes are an example of PTM cross-talk, as ubiquitination of the H2B C-terminus (K123 in *Homo* sapiens) stimulates methylase activity[53,54].

Production of nucleosome substrates bearing post translational modifications.

Structural and biochemical studies strongly benefit from homogeneous samples, and often require milligram amounts of protein of interest. Additionally, most studies relying on NMR spectroscopy require isotopic enrichment. However, producing mg amounts of both homogeneously site-specifically modified and isotope-labelled nucleosomes can be cumbersome. In particular, the incorporation of histone modifications requires dedicated approaches to ensure specific and complete modification. Since this is typically not possible throughout enzymatic modification[55], other techniques have been introduced[53,56], some of which are compatible with isotopic labelling[55,57,58]. Due to its simplicity and the rare occurrence of cysteines in histones, the most attractive of these is to mimic post translational modifications using a thiol-specific chemical modification of a site-specifically introduced cysteine[52,57].

A primer of NMR

Even if NMR is intrinsically a quantum-mechanical phenomenon[59-62], this section attempts to introduce, in layman's terms, the main principles of NMR and how these are exploited in biological studies. Biological NMR being in the vast majority focused on spin ½ nuclei, the following part will omit the multiple complications arising from nuclei with spin of higher absolute values.

Nuclear magnetism and NMR signal

Atomic nuclei have an intrinsic magnetic propensity called spin, being either "up" or "down". When submitted to a magnetic field, a given spin will very weakly align with that field and precess relative to it – as for a spinning top toy under gravity. The frequency of precession, also termed the *Larmor frequency*, depends on the type of nucleus (^{1}H, ^{13}C, ^{15}N, ^{19}F, …) – essentially how magnetically potent it is – and on the strength of the magnetic field.

In NMR spectroscopy, to measure a signal, a sample is put in an intense, as homogeneous as possible, magnetic field. A coil of wire linked to a current detector is positioned directly around the sample, perpendicular to the magnetic field. The detected NMR signal is thus the electric current induced in the coil by the precession of the spins in the sample. However, since the precession of individual spins is asynchronous, their individual effects cancel each other out. To detect a signal, the sample is first excited by applying an intense alternating electric current to the coil – with a frequency corresponding to the Larmor frequency – over a very brief period of time (typically microseconds). This "pulse" of current induces a circularly oscillating magnetic field that is on-resonance with the precessing spins. This pulse forces the net alignment of the spins to rotate from the axis of the main (static) magnet to axis of the detection coil – while the spins are still precessing around the same axis of the main magnet. Thus, the spins now induce a current in the detection coil.

The degree of magnetic alignment for the spins being very weak – despite the magnetic fields in spectrometers typically reaching 10 to 30 T – means the net magnetization

obtained is also weak. To detect enough current in the wire, very high amounts (concentrations typically ranging from 10 to 10,000 μM) of spins need to be measured at the same time. Additionally, most experiments try to focus on the direct detection of the magnetically most potent, stable element, ^{1}H.

The magnetic field alignment of the spins being a very weak effect also implies that once the spins are excited, the restoration of thermal equilibrium is very slow (typically milliseconds to seconds), allowing to detect the signal for a long time. The signal usually fades out faster than we recover the equilibrium state, for reasons described further in the text.

In order to be able to discriminate the various frequencies arising from a sample, sometimes differing by as little as 1 part per billion[63], the detector constantly runs a reference current of specific frequency, called the "carrier" frequency. Upon acquiring signal, the detector mixes the current from the coil with its reference current. What is detected is thus the interference between the carrier and the signals, causing a beat-effect which, coupled to a Fourier transform, allows the very fine discrimination of frequencies. The longer the signal remains detectable, the more accurate the measurement of its frequency and thus the sharper the linewidth of the peak in a spectrum.

The chemical shift

As said above, the Larmor frequency of a spin depends on the magnetic field. To be precise, it depends on the magnetic field the spin *perceives*. This field not only includes the immense field from the superconducting magnet composing the spectrometer, but also fine contributions originating from electron cloud surrounding the observed spin. This causes an additional, small but significant, magnetic field. Variations in the shape and density of the electron cloud thus changes this induced magnetic field, and consequently the Larmor frequency of the nuclear spin. Therefore, the presence and type of chemical bond(s), presence of a nearby strongly (or weakly) electronegative elements, or aromatic rings, all influence the effective magnetic field perceived by a given nuclear spin. Since the chemistry surrounding the spin of interest thus alters a

spin's Larmor frequency, the precise frequency is also called the "chemical shift" (of the Larmor frequency). When two spins have the exact same electronic environment, they will resonate at the same chemical shift. This can happen in symmetrical chemical environments. In such case, the spins are called *chemically equivalent*. The chemical shift is directly proportional to the strength of the main magnetic field, yet in the order of a few millionths of the strength thereof. The frequency of a signal is thus generally expressed in parts per million (ppm) compared to a reference frequency.

Exchange in NMR

When a dynamic process occurs in the sample (such as association and dissociation with an element, molecule or domain, or alternating between conformations), a given spin exchanges between surrounding conditions (ligand-bound/free, a protein domain being open/close, multiple conformations because of flexibility, *etc.*) with each corresponding a chemical shift. If the exchange between two conditions is *fast enough*, one given spin will alternate between the conformations, and thus frequencies, sufficiently often for its chemical shifts to be detected as a single signal corresponding to the population-weighted average. Similar logic applies if the exchange rate is *slow enough*. Then, one given spin will be constrained to one specific chemical shift during the measurement. Thus, over the sample, the signal given by the spin will be split between one peak per present conformation, whose intensities are related to their respective populations.

Whether exchange is fast or slow enough depends on ratio between the difference of chemical shift between the two states (expressed in Hertz) and the rate of their interconversion. When the difference is much smaller than the interconversion rate, exchange is said to be fast. When the difference is much larger, exchange is slow. When the exchange rate and chemical shift difference are similar, exchange is intermediate. In this case the frequency is not well-defined, and thus will yield a broad peak after Fourier transform around the population-weighted average frequency.

In NMR titration experiments, there is exchange between free and ligand-bound protein. As ligand is added, the population of the observed molecule will be shifted

towards a ligand-bound conformation. A spin experiencing a *fast exchange* type of interaction will be seen as a signal changing in frequency in each spectrum (Figure 2A). A spin experiencing a *slow exchange* will be seen as a signal disappearing at the initial, free-state position, while another signal will appear at a chemical shift corresponding to the population of the bound conformation (Figure 2B). In the *intermediate exchange* case, the signal changes in frequency towards the bound state, while broadening and dimming out. After the mid-point, the signal increases again in intensity as it shifts towards the bound state, eventually reaching it at full intensity upon complete saturation (Figure 2C).

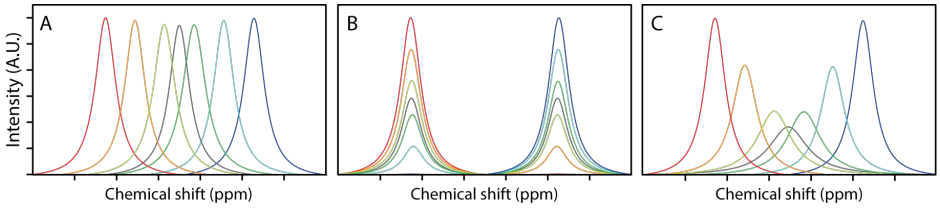


Figure 2: Chemical exchange. Fictive 1D spectra following an NMR titration between free (red) and ligand-saturated (blue) states. The mid-point is colored in grey. (A) Fast exchange case. (B) Slow exchange case. (C) Intermediate exchange case

Couplings: how nuclei influence each other

As discussed above, the main magnetic field allows us to observe the effect of the chemical environment of one nuclear spin through the electron cloud. Similarly, the nuclear spin being essentially a minuscule magnet, one spin can directly influence a nearby other spin in a very fine way.

A spin "X" can be in an either "up" or "down" state, which will give rise to additional contributions to the global magnetic field felt by a nearby spin "Y", and thus either increase or decrease the Larmor frequency of spin "Y". That effect being directly caused by the state of the nuclear spins means that it does not depend on the strength of the main magnetic field but on how strong the magnetic propensities of the "X" and "Y" spins are (thus what atom/isotope they are), and how strongly the spin "Y" feels the spin "X". This effect is reciprocal, and these nuclear-spin-to-nuclear-spin interactions are termed *couplings*. The observed chemical shift of one spin "Y" will thus

be different whether the coupled spin "X", is in an "up" state or "down" state. The difference between the "X-up"-shifted and the "X-down"-shifted frequencies of "Y" is called the coupling constant and is expressed in Hz. There are two ways in which two spins can be coupled: scalar (also called J-) couplings, and dipolar couplings. In a molecule, spins experiencing identical couplings are called *magnetically equivalent*.

Scalar couplings arise from the termed *hyperfine* interaction a nucleus has with its own electrons. Through that interaction, the spin state (up/down) of a nuclear spin "X" influences its electrons. Electrons being shared in chemical bonds, this results in a "X"-spin-state-based perturbation to a neighboring, chemically bonded spin "Y". This translates to an additional magnetic contribution to its Larmor frequency. Being transmitted through the chemical bonds, the strength of the coupling depends on how the two spins are bound. The number of bonds, bond length(s) and angles are thus influencing the coupling strength. In the case of one spin "X" coupled to several magnetically equivalent spins, the various possible combinations of coupled spin states dictate the number and relative intensities of signals detected from the spin "X".

Dipolar couplings arise from the two nuclear spins perceiving each other's induced magnetic field, through space. This is akin to how two bar magnets put one near the other are perturbing each other. Thus, the strength of the perturbation strongly depends on the distance between the two magnets, how magnetically potent they are and the position of one magnet within the field caused by the second. Similarly, the magnitude of the dipolar coupling depends on the atom types and the distance between their magnetic dipoles. This is scaled by the position of spin "X" within the magnetic field from "Y" and thus highly orientation dependent.

In solution, molecules tumble rapidly through Brownian motions. Doing so, they sample all possible orientations relative to the magnetic field. Thus, the dipolar couplings are also rapidly sampling all orientations and thus are thus effectively averaged out. In contrast, for a solid powder, these different orientations are now static. Thus, a specific orientation of the coupled spin pair in one molecule will have its signals differently split, compared to those of another identical molecule with different orientation. This results in a very broad 'powder spectrum' in static solid-

state NMR. In a way, this can be seen as the different orientations being in fast exchange in solution, and in slow exchange in solids.

From individual spin(s) to whole molecules: "atomic resolution"

Most of what was described above assumed only one distinct spin or set of spins. When considering larger molecules such as proteins, the presence of many different spins in the molecule reduces the resolution of the NMR spectrum in two ways. First, the dense network of dipolar coupled protons will broaden the signals, as discussed in a bit more detail below. Second and more obvious, the presence of many different signals will create overlap in a 1D NMR spectrum. This necessitates the use of multidimensional NMR in which spectra contain two or more frequency dimensions such that each signal can encode the Larmor frequencies of multiple, different spins. Usually this is combined with isotopic enrichment during sample preparation to introduce additional magnetically active elements in the sample. This way for instance a 2D NH correlation spectrum can be recorded, in which roughly one signal per amino acid is obtained, encoding both the amide backbone ^{1}H and ^{15}N chemical shift. Since these chemical shifts are very sensitive to the conformation of a protein, the NH spectrum is the workhorse spectrum to evaluate the effect of a perturbation, e.g. ligand binding, on the protein. Provided one can resolve and pinpoint the signals to their corresponding specific atom in the molecule, information at atomic resolution can be obtained in straightforward fashion.

Since the Larmor frequency is influenced in such subtle manner by the protein conformational landscape, its direct use in understanding protein structure and dynamics is limited. While knowledge of backbone chemical shifts can be used to determine secondary structure, more detailed study of protein structure and dynamics requires more sophisticated NMR experiments.

The accurate measurement of NMR spectra typically takes between hours to days. Radio waves being of very low energy, they do not affect the sample. Similarly, the usual aim is to have a sample in biologically relevant conditions of buffer and temperature, or by default in stable conditions. NMR spectroscopy itself is thus

typically a non-destructive technique and an NMR sample can potentially be recovered afterwards for other uses.

Manipulating the NMR signal

As mentioned above, once perturbed (excited) by a pulse, the spins remain perturbed for a relatively long time. This means that instead of just detecting the NMR signal, one can further influence the spins before the acquisition of the signal. Through the proper use of pulses and delays, specific magnetic effects can be suppressed and (re-)added or induced specifically in various NMR experiments. One can even use J-couplings and dipolar couplings to bring an excited spin to excite one other.

One of the main approaches used to transfer magnetization in solution NMR is to exploit J-coupled spins. This allows, for instance, to transfer the magnetization of a ^{1}H spin to a neighboring ^{15}N spin, resulting in a ten-fold enhancement of the ^{15}N magnetization compared to ^{15}N direct excitation. The most frequently used concept and the set of pulses to exploit it have thus been named *Insensitive Nuclei Enhanced by Polarization Transfer* (INEPT)[64] and are the basic building block of most multidimensional NMR experiments.

The main approach to transfer magnetization in solid state NMR relies on dipolar couplings. Akin to two similar pendula hanging from a same rope, one pendulum's oscillation being transferred to the other, then back; dipolar couplings cause one excited spin to transfer its magnetization to a non-excited one back and forth. As said before, the strength of the coupling effect depends on the proximity of the spins, their magnetic potency and the relative orientations compared to their magnetic fields. Two spins – one being excited – would thus exchange magnetization back and forth, with an efficiency that depends on their dipolar coupling. If the two spins are very mobile, their relative orientation with the magnetic field varies over time. If that motion is fast enough, the two spins would end up being statistically in one orientation and its exact opposite during the time scale required for the magnetization to be exchanged, constantly modulating the coupling and thus hindering the magnetization exchange.

The problems with larger molecules

With increasing protein size arises an increasing number of distinct detectable spins. To avoid overcrowding the spectrum with overlapping signals, site-specific isotope-labelling strategies are typically used. This incorporates new targets in the protein for specific magnetization transfers, described above, to be used and thereby select the systems of coupled spins to be observed in a specific spectrum.

More fundamentally, larger proteins pose a significant challenge for solution NMR because their molecular tumbling due to Brownian motions becomes slower. At some point this breaks down the fast exchange averaging of the dipolar interactions. As tumbling slows down, the exchange effectively takes place in the intermediate regime and as a result the signal will broaden. A further compounding problem in large proteins is that the ^{1}H nuclei form a tight, extended network of dipolar coupled spins. This allows a very efficient, uncontrolled magnetization exchange throughout the network, effectively *diluting* the magnetization of a spin of interest away through the protein.

An additional source for broadening comes from the chemical shift itself. As stated earlier, the shape of the electron cloud surrounding one spin in a molecule is not necessarily symmetric. From this follows the fact that each possible orientation of a molecule, compared to the main magnetic field, will yield a different chemical shift. This effect is called *chemical shift anisotropy* (CSA) and its magnitude scales with the magnetic field. Again, in solution the different orientations are averaged out and there is no effect from the anisotropy. When the orientations are no longer averaged in fast exchange, also the CSA results in broadening.

For very large assemblies such as nucleosomes, these broadening effects are so prominent that only the histone tails, which experience fast local tumbling due to their very flexible nature, can be detected while signals from the histone core are unobservable.

NMR on large molecules: solution NMR and methyl-TROSY

To work around this issue, several approaches have been undertaken for biological NMR. Working with small, biologically relevant domains instead of a whole protein can allow to remain within the fast tumbling regime. Extensive deuteration attenuates the 'magnetization dilution' effect of the proton network, thus extending the range of molecular sizes accessible for observation. Indeed, replacing most of the ^{1}H by a 7-fold magnetically weaker ^{2}H strongly affects the corresponding dipolar couplings and their effective range[65]. Increasing the temperature allows for faster molecular tumbling, within the limits of stability of the sample.

Another approach, often combined with deuteration, relies on using molecular symmetry and interference between relaxation effects to counter the signal damping. The various techniques making use of these phenomena are referred to as TROSY (transverse relaxation optimized spectroscopy)[66]. A particularly powerful variant of this approach is methyl-TROSY, which has already been extensively used for the study of high molecular weight complexes[67-71]. In this approach, methyl groups of the proteins (typically Isoleucine, Leucine and Valine, ILV) are specifically ^{13}C labelled and any ^{1}H aside from those methyls is deuterated. Methyl TROSY uses the high mobility and symmetry of methyl groups, specific magnetic properties thereof, and proper pulse sequence to obtain a sharp and intense signal. At the same time, the very high degree of deuteration counters the magnetization diffusion issue. So far, this allowed to perform quantitative studies in up to one megadalton complexes. Through extensive pulse sequence development, it has now become a state-of-the-art approach for detailed characterization of intermolecular interactions and very fine characterization of the dynamics of a protein or complex[72].

NMR on large molecules: solid-state NMR

Most of the aforementioned size problems arise from the too slow tumbling of the molecules of interest. Solid state NMR thus uses the simple concept of physically

turning the sample within the magnet, at a specific angle relative to the magnetic field (called "Magic Angle", 54.74 °) that ensures optimal averaging[73].

A fast-enough MAS (magic angle spinning) is able to re-average most of the orientation-dependent effects (with spinning speeds up to 111 kHz – 6.7 x 10^6 rpm in the newest setups[74]). The most frequently used spins in protein studies with very fast MAS being the amide ^{15}N-^{1}H pair in the peptidic bond, removal of the nearest ^{1}Hα through deuteration effectively removes one the strongest sources of dipolar couplings causing relaxation and thus signal damping[75]. More extensive deuteration further hinders relaxation and increases resolution. Additional pulsing is used to further dampen dipolar couplings and thus increase resolution. This setup allows for sufficient resolution and sensitivity to observe the backbone of most of the protein of interest. Compared to methyl-TROSY, this setup yields one signal per amino acid and thus increases the number of observables to characterize an interaction.

To transfer magnetization, again the two types of couplings are available. J-coupling always allows an efficient transfer, but unless very high spinning rates are used, remaining orientation-dependent effects cause a too strong relaxation for the ^{1}H-dense regions to be properly observed. Typically, only the most flexible regions are effectively observed.

To use dipolar couplings, one must counter the effect of the magic angle spinning. To do so, one can either use rotor-spinning-synchronized pulses to counter the effect of MAS[76] or use long constant pulses[77-79]. The latter option creates an on-resonance magnetic field that is tuned to either bring two spins of interest to similar on-resonance frequencies, or constantly affect the precession in a rhythmic way, tuned to match and counter the MAS frequency. The combined effects of pulse strengths, durations and delays between pulses allow to select for ranges dipolar couplings to reintroduce. In all cases, the overall magnetization will equilibrate between the concerned spins as described above, allowing the further observation or manipulation of newly excited spins. This transfer however only works if molecular motions are not modulating the dipolar coupling too much, through either varying the orientation or distance between the two spins.

This approach thus requires a sample to be sufficiently immobile that the random molecular tumbling does not counter the magic angle spinning and concentrated enough for a signal to be detected. To that extent, multiple approaches have been used, notably lyophilization, native precipitation or (micro-)crystallisation[80], before loading the sample in a cylindric rotor for MAS. One later approach was to simply use soluble protein complexes that, through the very high spinning speeds and resulting internal centrifuge forces obtained in MAS, would form a sediment against the walls of the rotor and thus be *de facto* immobilised[81].

Scope of the thesis

A wealth of methods is available for the study of the nucleosome throughout the range of size scales it is involved in. Crystallography, cryo-EM, small angle X-ray scattering (SAXS) and NMR allow an increasing level of understanding of the structural details on which nucleosome biology relies on, and computational methods allow the integration of further techniques such as mass spectrometry, mutagenesis and Förster resonance energy transfer (FRET)[82]. In this work, we aimed to extend the toolbox of NMR-based techniques available for the study of the nucleosome.

Chapter 2 describes the unexpected discovery that occurred upon aiming to expand the methyl TROSY toolbox. Isotope-labelling the methyl groups of a trimethyl-lysine would allow the obtention of an NMR probe directly at the site of the modification. To that aim, an isotope-labelled version of the compound, used to alkylate a cysteine into a trimethyl lysine analogue, was characterized by NMR. Due to the symmetry of the trimethyl moiety, all methyl protons are chemically equivalent, yet due to the isotope-labelling their magnetic equivalence was broken. This yielded a very striking ^{1}H spectrum showcasing the sometimes unexpected effects of magnetic inequivalence.

The following chapters describe the addition of co-sedimentation coupled with ^{1}H-detected solid-state NMR in the toolbox for the study of nucleosome-protein interactions. This approach allows to increase the number of observables available

from the nucleosome perspective, and at the same time lower the requirements for deuteration.

Chapter 3 describes a proof-of-principle experiment focusing on the interaction of a well-characterized nucleosome-binding peptide. Sedimentation of nucleosomes is shown to yield high-quality spectra, and allowing assignment and secondary structure determination of histone H2A. Co-sedimentation of the peptide resulted in clear and specific changes, permitting the determination of the binding site of the peptide.

In **Chapter 4**, the interaction between the yeast protein Dot1p and the nucleosome was studied through biochemical and NMR methods, including ^{1}H-detected solid state NMR on H3. Similarly to what was described in chapter 3, H3 was assigned and characterized before proceeding to binding experiments. Using a divide and conquer approach, we explored the complex mechanism of Dot1p interaction with the nucleosome.

In **Chapter 5**, we attempted to better characterize the sedimentation method and its limitations. Through careful inspection of our previously published H2A data and submitting that same sample to SAXS analysis, we investigated for the presence of nucleosome superstructures that could hinder further use. Additionally, we used the second PHD finger of CHD4 as a worst-case test of the limits of the co-sedimentation approach.

References

1 Hopcraft, S. E. *et al.* Chromatin remodeling controls Kaposi's sarcoma-associated herpesvirus reactivation from latency. *PLoS Pathog* **14**, e1007267, doi:10.1371/journal.ppat.1007267 (2018).

2 McAnena, P., Brown, J. A. & Kerin, M. J. Circulating Nucleosomes and Nucleosome Modifications as Biomarkers in Cancer. *Cancers (Basel)* **9**, doi:10.3390/cancers9010005 (2017).

3 Lampi, Y. *et al.* Targeted editing of the PSIP1 gene encoding LEDGF/p75 protects cells against HIV infection. *Sci Rep* **9**, 2389, doi:10.1038/s41598-019-38718-0 (2019).

4 Horn, V. & Ingen, H. v. Recognition of nucleosomes by chromatin factors: Lessons from data-driven docking-based structures of nucleosome-protein complexes. **Epigenetics and chromatin**, doi:10.5772/intechopen.81016 (2019).

5 Hunt, S. E. *et al.* Ensembl variation resources. *Database (Oxford)* **2018**, doi:10.1093/database/bay119 (2018).

6 Wang, J. C. Helical repeat of DNA in solution. *Proc Natl Acad Sci U S A* **76**, 200-203, doi:10.1073/pnas.76.1.200 (1979).

7 Ichikawa, Y., Nishimura, Y., Kurumizaka, H. & Shimizu, M. Nucleosome organization and chromatin dynamics in telomeres. *Biomol Concepts* **6**, 67-75, doi:10.1515/bmc-2014-0035 (2015).

8 Schalch, T. & Steiner, F. A. Structure of centromere chromatin: from nucleosome to chromosomal architecture. *Chromosoma* **126**, 443-455, doi:10.1007/s00412-016-0620-7 (2017).

9 Takizawa, Y. *et al.* Cryo-EM structure of the nucleosome containing the ALB1 enhancer DNA sequence. *Open Biol* **8**, doi:10.1098/rsob.170255 (2018).

10 Whitehouse, I., Rando, O. J., Delrow, J. & Tsukiyama, T. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**, 1031-1035, doi:10.1038/nature06391 (2007).

11 Fenley, A. T., Anandakrishnan, R., Kidane, Y. H. & Onufriev, A. V. Modulation of nucleosomal DNA accessibility via charge-altering post-translational modifications in histone core. *Epigenetics Chromatin* **11**, 11, doi:10.1186/s13072-018-0181-5 (2018).

12 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

13 Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *Journal of Molecular Biology* **319**, 1097-1113, doi:10.1016/s0022-2836(02)00386-8 (2002).

14 Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett* **589**, 2914-2922, doi:10.1016/j.febslet.2015.05.016 (2015).

15 McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem Rev* **115**, 2255-2273, doi:10.1021/cr500373h (2015).

16 Chua, E. Y., Vasudevan, D., Davey, G. E., Wu, B. & Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res* **40**, 6338-6352, doi:10.1093/nar/gks261 (2012).

17 Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145-150, doi:10.1038/nature01595 (2003).

18 Korolev, N., Allahverdi, A., Lyubartsev, A. P. & Nordenskiöld, L. The polyelectrolyte properties of chromatin. *Soft Matter* **8**, doi:10.1039/c2sm25662b (2012).

19 Stutzer, A. *et al.* 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, doi:10.1016/j.molcel.2015.12.015 (2016).

20 Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **7**, doi:10.7554/eLife.31481 (2018).

21 Arya, G. & Schlick, T. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J Phys Chem A* **113**, 4045-4059, doi:10.1021/jp810375d (2009).

22 Hsieh, T. H. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108-119, doi:10.1016/j.cell.2015.05.048 (2015).

23 Ricci, M. A., Manzo, C., Garcia-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145-1158, doi:10.1016/j.cell.2015.01.054 (2015).

24 Robinson, P. J., Fairall, L., Huynh, V. A. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A* **103**, 6506-6511, doi:10.1073/pnas.0601212103 (2006).

25 Bertin, A., Mangenot, S., Renouard, M., Durand, D. & Livolant, F. Structure and phase diagram of nucleosome core particles aggregated by multivalent cations. *Biophys J* **93**, 3652-3663, doi:10.1529/biophysj.107.108365 (2007).

26 Mangenot, S., Leforestier, A., Durand, D. & Livolant, F. Phase Diagram of Nucleosome Core Particles. *Journal of Molecular Biology* **333**, 907-916, doi:10.1016/j.jmb.2003.09.015 (2003).

27 Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* **13**, 436-447, doi:10.1038/nrm3382 (2012).

28 Meng, H., Andresen, K. & van Noort, J. Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers. *Nucleic Acids Res* **43**, 3578-3590, doi:10.1093/nar/gkv215 (2015).

29 Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138-141, doi:10.1038/nature03686 (2005).

30 Song, F. *et al.* Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380, doi:10.1126/science.1251413 (2014).

31 Maeshima, K. *et al.* Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *The EMBO Journal* **35**, 1115-1132, doi:10.15252/embj.201592660 (2016).

32 Maeshima, K., Ide, S. & Babokhov, M. Dynamic chromatin organization without the 30-nm fiber. *Curr Opin Cell Biol* **58**, 95-104, doi:10.1016/j.ceb.2019.02.003 (2019).

33 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041, doi:10.1073/pnas.181342398 (2001).

34 Fasci, D., van Ingen, H., Scheltema, R. A. & Heck, A. J. R. Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei. *Mol Cell Proteomics* **17**, 2018-2033, doi:10.1074/mcp.RA118.000924 (2018).

35 Henikoff, S. & Smith, M. M. Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* **7**, a019364, doi:10.1101/cshperspect.a019364 (2015).

36 Tolstorukov, M. Y. *et al.* Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol Cell* **47**, 596-607, doi:10.1016/j.molcel.2012.06.011 (2012).

37 Gaspar-Maia, A. *et al.* MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nat Commun* **4**, 1565, doi:10.1038/ncomms2582 (2013).

38 Regnier, V. *et al.* CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Mol Cell Biol* **25**, 3967-3981, doi:10.1128/MCB.25.10.3967-3981.2005 (2005).

39 Preparing for Transcription: The Role of Histone H2A.Z. *PLoS Biology* **3**, doi:10.1371/journal.pbio.0030413 (2005).

40 Rudnizky, S. *et al.* H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat Commun* **7**, 12958, doi:10.1038/ncomms12958 (2016).

41 Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421-434, doi:10.1038/cr.2011.14 (2011).

42 Yan, L., Wang, L., Tian, Y., Xia, X. & Chen, Z. Structure and regulation of the chromatin remodeller ISWI. *Nature* **540**, 466-469, doi:10.1038/nature20590 (2016).

43 Ayala, R. *et al.* Structure and regulation of the human INO80-nucleosome complex. *Nature* **556**, 391-395, doi:10.1038/s41586-018-0021-6 (2018).

44 Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* **7**, a025064, doi:10.1101/cshperspect.a025064 (2015).

45 Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Current Opinion in Cell Biology* **15**, 172-183, doi:10.1016/s0955-0674(03)00013-9 (2003).

46 Sadakierska-Chudy, A. & Filip, M. A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. *Neurotox Res* **27**, 172-197, doi:10.1007/s12640-014-9508-6 (2015).

47 Gatchalian, J. *et al.* Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nat Commun* **8**, 1489, doi:10.1038/s41467-017-01598-x (2017).

48 Musselman, C. A. *et al.* Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* **423**, 179-187, doi:10.1042/BJ20090870 (2009).

49 Andrews, F. H. *et al.* Multivalent Chromatin Engagement and Inter-domain Crosstalk Regulate MORC3 ATPase. *Cell Rep* **16**, 3195-3207, doi:10.1016/j.celrep.2016.08.050 (2016).

50 Arita, K. *et al.* Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc Natl Acad Sci U S A* **109**, 12950-12955, doi:10.1073/pnas.1203701109 (2012).

51 Cho, H. J. *et al.* GAS41 Recognizes Diacetylated Histone H3 through a Bivalent Binding Mode. *ACS Chem Biol* **13**, 2739-2746, doi:10.1021/acschembio.8b00674 (2018).

52 van Nuland, R. *et al.* Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin* **6**, 12, doi:10.1186/1756-8935-6-12 (2013).

53 Worden, E. J., Hoffmann, N. A., Hicks, C. W. & Wolberger, C. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**, 1490-1501 e1412, doi:10.1016/j.cell.2019.02.002 (2019).

54 Valencia-Sanchez, M. I. *et al.* Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. *Mol Cell* **74**, 1010-1019 e1016, doi:10.1016/j.molcel.2019.03.029 (2019).

55 McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G. & Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453**, 812-816, doi:10.1038/nature06906 (2008).

56 van Hest, J. C. M., Kiick, K. L. & Tirrell, D. A. Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *Journal of the American Chemical Society* **122**, 1282-1288, doi:10.1021/ja992749j (2000).

57 Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003-1012, doi:10.1016/j.cell.2006.12.041 (2007).

58 Zhang, H. & van Ingen, H. Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. *Curr Opin Struct Biol* **38**, 75-82, doi:10.1016/j.sbi.2016.05.008 (2016).

59 Benesi, A. J. *A Primer of NMR Theory with Calculations in Mathematica*. (John Wiley & Sons, Inc., 2015).

60 Levitt, M. H. *Spin Dynamics : basics of nuclear magnetic resonance*. 2nd edition edn, (John Wiley & Sons, Inc., 2008).

61 Keeler, J. *Understanding NMR Spectroscopy*. 2nd edition edn, (John Wiley & Sons, Inc., 2010).

62 Teilum, K., Kunze, M. B., Erlendsson, S. & Kragelund, B. B. (S)Pinning down protein interactions by NMR. *Protein Sci* **26**, 436-451, doi:10.1002/pro.3105 (2017).

63 le Paige, U. B. *et al.* Weak coupling between magnetically inequivalent spins: The deceptively simple, complicated spectrum of a (13)C-labeled trimethylated amine. *J Magn Reson* **278**, 96-103, doi:10.1016/j.jmr.2017.03.016 (2017).

64 Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **101**, 760-762, doi:10.1021/ja00497a058 (1979).

65 Sattler, M. & Fesik, S. W. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **4**, 1245-1249, doi:10.1016/s0969-2126(96)00133-5 (1996).

66 Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-12371, doi:10.1073/pnas.94.23.12366 (1997).

67 Ollerenshaw, J. E., Tugarinov, V. & Kay, L. E. Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* **41**, 843-852, doi:10.1002/mrc.1256 (2003).

68 Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618-622, doi:10.1038/nature05512 (2007).

69 Horn, V. *et al.* Structural basis of specific H2A K13/K15 ubiquitination by RNF168. *Nat Commun* **10**, 1751, doi:10.1038/s41467-019-09756-z (2019).

70 Kato, H. *et al.* Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A* **108**, 12283-12288, doi:10.1073/pnas.1105848108 (2011).

71 Hu, Q., Botuyan, M. V., Cui, G., Zhao, D. & Mer, G. Mechanisms of Ubiquitin-Nucleosome Recognition and Regulation of 53BP1 Chromatin Recruitment by RNF168/169 and RAD18. *Mol Cell* **66**, 473-487 e479, doi:10.1016/j.molcel.2017.04.009 (2017).

72 Rosenzweig, R. & Kay, L. E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu Rev Biochem* **83**, 291-315, doi:10.1146/annurev-biochem-060713-035829 (2014).

73 Andrew, E. R., Bradbury, A. & Eades, R. G. Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed. *Nature* **182**, 1659-1659, doi:10.1038/1821659a0 (1958).

74 Xue, K. *et al.* Limits of Resolution and Sensitivity of Proton Detected MAS Solid-State NMR Experiments at 111 kHz in Deuterated and Protonated Proteins. *Sci Rep* **7**, 7444, doi:10.1038/s41598-017-07253-1 (2017).

75 Mance, D. *et al.* An Efficient Labelling Approach to Harness Backbone and Side-Chain Protons in (1) H-Detected Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **54**, 15799-15803, doi:10.1002/anie.201509170 (2015).

76 Bennett, A. E., Griffin, R. G., Ok, J. H. & Vega, S. Chemical shift correlation spectroscopy in rotating solids: Radio frequency-driven dipolar recoupling and longitudinal exchange. *The Journal of Chemical Physics* **96**, 8624-8627, doi:10.1063/1.462267 (1992).

77 Pines, A., Gibby, M. G. & Waugh, J. S. Proton-Enhanced Nuclear Induction Spectroscopy. A Method for High Resolution NMR of Dilute Spins in Solids. *The Journal of Chemical Physics* **56**, 1776-1777, doi:10.1063/1.1677439 (1972).

78 Nielsen, N. C., Bildsøe, H., Jakobsen, H. J. & Levitt, M. H. Double-quantum homonuclear rotary resonance: Efficient dipolar recovery in magic-angle spinning nuclear magnetic resonance. *The Journal of Chemical Physics* **101**, 1805-1812, doi:10.1063/1.467759 (1994).

79 Verel, R., Ernst, M. & Meier, B. H. Adiabatic dipolar recoupling in solid-state NMR: the DREAM scheme. *J Magn Reson* **150**, 81-99, doi:10.1006/jmre.2001.2310 (2001).

80 Bertini, I. *et al.* On the use of ultracentrifugal devices for sedimented solute NMR. *J Biomol NMR* **54**, 123-127, doi:10.1007/s10858-012-9657-y (2012).

81 Sarkar, R. *et al.* Immobilization of soluble protein complexes in MAS solid-state NMR: Sedimentation versus viscosity. *Solid State Nucl Magn Reson* **76-77**, 7-14, doi:10.1016/j.ssnmr.2016.03.005 (2016).

82 van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725, doi:10.1016/j.jmb.2015.09.014 (2016).

Chapter 2

Weak coupling between magnetically inequivalent spins: the deceptively simple, complicated spectrum of a ^{13}C-labeled trimethylated amine

Contributions:
The Composite Particle Approach were done by B. Smits.
Compound 1 was synthetized by Peter 't Hart (Utrecht University).

Published as:
Ulric B. le Paige, Bauke Smits*, Peter 't Hart, Fons Lefeber, Nathaniel L. Martin & Hugo van Ingen (2017). J Mag Res, 278, 96-103*

Abstract

Magnetic inequivalence of nuclear spins is well known to cause additional splittings that complicate spectral analysis. Here, we present an extraordinary case of magnetic inequivalence, manifested in the 13-spin system of a $^{13}C,^{15}N$-labeled trimethylated amine. All methyl group protons are chemically equivalent due to the molecular symmetry, but not all are magnetically equivalent as they have different $^{1}J_{CH}$ and $^{3}J_{CH}$ couplings. In general, spectra of such a large spin system can be expected to be extremely complicated by the presence of hundreds if not thousands of extra lines, caused by the strong coupling between inequivalent nuclei. Surprisingly, the ^{1}H spectrum presented consists of very few lines, in a pattern of the utmost simplicity. Using sub-spectral analysis, we show that this is due to weak coupling between the magnetically inequivalent nuclei, as a consequence of the particular combination of coupling constants. We find that the $^{4}J_{HH}$ geminal methyl coupling constant is 0.43 Hz and $^{2}J_{CC}$ is ~0 Hz. In addition, we demonstrate that homo-decoupling can be used to transform the spin system to a set of fully equivalent spins, resulting in disappearance of $^{4}J_{HH}$-splittings. We believe this curious case is a highly instructive example of magnetic inequivalence. The spectra may be considered deceptively simple, as fewer lines are observed than one would anticipate. At the same time, the spectra are deceptively complicated, as they can very well be approximated by intuitive reasoning.

Introduction

Scalar coupling between nuclear spins is one of the defining features of solution NMR. It is used as a rich source of information on molecular structure[1], and encountered in the everyday practice of spectroscopists through the fine structure of resonance lines. The exact shape of the multiplet pattern that is generated by scalar coupling depends critically on the ratio of the frequency separation between the two nuclei, $\Delta\nu$, and their coupling constant, *J*. In the weak coupling regime ($\Delta\nu >> J$) simple rules can predict the observed multiplet patterns. Spins that are strongly coupled *($\Delta\nu \lesssim J$)* give rise to more complicated multiplets: the intensity of individual components depends strongly on $\Delta\nu/J$, and in larger spin systems additional, so-called combination lines may be observed. At the extreme end of strong coupling is the coupling between spins with identical chemical shifts ($\Delta\nu = 0$). In this case, the spins are either magnetically equivalent, when they have identical coupling constants to all other (non-equivalent) spins in the molecule, or magnetically *in*equivalent, when this is not the case[2]. The coupling between magnetically equivalent spins is inconsequential, as it cannot be observed. On the contrary, the coupling between magnetically inequivalent spins is observable and responsible for additional line splitting, a fact well established since the pioneering work of McConnell, McLean and Reilly[3]. The theory of magnetically inequivalent spin systems and analysis of their spectra has been thoroughly described in the 50's and 60's of last century[4-10]. These and other analyses show that, in particular for large spin systems, the strong coupling between magnetically inequivalent nuclei can generate so many additional splittings and additional lines that the extraction of accurate spin frequencies and scalar coupling constants is severely challenged[11-13].

Here, we revisit the analysis of magnetically inequivalent spin systems to describe an exceptional scenario where this spectral complexity is lost and a simple, essentially weakly coupled spectrum is obtained. We demonstrate this effect experimentally in a 13-spin system, a ^{13}C,^{15}N-labeled trimethylated amine, in which the three methyl groups are magnetically inequivalent. We show that the combination of large and small coupling constants in this spin system results in a system that can very well be treated

as weakly coupled magnetically inequivalent spin pairs. We further demonstrate that the effects from magnetically inequivalence can be completely removed in this case by ^{13}C-spin state selective proton homo-decoupling. The curious spectrum of this compound and the controlled transformation to magnetic equivalence highlight the richness of scalar coupling phenomena in NMR.

Theory

In this section, we will, after a brief introduction, use the sub-spectral analysis approach to derive the boundary conditions that result in weakly coupled spectra of magnetically inequivalent spins.

Sub-spectral analysis

The sub-spectral analysis framework was developed by Diehl, Bernstein and co-workers to ease the accurate determination of *J*-coupling values and chemical shifts in magnetically inequivalent systems[8,10]. The basic principle of this approach is to decompose the spectrum into much simpler sub-spectra, each with well-defined (apparent) chemical shifts and coupling constants. This decomposition reflects the factorization of the Hamiltonian, or equivalently the energy level diagram, into independent parts. Considering the A-spin spectrum of AX-type spin systems, the energy level diagram can be factorized based on the total magnetic quantum number $m_{T,X}$ for the X-spins: transitions of spin A are only allowed between wave functions that have $\Delta m_{T,X} = 0$ (the so-called X-approximation), in addition to the usual $\Delta m = +/- 1$ selection rule. Further factorization can be obtained by using wave functions that are constructed using the molecular symmetry, since selection rules only allow transitions between states of the same symmetry class[4]. Thus, the sub-spectra in the A-spectrum correspond each to a *sub*-energy level diagram formed by wave functions of one particular $m_{T,X}$ value and one particular symmetry class.

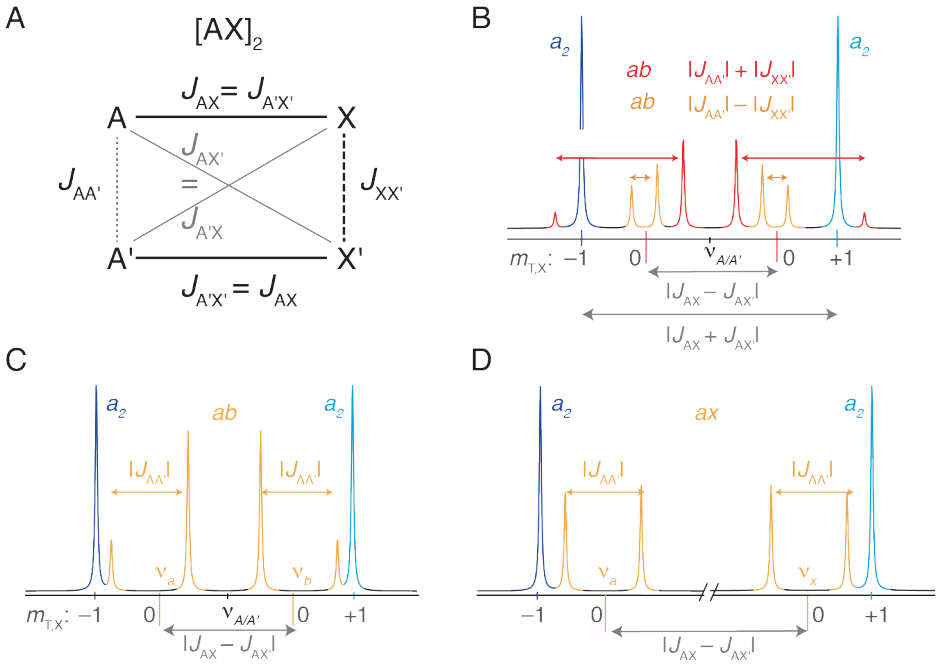


Figure 1. Sub-spectral analysis of the $[AX]_2$ spin system illustrating the spectral simplification when $J_{XX'}$ couplings are negligible and J_{AX} couplings are large. (A) Spin topology of the $[AX]_2$ spin system with all *J*-couplings indicated (B,C,D) Simulated spectra of the AA'-part of the AA'XX' spectrum using the following parameters: $J_{AX} = 7.5$, $J_{AX'} = 2.5$, $J_{AA'} = 3$, and $J_{XX'} = 2$ Hz in (B); $J_{XX'} = 0$ Hz in (C) and in (D) $J_{XX'} = 0$ Hz and $J_{AX} = 150$ Hz. Decomposition of the spectrum into sub-spectra is indicated (a_2 in blue/cyan; *ab* in red/orange). Effective Larmor frequency differences and coupling constants between the *a,b(x)* spins in the *ab(x)*-sub-spectra are indicated. The total spin state of the X-spins, $m_{T,X}$ is indicated as –1, 0 or +1.

Conditions for weakly coupled sub-spectra

To explain how weakly coupled sub-spectra can arise, we turn to the textbook case of a 4-spin system with two pairs of magnetically inequivalent spins[3,5,20,21]. This spin system is labeled AA'XX' in the extended notation of Pople[22,23], or $[AX]_2$ in the notation of Haigh[24]. The spin system is schematically illustrated in Figure 1A together with definition of the coupling constants. The spectrum of the inequivalent A/A' spins

consists of eight lines, decomposed into two *a*$_2$ and two *ab* sub-spectra (sub-spectra are denoted in lowercase italics) (Figure 1B). The two *a*$_2$ sub-spectra originate from sub-energy level diagrams that have the same functional shape as that for two equivalent spins, hence the *a*$_2$-designation. These corresponds to $m_{T,X}$ +1, i.e. X-spins in the αα-state, and to $m_{T,X}$ –1, i.e. X-spins in the ββ-state. In both subspaces, the A/A'-spins are described by fully equivalent, virtual *a*-spins with frequency $\nu_a = \nu_A \pm \left|J_{AX} + J_{AX'}\right|$. Their equivalence may be intuitively understood by imagining a molecule with the X/X'-spins both in the α or both in the β-spin state: the A/A'-spins are indistinguishable due to the two-fold symmetry of the molecule. This is not true for mixed α/β–states; hence the A/A'-spins in the $m_{T,X}$=0 space are magnetically *in*equivalent and subject to J_{AA}/J_{XX} -coupling. In this subspace, the A/A'-spins form a strongly coupled *ab*-type spin system, with frequency difference $\Delta\nu_{ab} = \left|J_{AX} - J_{AX'}\right|$ and *J*-coupling $J_{ab} = \left|J_{AA'}\right| \pm \left|J_{XX'}\right|$ between the virtual *a* and *b*-spins. Since this space is formed by both symmetrical and anti-symmetrical wave functions there are two *ab*-sub-spectra. An explicit presentation of the wave functions and energy diagram of the AA'XX'-system, based primarily on the analysis by Flynn *et al.*[20], is presented in Figure S2.

We will now set two boundary conditions that will result in a dramatic simplification of the spectrum. First, either J_{XX} or J_{AA} is near zero. In this case the two *ab*-sub-spectra in Figure 1B have the same effective splitting and collapse into a single *ab*-sub-spectrum, as shown in Figure 1C. Second, the frequency separation between the virtual *a* and *b*-spins $\Delta\nu_{ab}$ is much larger than their effective coupling constant J_{ab} , i.e. when J_{XX} ~0:

$$\left|J_{AX} - J_{AX'}\right| >> \left|J_{AA'}\right| \qquad [4]$$

Because the *a,b*-spins are now no longer strongly coupled, we redefine them as weakly coupled *a,x*-spins. The $m_{T,X}$=0 sub-spectrum is a symmetric doublet, split by J_{AA}, and is best labeled as an *ax*-sub-spectrum (Figure 1D). Thus, under the above two

conditions the spectrum of 8 lines and 5 different intensity levels of Figure 1A is reduced to 6 lines with only 2 different intensities, the sum of two a_2 and two overlapping *ax*-sub-spectra in Figure 1D. This dramatic simplification may arise if J_{AX} is very large compared to the other coupling constants.

Extension to $[AX]_3$ and $[A_3X]_3$ spin systems

The simplified pattern of Figure 1D in the $[AX]_2$ spin system can be extended to more complicated AX-type spin systems in a predictable manner. As an intermediate step towards the $[A_3X]_3$ system, it is useful to consider the A-spin spectrum in the 6-spin system $[AX]_3$, first described by Jones *et al.*[9]. In general, this spectrum consists of 56 transitions that can be grouped in six sub-spectra of three virtual spins: two a_3-patterns for $m_{T,X}=\pm^3/_2$, i.e. where the X-spins are in the αααβββ-spin state, two ab_2- and two *abc*-patterns for $m_{T,X}= \pm^1/_2$. In the special case where $J_{XX} = 0$, it can be shown that the *abc*-pattern is reduced to another ab_2-pattern[9], a situation depicted in Fig. 2A. The frequency difference between the virtual *a* and *b* spin is $| J_{AX} - J_{AX'} |$, with apparent coupling constant J_{AA} (when $J_{XX} = 0$). Here, we label the $m_{T,X}=\pm^1/_2$ sub-spectra on one side of A-multiplet as ab_2 and as a_2b on the other side, reflecting the different orientations of the ab_2-pattern. This complicated multiplet is again dramatically simplified when $| J_{AX} - J_{AX'} | >> | J_{AA} |$, as shown in Fig. 2B. Under these conditions, the *a* and *b* spins are effectively weakly coupled and the sub-spectra reduce to a_2x and ax_2-patterns. The logical progression from a_3 to a_2x and ax_2 can also be appreciated considering the X spin states: from ααα ($m_{T,X}=+^3/_2$, a_3) to states of the ααβ ($m_{T,X}=+^1/_2$, a_2x) to αββ-type ($m_{T,X}=-^1/_2$, ax_2).). We note that while the $[AX]_2$ system is symmetrical with respect to J_{AA} and J_{XX}, this is not true for $[AX]_3$ systems. The A-spin spectra for $J_{AA}=0$ and $J_{XX} << | J_{AX} - J_{AX'} |$ in $[A_nX]_3$ spin systems feature very distinct multiplets of five lines in addition to the singlet a_n sub-spectrum[25]. In addition, it can be noted that weakly coupled sub-spectra arise in the $[AX]_2$ system as

long as $| J_{AX} - J_{AX'} | >> | J_{AA} | + | J_{XX} |$ (see also Figure S2). This is not sufficient in $[A_nX]_3$ spin systems, where J_{XX} must be negligible.

We now turn to the $[A_3X]_3$ system. In general, without simplifying assumptions about the values of the coupling constants, there are 9312 allowed A-transitions in this spin system out of which many hundreds can result in distinct lines[26]. Even with negligible J_{XX}, the spectrum is characterized by the presence of many lines, with wildly varying intensities (Fig. 2C). Under the simplifying conditions $J_{XX} \sim 0$ and $| J_{AX} - J_{AX'} | >> | J_{AA} |$ the spectrum can now again be decomposed in simple sub-spectra. As there are three possible $m_{T,X}$-states for each side of the main J_{AX} doublet (e.g. $-3/2, -1/2, +1/2$ for the downfield triplet), there are again three sub-spectra. Recognizing that there are nine A-spins, the outer line is of the a_9-type. Since the spins within one A_3 group are magnetically equivalent, the center line becomes an a_6x_3 quartet, and the inner line is part of an a_3x_6 sub-spectrum, resulting in a septet (1:6:15:20:15:6:1) (Fig. 2D).

Similar reasoning can be used to show that $[A_2X]_2$ systems would have a_4 and a_2x_2 sub-spectra, and that spectra of $[A_2X]_3$ systems reduce to a_6, a_4x_2, and a_2x_4 sub-spectra. In short, we show here that in case of large $|J_{AX} - J_{AX'}|$ with respect to $J_{AA'}$ and negligible J_{XX}, the *a* and *b* species in sub-spectral analysis are in fact weakly coupled and the sub-spectra are simple multiplet patterns as in the case of weakly coupled spins.

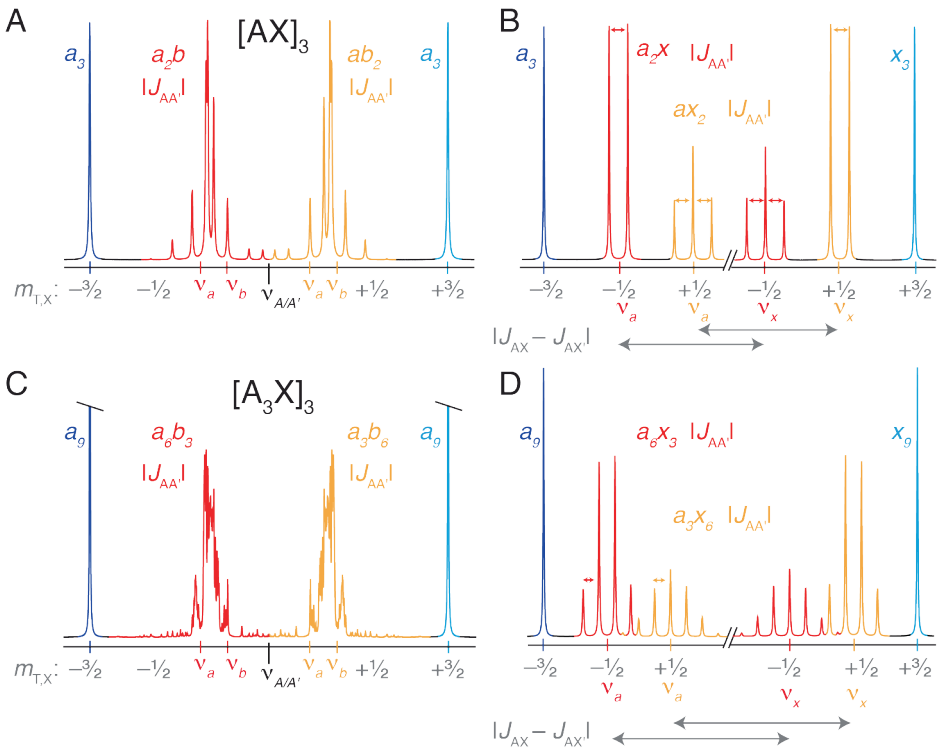


Figure 2. Sub-spectral analysis of $[AX]_3$ and $[A_3X]_3$ spin systems when J_{XX} couplings are negligible. Simulated spectra of the A-part of the $[AX]_3$ (A,B) and the $[A_3X]_3$ spin system (C,D) spectrum using the following parameters: $J_{AX} = 10$, $J_{AX'} = 8$, $J_{AA'} = 2$, $J_{XX'} = 0$ Hz (A,C); in (B,D) J_{AX} is set to 150 Hz resulting in weakly coupled *ax*-type sub-spectra. In the $[AX]_3$ spin systems this results in a_2x and ax_2 sub-spectra, each composed of a doublet and a triplet. In the $[A_3X]_3$-spin system this results in a_6x_3 and a_3x_6 sub-spectra, each composed of a quartet and a septet. Total spin states of the X-spin is indicated for each sub-spectrum. Decomposition of the spectrum into sub-spectra is indicated (*a*-type in blue/cyan; *ab*/*ax*-type in red/orange). Effective Larmor frequencies of the virtual *a* and *b*(*x*) spins and their coupling constants are indicated. In (B,D) the Larmor frequency differences between the *a* and *x*-spin within one sub-spectrum is indicated.

Results

The focus of our study is a ^{13}C,^{15}N-labeled trimethyl amine, abbreviated as compound **1** (Figure 3). While all methyl protons are chemically equivalent due to the molecular symmetry, the introduction of ^{13}C-isotopes breaks their magnetic equivalence. This 13-spin system is thus best described as a $A_3A_3'A_3''MXX'X''$, or more compactly as a $[A_3X]_3M$ system, where A_3 represents the three ^{1}H spins in a methyl group, M is the ^{15}N spin, X is the ^{13}C carbon spin and the quotes are used to indicate magnetically inequivalent spins. As a result of the inequivalence, the proton 1D NMR spectrum of compound **1** can be expected to show splittings due to the four-bound coupling between the geminal methyl groups ($^{4}J_{HH}$) and the two-bond carbon-carbon coupling ($^{2}J_{CC}$).

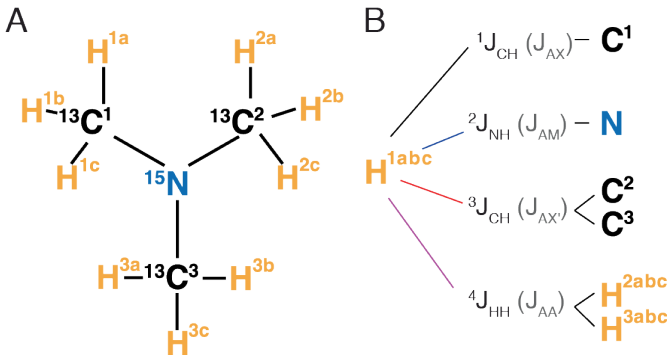


Figure 3. Structure and spin topology of the amine group in the molecule studied. (A) The amine head group of 2-bromo-d_4-ethyl-^{13}C-trimethyl-^{15}N-ammonium is shown, featuring three geminal methyl groups. The deuterated bromo-ethyl group attached to the amine ^{15}N is omitted for clarity. Atoms are labeled with superscripts to allow their distinction in the main text. The full structure of the compound is shown in supplemental Figure S1. (B) Scalar couplings experienced by the $H^{1a,b,c}$ spins, including to the geminal methyl groups H^2 and H^3. Corresponding AX nomenclature is shown between brackets. The coupling networks of the $H^{2a,b,c}$ and $H^{3a,b,c}$ spins can be found by permuting the atom labels. Protons in different methyl groups are magnetically inequivalent: protons in methyl group 1 have a large one-bond coupling ($^{1}J_{CH}$) to the directly attached C1 spin and much smaller three-bond coupling ($^{3}J_{CH}$) to C2, whereas protons in methyl groups 2 have three-bond couplings to C1

and a one-bond coupling to C2. Similar arguments can be made for all pairs of methyl groups, and for the ^{13}C spins.

As set out in the Theory section above, the appearance of AA'...XX'...–type spin systems depends strongly on the relative magnitudes of the coupling constants J_{AX}, $J_{AX'}$, J_{AA}, J_{XX}. In general, when all couplings are of comparable strength, the resulting spectrum will feature many lines with varying intensities, reflecting the strong coupling between magnetic inequivalent spin pairs. Moreover, the spectral complexity will increase rapidly with increasing size of the spin system, resulting in hundreds to thousands of distinct lines for an $[A_3X]_3$ spin system[26]. On the other hand, when some coupling constants dominate and others are negligible, the spectrum is greatly simplified. In particular, in case $J_{XX}\sim0$ and $|\ J_{AX} - J_{AX'}\ | >> |\ J_{AA}\ |$, the magnetically inequivalent nuclei become weakly coupled and the spectrum will reduce to a simple pattern of *ax*-type sub-spectra (see Theory section). In our case $|\ J_{AX} - J_{AX'}\ |$ corresponds to $|{}^1J_{CH} - {}^3J_{CH}|$, while J_{AA} and J_{XX} correspond ${}^4J_{HH}$ and ${}^2J_{CC}$ respectively. Thus, we may expect this simplifying condition to hold true for compound **1**, resulting in a spectrum similar to that in Figure 2D.

Assignment of splittings in the ^1H spectrum

The proton 1D NMR spectrum of compound **1** is characterized by two highly asymmetric multiplet patterns separated by the large ${}^1J_{CH}$ (Figure 4A). Ignoring the asymmetry, these patterns have the appearance of triplets, as expected considering the methyl protons are coupled to two ^{13}C spins by ${}^3J_{CH}$. Peak integration of the three triplet components shows the expected 1:2:1 ratio. As we will show below, the asymmetry is caused by unresolved ${}^4J_{HH}$ splittings.

The outer triplet component appears as a sharp doublet with 0.75 Hz splitting (Figure 4B). This splitting is removed upon ^{15}N decoupling (Figure 4C,D), proving that it is due to the ${}^2J_{HN}$ coupling interaction. This splitting is masked on the central and inner

line of the triplet by their much broader appearance (Figure 4B), but clearly present as can be inferred from the peak sharpening upon ^{15}N-decoupling (Figure 4D).

The center and inner lines in the spectra of Figure 4 show subtle signs of additional splittings. In an attempt to resolve these splittings, a high-resolution spectrum was recorded with 8 sec acquisition time and processed using different degrees of resolution-enhancing Lorentz-to-Gauss transformations (Figure 5). Without any window-function the spectrum shows a near perfect triplet, although the $^{2}J_{HN}$- splitting is only visible on the outer line (Figure 5A). Processing with moderate or strong resolution enhancement uncovers small splittings on the center and inner lines only (Figure 5B,C). This fine structure is different for the center and inner component, as is clear from their different shapes. This is consistently visible on both sides of the $^{1}J_{CH}$ doublet. The apparent coupling constant is on the order of 0.4 Hz, in line with *J*-coupling values expected for $^{4}J_{HH}$ or $^{2}J_{CC}$, and much smaller than $|^{1}J_{CH} - {}^{3}J_{CH}|$ (141.3 Hz). The increase in additional splittings going from the outer to center to the inner lines, i.e. there is a ^{13}C spin state dependent pattern of splittings, obeys the pattern shown in Figure 2D fully.

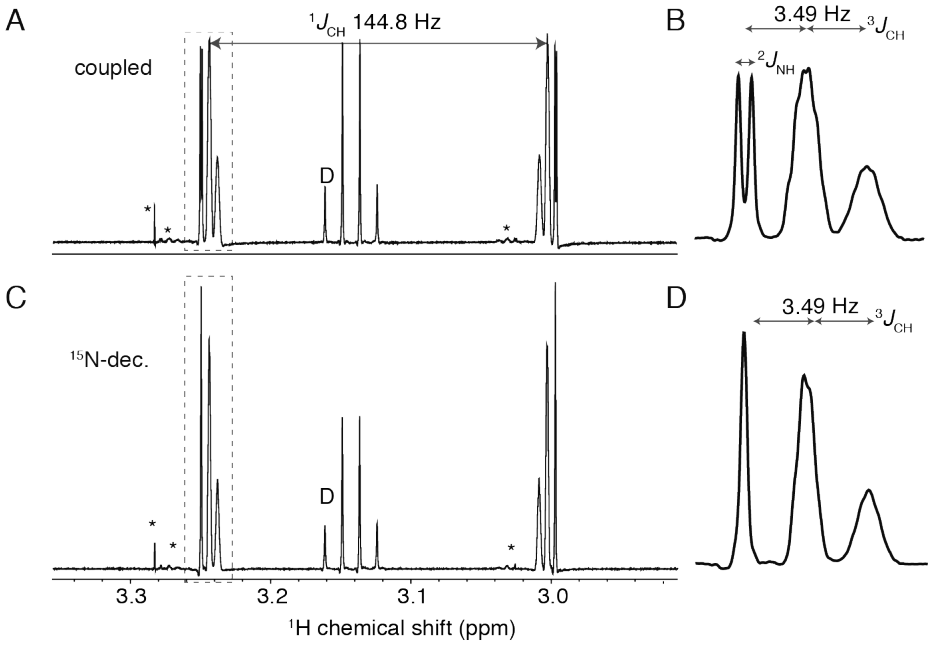


Figure 4. Proton 1D NMR spectra of a $^{13}C/^{15}N$-labeled trimethyl quaternary amine. Sections from the ^{1}H 1D NMR spectra without decoupling (A,B) and with ^{15}N decoupling (C,D) showing the signals from the nine chemically equivalent trimethyl protons. The boxed regions in (A,C) are expanded in (B,D) with indicated splittings in Hz. Signals from DIPEA (D) and minor impurities (*) are indicated.

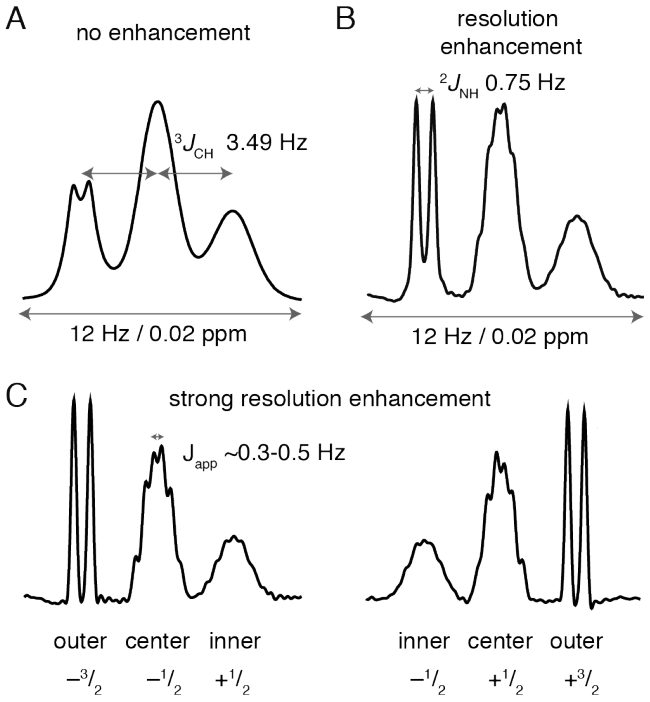


Figure 5. Different fine structures are observed for each component of the $^{3}J_{CH}$ triplet. (A, B) show the downfield component of the main $^{1}J_{CH}$ doublet, (C) also shows the upfield $^{1}J_{CH}$ component. The different fine structure for each triplet component is already visible without any processing to enhance resolution (A), but more clear upon moderate (B) or strong (C), resolution enhancement through Lorentz-to-Gauss transformation. The additional splittings visible on the center and inner line are indicated in (C), together with the total spin state of the ^{13}C spins.

Estimation of $^4J_{HH}$ and $^2J_{CC}$

We next applied the theory described above and decomposed the experimental spectrum into *ax*-type sub-spectra to extract an estimate for the $^4J_{HH}$ between the geminal methyl groups, thus assuming $^2J_{CC}$ is zero. Since compound **1** is an $[A_3X]_3M$ spin system, each line of the a_9, a_6x_3 and a_3x_6 sub-spectra in Figure 2D will now be split into a doublet due to $^2J_{HN}$. The downfield component of the main $^1J_{CH}$ doublet was thus fitted as a sum of a doublet, a quartet of doublets and a septet of doublets. The resulting stick-plot and a simulated spectrum with the experimental lineshapes is shown in Figure 6A,B. The best fit is obtained using 0.43 Hz for $|^4J_{HH}|$, and agrees very well with the experimental spectrum. An exact calculation of the ^{1}H spectrum of the $[A_3X]_3M$ system using the composite particle approach is shown in Figure 6C, underscoring the validity of the weak coupling approximation. The numerical approach also allows evaluating the impact of non-zero values for $^2J_{CC}$ (Figure 6C). Significant deviations from the experimental lineshape are already visible with very small non-zero values (0.4 Hz), most notably around the center of the inner multiplet. Furthermore, strong inequivalence effects are absent from the ^{13}C 1D spectrum (Figure S3). Together, the exact simulations of both the ^{1}H and ^{13}C spectra strongly support a near-zero value for $^2J_{CC}$, at least $|^2J_{CC}| < 0.4$ Hz.

Notably, comparison of the experimental and numerically simulated spectrum highlights that the observed linewidths of the outer, center and inner line are unequal (Fig. 6A,C). There is a small, but significant, increase in linewidth going from the outer to the center and inner line. This effect has been accounted for in the fit of Fig. 6B, where the center and inner lines are broadened by ~0.03 and 0.06 Hz, respectively, compared to the outer line.

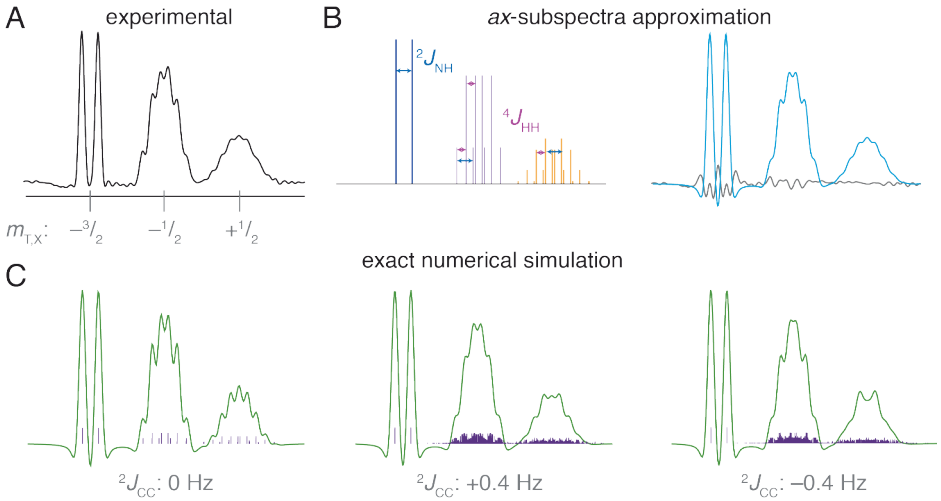


Figure 6. Estimation of $^4J_{HH}$ and $^2J_{CC}$ in a quaternary trimethyl amine. (A) Downfield part of the experimental spectrum showing the $^1J_{CH}$-doublet. (B) Stick (left) and simulated spectrum (right) of using the weakly coupled *ax*-type sub-spectra approximation (see Materials and Methods for details) using $^1J_{CH}$ = 144.8; $^3J_{CH}$ = 3.49; $^2J_{NH}$ = 0.749; $^4J_{HH}$ = 0.425; $^2J_{CC}$ = 0 Hz. The difference with the experimental spectrum is shown in gray on the right. In the stick spectrum, the $^2J_{NH}$ and $^4J_{HH}$ splitting are indicated for a number of lines. (**C**) Simulated spectrum using the composite-particle approach with the left spectrum using the *J*-values as in B as input, the middle and right spectrum have non-zero values for $^2J_{CC}$ as indicated. Stick-plot is shown as purple lines, overlaid with the summed spectrum in green.

Transformation to magnetic equivalence

The close agreement between the experimental spectrum and a fit based on a decomposition into *ax*-type sub-spectra suggests that the different methyl groups are in fact weakly coupled to each other due to $^4J_{HH}$. Phrased differently, the downfield center line and the upfield inner line form a weakly coupled *ax* spin system as do the downfield inner and upfield center line (cf. Figure 2D). This effect can also be seen in the 2D DQF-COSY spectrum, which shows cross-peaks between the center and inner line (Figure S4). Since spins *a* and *x* are weakly coupled and since there is a large

frequency separation between them (cf. Figure 2D), one may hypothesize the possibility to decouple their interaction. Indeed, a homo-decoupling experiment where the upfield component of the main $^{1}J_{CH}$ doublet, corresponding to the x spin species, are irradiated during acquisition results in the removal of all $^{4}J_{HH}$ splittings (Figure 7A,B). The multiplets are collapsed to the expected triplet for a set of three fully equivalent methyl groups. Composite particle based calculation of this spin state-selective homo-decoupling experiment confirms the controlled transformation to a magnetically equivalent spin system (Figure 7C). It is also of interest to note that the linewidths for each of the triplet components are very similar upon homo-decoupling.

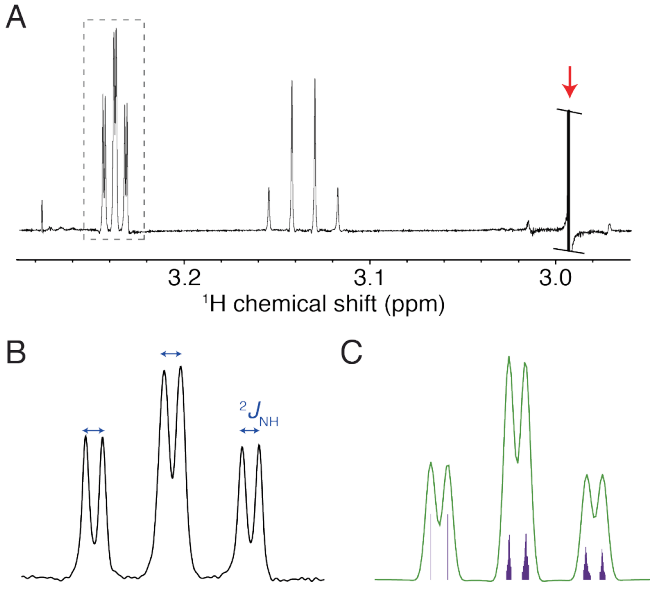


Figure 7. Reduction to a fully equivalent spin system by spin state-selective homo-decoupling. (A) Experimental spectrum showing that the additional splittings on the center and inner lines of the downfield $^{1}J_{CH}$ component can be removed by selective homo-decoupling (~70 Hz / 25 μW γB_2 CW irradiation) of the upfield component (indicated by the arrow). The boxed region is expanded in (B), showing a pure $^{3}J_{CH}$ triplet with each component split in a doublet due to $^{2}J_{NH}$. (C) Simulated spectrum using composite particle analysis. Spectrum was acquired and processed as those in Fig. 2. Stick-plot is shown as purple lines.

Discussion

Magnetic inequivalent nuclei are regularly encountered in organic or inorganic compounds, where their presence can be used to infer molecular symmetry. Their NMR signals feature more lines and more intensities levels compared to a first-order coupling pattern, due to the strong coupling between magnetically inequivalent spin pairs. In this paper, we have presented an exceptional case of how magnetic inequivalence can affect the appearance of spectra, resulting in a sum of first-order coupling patterns. Compound **1**, an isotope-labeled trimethyl amine, presents an $[A_3X]_3M$ spin system where the coupling constants produce an extreme simplification of a spectrum that in typical cases would consist of hundreds to thousands of lines. We showed that this simplification depends on two conditions: *i*) J_{XX} is near-zero; and *ii*) $|\ J_{AX} - J_{AX'}\ | >> |\ J_{AA}\ |$. We derived that under these conditions the spectrum can be decomposed into sub-spectra formed by fully equivalent spins (a_9-sub-spectra), and sub-spectra formed by weakly coupled *a* and *x* spins with coupling constant J_{AA} (a_6x_3/a_3x_6-sub-spectra).

We presented a stepwise analysis of $[AX]_2$, $[AX]_3$ and $[A_3X]_3$ spin systems, to understand and illustrate the salient features of the experimental spectrum, rather than a complete analytical description of the $[A_3X]_3M$ system. As noted by Diehl *et al.*, the composite particle approach offers the most convenient route to solving the Hamiltonian[10]. In particular, this approach does not rely on molecular symmetry to factorize the Hamiltonian but rather factorizes based on spin states of the (composite) particles. In this description, the ^1H spectrum is described as a sum of 16 non-identical sub-spectra, where each sub-spectrum corresponds to one $m_{T,X}$ state and one particular combination of quartet and doublet spin states for the three methyl groups (see Figure S5). Notably, all sub-spectra with the same $m_{T,X}$ value overlap and add up to the observed a_9, a_6x_3 or a_3x_6 patterns.

The overall spectrum can be reproduced using just 48 lines, thus classifying as a 'deceptively simple' spectrum where many transitions are (nearly) degenerate and/or have low transition probabilities[27]. This spectral simplicity made it possible to fit the

experimental spectrum and extract a value of 0.43 Hz for the $^4J_{HH}$ between geminal methyl group in compound **1**. This is well in line with previous reports on geminal methyl coupling constants[28], in particular with a value of 0.41 Hz for the primary amine tert-butyl-amine[29]. Notably, direct measurement of $^4J_{HH}$ from the ^{13}C satellite peaks in an unlabeled compound failed since the presence of the quadrupolar ^{14}N nucleus induces ^{1}H line broadening through scalar relaxation of the second kind (data not shown). A near-zero value for the $^2J_{CC}$ matches the observation that these couplings are generally small in saturated aliphatic groups, and that they can be both positive and negative[30]. Notably, a reverse assignment of J_{HH} and J_{CC} is not possible since the A-spin spectrum of a $[A_3X]_3$ systems has different multiplet structures when $J_{AA}=0$[25].

When fitting the experimental lineshape with the weak coupling approximation, we found small but significant linewidth differences between the outer, center and inner line. This differential line broadening effect is removed in the homo-decoupled experiment, where the *x*-spins are decoupled from the *a*-spins. The observed broadening is thus most likely due to relaxation interference between the different ^{1}H-^{1}H dipolar interactions. A full analysis of this effect is beyond the scope of this paper.

The peculiar spectral manifestation of magnetic inequivalence in compound **1** may also be found in other compounds as long as the two limiting conditions are met. We thus expect that isotope labeled compounds with carbon/nitrogen-atoms separated by another hetero-nucleus such as C-O-C or C-S-C, or N-C-N type structures, as for instance ^{13}C-labeled dimethyl-ether, will show similar spectra with *ax*-type sub-spectra. The presence of the hetero-nucleus is required to make sure J_{XX} is negligible.

To the best of our knowledge this report is the first explicit analysis and experimental demonstration of a completely first-order spectrum for a magnetically inequivalent spin system. Other limiting cases resulting in spectral simplification have been described before, particularly in the work of Harris *et al.*[25,31-33]. In the case of $[PF_2]_2$-containing complexes, they showed that some sub-spectra may become first-order when one of the J_{PF} constants is very large[33]. Since in that case both J_{PP} and J_{FF} were

non-zero, additional second-order sub-spectra remain. Another report on phosporous-fluorine complexes shows, without analysis, the ^{19}F spectrum of $Ni[PF_3]_4$ which is reminiscent of the spectrum of compound **1**[34]**.** Even though not all splittings can be resolved, it is clear that the spectrum deviates from the expected a_9x_3 quartet, suggesting that $^2J_{PP}$ is not negligible. The case described here stands out from such previous reports, as it presents an extreme simplification afforded by a 'perfect storm' of coupling constants.

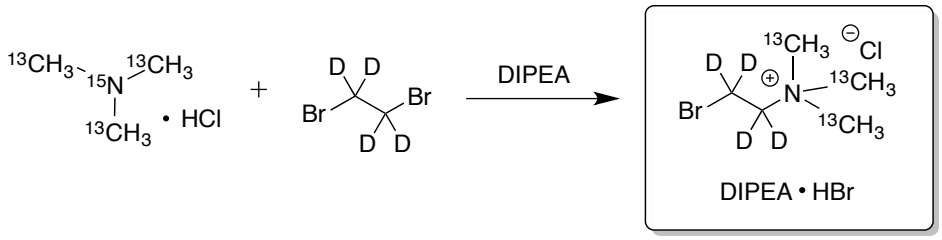
Lastly, our work demonstrates a remarkable consequence of the weak coupling between the magnetically inequivalent nuclei: they can effectively be decoupled by spin state selective homo-decoupling. This transforms the spin system into a fully equivalent system, removing all $^4J_{HH}$ splittings from the multiplet. In our opinion, the spectrum of the ^{13}C-labeled trimethyl amine presented here is an extremely instructive illustration of the effects of magnetic inequivalence. Depending on your viewpoint the spectrum may be considered deceptively simple or deceptively complicated.

Materials and Methods

Synthesis

Compound **1**, systematic name 2-bromo-N,N,N-tri(methyl-^{13}C)ethan-1-aminium-1,1,2,2-d_4-N-^{15}N-chloride, was synthesized in a straightforward manner from fully isotope labelled trimethylammonium chloride and fully deuterated 1,2 dibromo-ethane (see Scheme 1). In a 10 ml round-bottom flask, tri(methyl-^{13}C)-ammonium-^{15}N chloride (100 mg, 1.05 mmol) was mixed with a 10 fold excess of 1,2-dibromoethane-1,1,2,2-d_4 (1.98 g, 0.91 ml, 10.5 mmol). DIPEA (365 µl, 2.1 mmol) was then added and the mixture was stirred for 15 minutes. A few drops of MeOH were added to ensure complete dissolution. The mixture was placed under an N_2 atmosphere and stirred for 96 hours at 30°C. The excess 1,2-dibromoethane-d_4 was removed by rotary evaporation and the residue suspended in cold EtOAc. After filtration, the solids were washed with cold EtOAc and dried under high vacuum. The yield was 413 mg of compound/DIPEA-HBr (94%). The resulting compound was

characterized by HRMS ESI-TOF and NMR. m/z: expected mass for $C_2{}^{13}C_3H_9D_4Br^{15}N$ ($[M]^+$): 174.0548, experimentally obtained 174.0522 Da (Figure S1).



Scheme 1. Synthesis of compound 1.

NMR spectroscopy

NMR spectra of compound **1** (~20 mM) in 100% D_2O were obtained at 298K on a Bruker Avance III HD spectrometer operating at 600 MHz ^1H Larmor frequency and equipped with a cryoprobe. All 1D ^1H spectra were acquired in a single scan and zero-filled two times before Fourier transformation. Highest resolution spectra were obtained with 8 sec. acquisition time and Lorentz-to-Gauss apodization with negative exponential broadening -1 Hz and the center of the Gaussian placed at 30% of the FID. Other spectra, including $^{15}N/^{13}C$ decoupled spectra, were recorded with 5 sec acquisition time and apodized with -0.8 Hz negative exponential broadening and the center of the Gaussian shifted to 40% of the length of the FID. In order to obtain a high-resolution $^{15}N/^{13}C$-decoupled 1D, a low power on-resonance WALTZ16-decoupling field was used: 260 Hz γB_2 at 49.9 ppm for ^{15}N, and 640 Hz γB_2 at 53.0 ppm for ^{13}C. The ^{15}N and ^{13}C chemical shifts were obtained from 2D H(C)N and 2D HMQC spectra. The homo-decoupled spectrum was recorded using the zghd.2 pulse program with continuous wave irradiation (~70 Hz γB_2) centered on the upfield $^1J_{CH}$ doublet component and 5 seconds acquisition time.

Simulations

There are several powerful programs that can simulate solution-state NMR spectra, e.g.[14-16]. We preferred to write our own simulation script as part of our learning

process and to be able to tailor the simulations to our needs and requirements. Spectra of various AX-type spin systems were simulated using an in-house written GNU Octave[17] script that implements the composite particle approach (CPA)[18,19], as outlined by Diehl and co-workers[10]. Briefly, magnetically equivalent spins are grouped into a single composite particle, with total spin F_{max} equal to the sum of the grouped spins, e.g. F_{max} = 3/2 for three equivalent methyl protons. This particle can exist in independent spin states with a maximum spin F equal to F_{max}, F_{max} –1, ... ; e.g. for a system of one methyl group the allowed spin states are quartet (F = 3/2, Q) and doublet (F = 1/2, D), such that a trimethyl group can exist in any of 8 states such as QQQ, QQD, QDQ, *etc.* To solve the A-spin (^{1}H) spectrum, the Hamiltonian is factorized according to the total spin of the system m_T, the spin state of each particle, and the total spin $m_{T,X}$ of the X-nuclei (^{13}C). This procedure breaks down the 4096 x 4096 matrix for the 12-spin case into much smaller matrices of at most 36 x 36. The transition frequencies and intensities are then calculated using standard approaches. Each sub-spectrum is multiplied with a statistical weight to account for the degeneracy of the involved spin states. Lines corresponding to transitions with extremely low probability (intensity lower than $1 \cdot 10^{-4}$, whereas the largest peak has intensity 36) are omitted to ensure efficient drawing of the spectrum. For all remaining transitions an FID is simulated, these are subsequently added, an exponential decay to match the experimental lineshape is added, and the final FID is processed according to the experimental settings to yield the final spectrum.

In the weak-coupling approximation, the experimental spectrum was simulated by calculating FIDs according to the experimental settings for the outer, center and inner lines individually. For all three signals a weak $^2J_{NH}$ coupling was added. The effect of inequivalence was simulated by adding a weak $^4J_{HH}$ coupling to either 0, 3, or 6 equivalent protons for the outer, center and inner line, respectively. An exponential decay was added to match the experimental lineshape. Processing was identical to the experimental situation. All simulation scripts are available upon request.

Acknowledgements

We thank Dr. Dennis Hetterscheid for discussions on group theory.

References

1 M. Karplus, Contact Electron-Spin Coupling of Nuclear Magnetic Moments, The Journal of Chemical Physics. 30 (1959) 11–15. doi:10.1063/1.1729860.

2 H.S. Gutowsky, Chemical shifts and electron-coupled spin-spin interactions, Ann N Y Acad Sci. 70 (1958) 786–805. doi:10.1111/j.1749-6632.1958.tb35431.x.

3 H.M. McConnell, A.D. McLean, C.A. Reilly, Analysis of Spin-Spin Multiplets in Nuclear Magnetic Resonance Spectra, The Journal of Chemical Physics. 23 (1955) 1152. doi:10.1063/1.1742204.

4 E.B. Wilson Jr, Analysis of Spin-Spin Interaction in the Nuclear Magnetic Resonance Spectra of Symmetrical Molecules, The Journal of Chemical Physics. 27 (1957) 60–68. doi:10.1063/1.1743719.

5 J.A. Pople, W.G. Schneider, H.J. Bernstein, The analysis of nuclear magnetic resonance spectra: II. Two pairs of two equivalent nuclei, Can. J. Chem. 35 (1957) 1060–1072. doi:10.1139/v57-143.

6 F.S. Mortimer, The NMR spectrum of bicycloheptadiene, Journal of Molecular Spectroscopy. 3 (1959) 528–535. doi:10.1016/0022-2852(59)90045-1.

7 P.L. Corio, The Analysis of Nuclear Magnetic Resonance Spectra, Chemical Reviews. 60 (1960) 363–429. doi:10.1021/cr60206a003.

8 P. Diehl, R.G. Jones, H.J. Bernstein, A simplified procedure for the analysis of complex nuclear magnetic resonance spectra: I. The principles of sub-spectral analysis, Can. J. Chem. 43 (1965) 81–93. doi:10.1139/v65-011.

9 R.G. Jones, R.C. Hirst, H.J. Bernstein, A simplified procedure for the analysis of complex nuclear magnetic resonance spectra: II. The analysis of symmetrical trifluorobenzene, the AA'A"XX'X" system, Can. J. Chem. 43 (1965) 683–699. doi:10.1139/v65-087.

10 P. Diehl, R.K. Harris, R.G. Jones, Sub-spectral analysis, Progress in Nuclear Magnetic Resonance Spectroscopy. (1967). doi:10.1016/0079-6565(67)80011-6.

11 A. Connelly, R.K. Harris, D.J. Loomes, P.J. Beynon, Nuclear magnetic resonance spectral analysis for spin systems involving threefold symmetry. The spectrum of N(PF2)3, J. Chem. Soc., Faraday Trans. 2. 77 (1981) 915–928. doi:10.1039/F29817700915.

12 C.G. Barlow, D.L. Miller, R.A. Newmark, Complete analysis of the 31P and 19F NMR spectra of [(PF3)3Mo(CO)3], Magn Reson Chem. 38 (2000) 38–42. doi:10.1002/(SICI)1097-458X(200001)38:1<38::AID-MRC546>3.0.CO;2-G.

13 L. Kapička, D. Dastych, V. Richterová, M. Alberti, P. Kubáček, Analysis and calculation of the 31P and 19F NMR spectra of hexafluorocyclotriphosphazene, Magn Reson Chem. 43 (2005) 294–301. doi:10.1002/mrc.1549.

14 H.J. Reich, WINDNMR-Pro: Simulating NMR spectra with WINDNMR-Pro, (2002). http://www.chem.wisc.edu/areas/reich/plt/windnmr.htm (accessed February 28, 2017).

15 H.J. Hogben, M. Krzystyniak, G.T.P. Charnock, P.J. Hore, I. Kuprov, Spinach--a software library for simulation of spin dynamics in large spin systems, J Mag Res. 208 (2011) 179–194. doi:10.1016/j.jmr.2010.11.008.

16 A.M. Castillo, L. Patiny, J. Wist, Fast and accurate algorithm for the simulation of NMR spectra of large spin systems, J Mag Res. 209 (2011) 123–130.

doi:10.1016/j.jmr.2010.12.008.
17 J.W. Eaton, D. Bateman, S. Hauberg, GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations, CreateSpace Independent Publishing Platform, 2009. http://www.gnu.org/software/octave/doc/interpreter.
18 J.S. Waugh, F.W. Dobbs, Strong Coupling in Nuclear Resonance Spectra. III. Systems Containing Many Equivalent Spins, The Journal of Chemical Physics. 31 (1959) 1235–1239. doi:10.1063/1.1730575.
19 D.R. Whitman, L. Onsager, M. Saunders, H.E. Dubb, Proton Magnetic Resonance Spectrum of Propane, The Journal of Chemical Physics. 32 (1960) 67–71. doi:10.1063/1.1700949.
20 G.W. Flynn, NMR Spectrum of 1,1-Difluoroethylene in the Gas Phase, The Journal of Chemical Physics. 38 (1963) 226. doi:10.1063/1.1733466.
21 P. Diehl, Über eine Vereinfachung der Beschreibung und Analyse von komplizierten kernmagnetischen Resonanzspektren mit Hilfe von Teilspektren, Helvetica Chimica Acta. 48 (1965) 567–580. doi:10.1002/hlca.19650480316.
22 H.J. Bernstein, J.A. Pople, W.G. Schneider, The analysis of nuclear magnetic resonance spectra: I. Systems of two and three nuclei, Can. J. Chem. 35 (1957) 67–83. doi:10.1139/v57-011.
23 R.E. Richards, T. Schaefer, High resolution hydrogen resonance spectra of trisubstituted benzenes, Mol Phys. 1 (1958) 331–342. doi:10.1080/00268975800100411.
24 C.W. Haigh, Notation for spin systems in nuclear magnetic resonance spectroscopy, Journal of the Chemical Society a: Inorganic. (1970).
25 E.G. Finer, R.K. Harris, Nuclear magnetic resonance spectra of the [AXn]3 spin system: calculations of X spectra when Jxx= 0, Journal of the Chemical Society a: Inorganic. (1969) 1972–1976. doi:10.1039/J19690001972.
26 J.F. Nixon, 19F nuclear magnetic resonance spectra of some trisubstituted metal trifluorophosphine complexes. Examples of [AX3]3 nuclear spin systems, Journal of Fluorine Chemistry. 3 (1973) 179–185. doi:10.1016/S0022-1139(00)84162-7.
27 R.J. Abraham, H.J. Bernstein, The analysis of nuclear magnetic resonance spectra: V. The analysis of deceptively simple spectra, Can. J. Chem. 39 (1961) 216–230. doi:10.1139/v61-025.
28 W.J. Mijs, Long range spin spin coupling between protons of geminal methyl groups in 1,3-dioxane derivatives, Recueil Des Travaux Chimiques Des Pays-Bas. 86 (1967) 220–224. doi:10.1002/recl.19670860213.
29 D.J. Sardella, Substituent effects on long-range couplings in substituted propanes, Journal of Molecular Spectroscopy. 31 (1969) 70–75. doi:10.1016/0022-2852(69)90340-3.
30 H.O. Kalinowski, S. Berger, S. Braun, 13C-NMR-Spektroskopie, Georg Thieme Verlag, Stuttgart, 1984.
31 R.K. Harris, Comments on N.M.R. sepctra of the XnAA'Xn' type, Can. J. Chem. 42 (1964) 2275–2281. doi:10.1139/v64-334.
32 E.G. Finer, R.K. Harris, M.R. Bond, R. Keat, Proton spectra of symmetrical cyclotriphosphazatrienes, Journal of Molecular Spectroscopy. 33 (1970) 72–83. doi:10.1016/0022-2852(70)90053-6.
33 R.K. Harris, M.I.M. Wazeer, NMR Spectra of the [AX2]2 and [A[X]2]2 Types: The X Transitions when |JAX| is Large, as in the Case of Tetrafluorodiazadiphosphetidines, Berichte Der Bunsengesellschaft Für Physikalische Chemie. 83 (1979) 697–702. doi:10.1002/bbpc.19790830711.
34 G.S. Reddy, R. Schmutzler, Phosphorus-fluorine chemistry. XVIII. Nuclear magnetic resonance studies on coordination compounds involving fluorine-containing phosphine ligands, Inorg Chem. 6 (1967) 823–830. doi:10.1021/ic50050a036.

Supplementary material

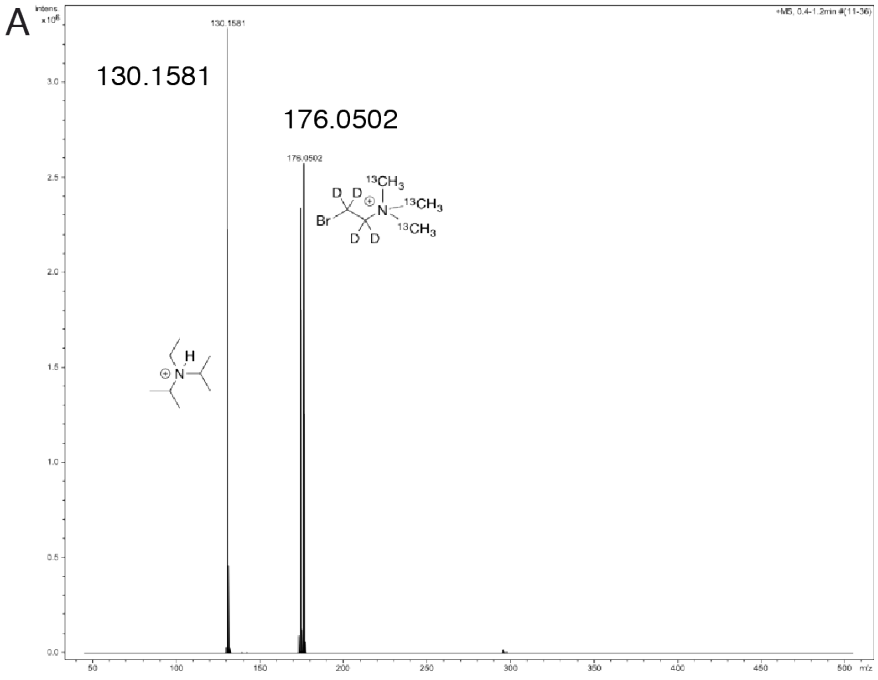


Figure S1. Quality control mass and NMR spectra of compound 1. (A) HRMS ESI-TOF analysis confirms purity. Expected mass of 2-bromo-N,N,N-tri(methyl-^{13}C)ethane-1-aminium-1,1,2,2-d$_4$-N-^{15}N-chloride ($C_2{}^{13}C_3H_9D_4Br_{15}N$) is 174.0548/ 176.0527 for the ^{79}Br (49% abundance) and ^{81}Br (51% abundance) isotopes. The purified product also contains DIPEA. (B) 1D ^{1}H NMR spectrum of compound 1/DIPEA-HBr in 100% D_2O. Indicated are the signals from compound 1, DIPEA (D1 (*sep*) = CH, D2 (*q*)= CH2, D3 (*d*) = CH3, D4 (*t*) = CH3), residual water (H_2O) and impurities (*).

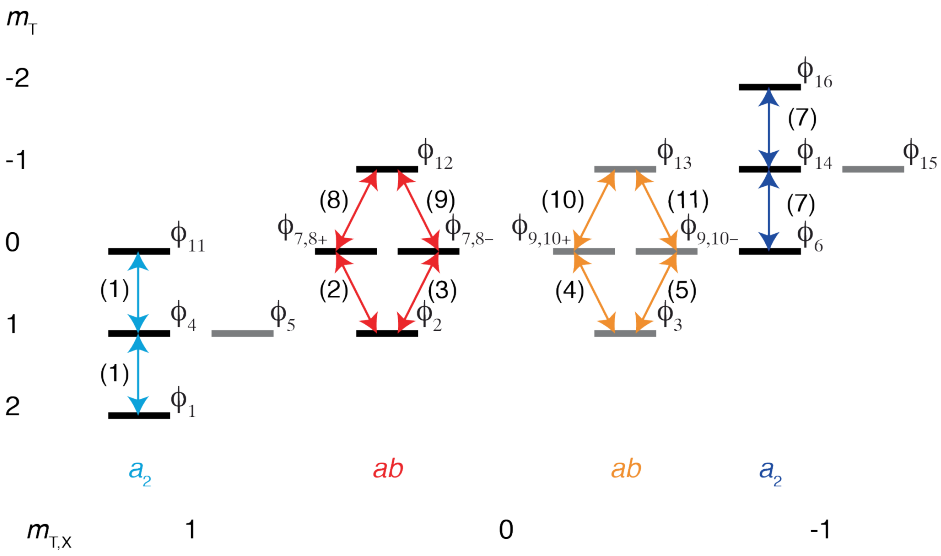


Figure S2. Energy level diagram of the AA'XX' spin system with allowed A-spin transitions indicated. The energy diagram shown is factorized for total spin m_T, total spin-state of the X-spins, $m_{T,X}$ and symmetry of the involved wave functions. Black bars represent energy levels of symmetric wave functions (group theory symbol A_1), grey bars those of anti-symmetric wave functions (B_1). The color-coding of the transitions matches the sub-spectra shown in Figure 1. The basis wave functions are, in order of sub-spectra:

$\phi_1 = \alpha\alpha\alpha\alpha$ (AA'XX'); $\phi_4 = (1/\sqrt{2})(\alpha\beta+\beta\alpha)\alpha\alpha$; $\phi_{11} = \beta\beta\alpha\alpha$;

$\phi_5 = (1/\sqrt{2})(\alpha\beta-\beta\alpha)\alpha\alpha$;

$\phi_2 = (1/\sqrt{2})\alpha\alpha(\alpha\beta+\beta\alpha)$;

$\phi_7 = (1/\sqrt{2})\ (\alpha\beta+\beta\alpha)(\alpha\beta+\beta\alpha)$; $\phi_8 = (1/\sqrt{2})(\alpha\beta-\beta\alpha)(\alpha\beta-\beta\alpha)$;

$\phi_{12} = (1/\sqrt{2})\beta\beta(\alpha\beta+\beta\alpha)$;

$\phi_{3} = (1/\sqrt{2})\alpha\alpha(\alpha\beta-\beta\alpha)$;

$\phi_{9} = (1/\sqrt{2})\ (\alpha\beta+\beta\alpha)(\alpha\beta-\beta\alpha)$; $\phi_{10} = (1/\sqrt{2})(\alpha\beta-\beta\alpha)(\alpha\beta+\beta\alpha)$;

$\phi_{13} = (1/\sqrt{2})\beta\beta(\alpha\beta-\beta\alpha)$;

$\phi_{6} = \alpha\alpha\beta\beta$; $\phi_{14} = (1/\sqrt{2})(\alpha\beta+\beta\alpha)\beta\beta$;

$\phi_{16} = \beta\beta\beta\beta$;

$\phi_{15} = (1/\sqrt{2})(\alpha\beta-\beta\alpha)\beta\beta$.

Mixing of ϕ_7 and ϕ_8 depends on the relative magnitudes of $J_{AA} + J_{XX}$ (K) to $J_{AX} - J_{AX'}$ (L), likewise mixing of ϕ_9 and ϕ_{10} depends on the relative magnitudes of $J_{AA} - J_{XX}$ (M) to L (for exact formulas see Flynn *et al.* ref. 20). The transition energies are: (1) = $-\frac{1}{2}(J_{AX}+J_{AX'})$; (2,3,8,9) = $\pm\frac{1}{2}K \pm \frac{1}{2}\sqrt{(K^2+L^2)}$; (4,5,10,11) = $\pm\frac{1}{2}M \pm \frac{1}{2}\sqrt{(M^2+L^2)}$; (7) = $+\frac{1}{2}(J_{AX}+J_{AX'})$. In the weak coupling case $|J_{AA}| + |J_{XX}| << L$, resulting in negligible mixing of central two energy levels in the *ab*-parts of the diagram, thus effectively creating an *ax* system. Since in this case $K << L$ and $M << L$, transitions (2,3,8,9) reduce to $\pm\frac{1}{2}K \pm \frac{1}{2}L$, and transitions (4,5,10,11) reduce to $\pm\frac{1}{2}M \pm \frac{1}{2}L$. Thus, in case both J_{AA} and J_{XX} are non-negligible, there will be two *ax*-sub-spectra with either splitting K or M, centered around $\pm \frac{1}{2}L$.

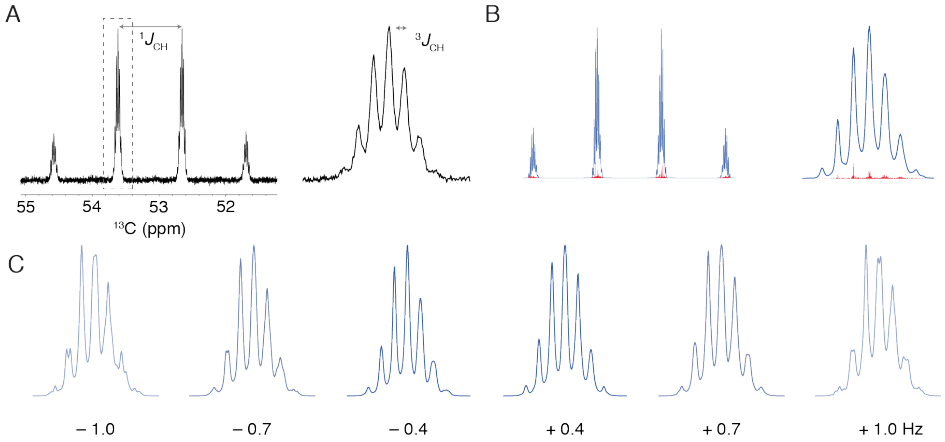


Figure S3. Absence of strong inequivalence effects in the ^{13}C spectrum of compound 1. (A) Experimental ^{15}N-decoupled 1D ^{13}C spectrum, recorded using 2.7 sec acquisition time. (B, C) Simulated spectrum using the composite-particle approach and best-fit values for the J-coupling constants as in Figure 6B, including $^2J_{CC} = 0$ Hz (B) or as indicated

(C). Spectra with $^{2}J_{CC}$ between –0.4 and +0.7 Hz show close correspondence with the experimental spectrum. Due to the lower resolution of the ^{13}C spectrum, these limits are better determined from the ^{1}H spectrum as in Figure 6C. When $^{2}J_{CC}$ is zero, the slight asymmetry in the quartet components is the only visible effect from magnetic inequivalence. The stick plot is shown in red in (B). Acquisition and processing parameters as for the experimental spectrum.

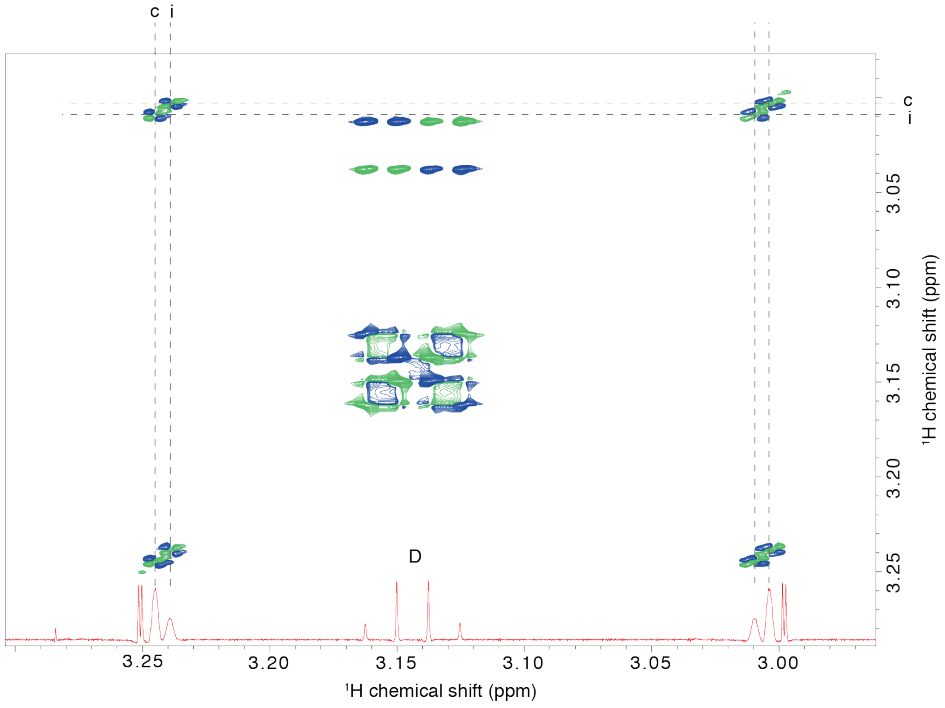


Figure S4. 2D DQF-COSY spectrum of compound 1. Cross-peaks are observed between the center (c) and inner line (i) on opposite sides of the $^{1}J_{CH}$ doublet. No cross-peaks are observed at the outer-line. This pattern is fully consistent with weak coupling between downfield center and upfield inner lines with coupling constant $^{4}J_{HH}$ and *vice versa*, as a result of magnetic inequivalence. The 1D spectrum is shown in red as a reference. The multiplet labeled "D" arises from DIPEA.

A

spin state XX'X"		spin state $A_3A_3'A_3''$								
$m_{T,X}$		DDD	DDQ	DQD	QDD	QQD	QDQ	DQQ	QQQ	sum
+1½	½; ½; ½	a_3	$a_2\boldsymbol{\alpha}$	$a_2\boldsymbol{\alpha}$	$a_2\boldsymbol{\alpha}$	$a\boldsymbol{\alpha}_2$	$a\boldsymbol{\alpha}_2$	$a\boldsymbol{\alpha}_2$	$\boldsymbol{\alpha}_3$	a_9
+½	½; ½; –½	aax	$a\alpha\chi$	$a\alpha x$	$\boldsymbol{\alpha}ax$	$\boldsymbol{\alpha}\alpha x$	$\boldsymbol{\alpha}a\chi$	$a\boldsymbol{\alpha}\chi$	$\boldsymbol{\alpha}\alpha\chi$	a_6x_3
		a_2x	$a_2\chi$	$a\alpha x$	$a\alpha x$	$\boldsymbol{\alpha}_2x$	$a\boldsymbol{\alpha}\chi$	$a\boldsymbol{\alpha}\chi$	$\boldsymbol{\alpha}_2\chi$	
–½	½; –½; –½	ax_2	$ax\chi$	$ax\chi$	$\boldsymbol{\alpha}x_2$	$\boldsymbol{\alpha}x\chi$	$\boldsymbol{\alpha}x\chi$	$\boldsymbol{\alpha}\chi_2$	$\boldsymbol{\alpha}\chi_2$	a_3x_6

B

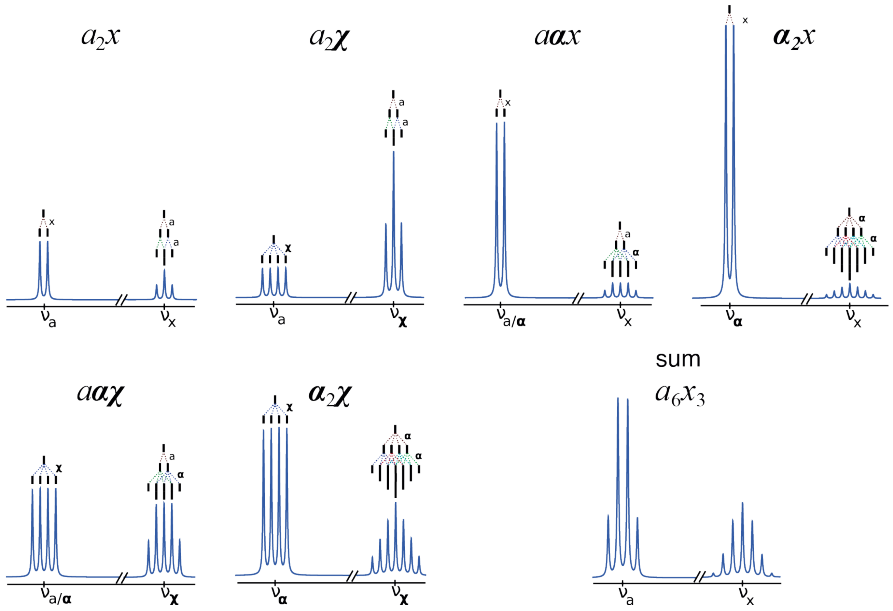
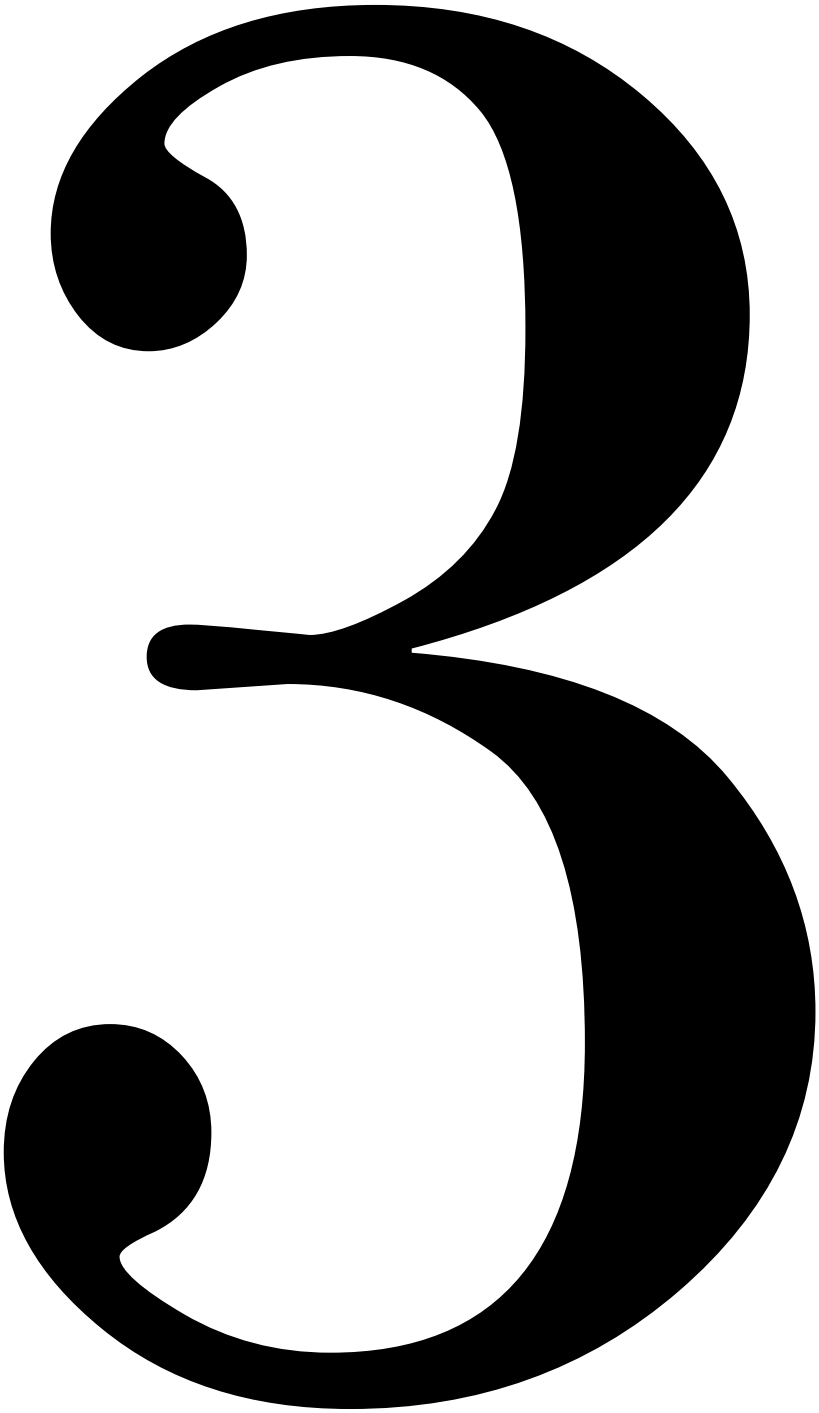


Figure S5. Sub-spectra in the composite particle approach. (A) Overview of different sub-spectra found in the A-spin spectrum of an $[A_3X]_3$ spin system as analyzed with the composite particle approach. The total and individual spin state of X-nuclei is shown in the first two columns. For each spin state of the A_3 groups (either D for doublet, $m_{T,A} = \pm½$, or Q for quartet, $m_{T,A} = \pm{}^3/_2$), shown in columns three to ten, the sub-spectra are classified, assuming *J*-coupling constants as in Figure 6B for compound **1,** where the weak coupling approximation holds. A-spin D-states are labeled as *a* when the attached X-spin is +½, as *x* when the attached X-spin is –½ (as in regular sub-spectral analysis used in the main text). To account for the different total spin, A-spin Q-states are labeled as $\boldsymbol{\alpha}$ when the attached X-spin is +½, as $\boldsymbol{\chi}$ when the attached X-spin is –½. The total sub-spectrum of each X-state is shown in the final column. For $m_{T,X}$ ±½, only one possible X-spin state is shown. Other X-spin states

with the same $m_{T,X}$ will give the same collection of sub-spectra but distributed across the A-spin states differently. The sub-spectra marked in gray are plotted in (B), assuming *J*-coupling constants as in compound **1.** The insets detail how the couplings between a/x (spin-½) and α/χ (spin-$^3/_2$) virtual spins give rise to the multiplet patterns.



Chapter 3

Site-Specific Studies of Nucleosome Interactions by Solid-State NMR

Contributions:
The ssNMR experiments and assignment were done by dr. ShengQi Xiang.
Velten Horn performed the data-driven docking (Leiden University).

Published as:
ShengQi Xiang, Ulric B. le Paige*, Velten Horn, Klaartje Houben, Marc Baldus, Hugo van Ingen (2018). Angew. Chem., Int. Ed. 54, 4571-4575*

Abstract:

Chromatin function depends on a dense network of interactions between nucleosomes and wide range of proteins. A detailed description of these protein-nucleosome interactions is required to reach a full molecular understanding of chromatin function in both genetics and epigenetics. Here, we show that the structure, dynamics and interactions of nucleosomes can be interrogated in a residue-specific manner using state-of-the-art solid-state NMR. Using sedimented nucleosomes, high-resolution spectra are obtained for both flexible histone tails and the non-mobile histone core. Through co-sedimentation of a nucleosome-binding peptide, we demonstrate that protein binding sites on the nucleosome surface can be determined. We believe that this approach holds great promise as it is generally applicable, extendable to include the structure and dynamics of the bound proteins, and scalable to interactions of proteins with higher-order chromatin structures, including isolated or cellular chromatin.

Introduction

Eukaryotic genomes are present in the cell as a complex of DNA and histone proteins, called chromatin. Chromatin's fundamental building block, the nucleosome, is an assembly of a histone octamer protein core around which ~147bp of DNA are wrapped[1,2] (Figure 1A). Nucleosomes compact and protect DNA and, importantly, serve as the binding platform for a multitude of proteins that regulate gene transcription, DNA repair and replication.[3-5] Recently, the structures of various nucleosome-protein complexes have been solved, offering new and exciting insights into chromatin function.[6-18] The wide range in interaction modes in terms of affinities, binding sites, and contribution of dynamics calls for a wide range of techniques to describe the molecular basis of these interactions. The combination of state-of-the-art solid-state NMR (ssNMR) and sedimentation has developed into an attractive method to a obtain detailed characterization of soluble biomolecular systems.[19-25] Sedimentation has been widely used to study folding and compaction of nucleosomes

and nucleosomal arrays.[26-30] The Jaroniec group were the first to realize the great potential of ssNMR in the context of nucleosomes through their study on histone tails in nucleosomal arrays.[31] Here, we introduce the use of nucleosome sedimentation, ultra-fast magic-angle spinning (MAS) and ^{1}H-detected ssNMR to characterize the structure and dynamics of nucleosomes and their protein-complexes. High-quality ssNMR spectra were obtained for both flexible and rigid parts of the nucleosome, allowing near-complete and residue-specific assignment. As proof-of-principle, we exploited the sensitivity of the amide backbone chemical shifts to map the binding surface of the N-terminal segment of the LANA protein on the nucleosome surface.[6,32]

Results

We reconstituted milligram amounts of mononucleosomes from recombinantly expressed core histones (H2A, H2B, H3 and H4) and the strong-positioning '601'-DNA template[33] according to established methods (see Supporting Information for detailed description and Figure S1).[34] The modular nature of the nucleosome (molecular weight ~210 kDa) afforded selective isotope-labelling of one histone protein per sample (here H2A or H3, each ~14 kDa) thereby alleviating signal overlap. Isotope-labeled histones were labeled uniformly with ^{15}N/^{13}C-isotopes in combination with fractional deuteration.[35] Typically, ~2 mg of nucleosomes were sedimented directly into a 1.3 mm ssNMR rotor by ultracentrifugation. Proton 1D spectra established that the sediment is highly hydrated (Supporting Information, Figure S1).

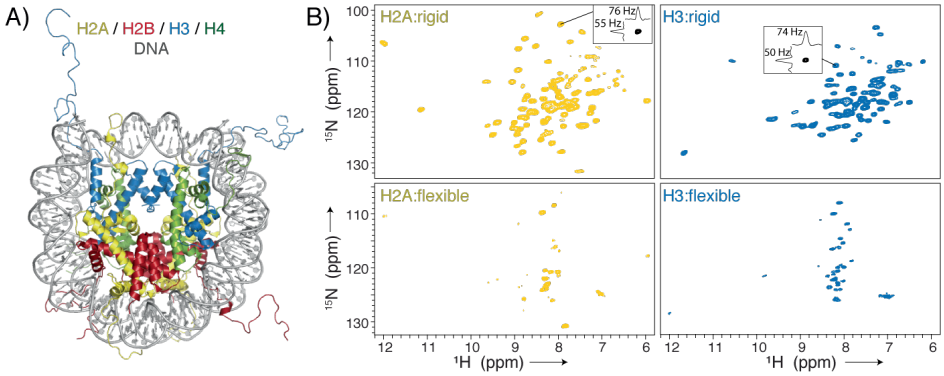


Figure 1. Nucleosome sedimentation yields high-quality protein 'fingerprint' NMR spectra covering both the globular core and flexible tails of histones. A) The nucleosome is formed by an octamer of four core histones, H2A, H2B, H3 and H4, that binds ~147 bp of DNA around its surface (crystal structure from PDBid 1KX5[41]). Color coding indicated in the Figure. B) ^{1}H-detected 2D NH spectra of sedimented nucleosomes with either ^{2}H/^{15}N/^{13}C-labeled H2A or H3, using either dipolar-based (top) or scalar based (bottom) magnetization transfer. Slices along the ^{1}H/^{15}N dimension in the dipolar spectra are shown to highlight typical linewidths.

Sedimented nucleosomes with either labeled H2A or H3 both give excellent 2D ^{1}H-^{15}N correlation spectra in terms of sensitivity, resolution and signal dispersion (Figure 1). The spectral quality suggests that the nucleosomes in the sediment are well-folded and homogeneously arranged on a microscopic scale. Spectra were recorded using either dipolar or scalar coupling-based magnetization transfer to identify and characterize rigid and dynamic parts in the histones, respectivel[36,37]. As expected for folded nucleosomes, the dipolar coupling-based spectra exhibit high chemical shift dispersion in line with a folded core of the histones (top panels in Figure 1B). The scalar coupling-based spectra show fewer resonances that are clustered around 8.2 ppm ^{1}H chemical shift, characteristic for dynamic and unstructured histone tails (bottom panels in Figure 1B). For both H2A and H3, the observed peak pattern is highly similar to that in the solution-state spectra of nucleosomes[38], allowing the assignment of nearly all observed resonances (Supporting Information, Figure S2).

For H3, all resonances correspond to the N-terminal tail (res 3-35), indicating that the H3 tail of sedimented nucleosomes is unstructured and highly flexible, as in solution[39] and in Mg^{2+}-precipitated nucleosomal arrays.[31] For H2A, part of the N- (res. 3-8), and C-terminal tail (res 117-123) are highly dynamic, as judged from their observation in the J-based ssNMR spectrum. Chemical shift changes to isolated H2A/H2B dimer or H3 peptide, where DNA is absent, suggest that both the N-terminal tail of H2A and H3 are (transiently) bound to nucleosomal DNA (Supporting Information, Figure S2), in line with previous NMR studies.[39,40]

The high quality of the spectra, in particular for the core region of the nucleosome, opens a route for the per-residue characterization of histone structure, dynamics and interactions.

We performed a de-novo sequential resonance assignment of H2A within the sedimented nucleosomes, aided by the assignments of H2A in the H2A-H2B dimer in solution (unpublished data).

Using 3D ^{1}H-detected CANH/CA(CO)NH and CONH/ CO(CA)NH experiments,[42] we assigned 93% of the amide backbone resonances between residues 18-112, corresponding to nearly all resonances in the dipolar-based ^{1}H-^{15}N spectrum of H2A (Figure 2A, B and Supporting Information, Figure S3). Notably, the significant resolution and sensitivity gains afforded by ^{1}H compared to ^{13}C-detection were instrumental herein (Supporting Information, Figure S4). Based on the assigned H, N, Cα and C' chemical shifts, the H2A secondary structure propensities were analyzed using TALOS.[43] The characteristic three-helix histone fold is clearly present, as well as the two β-strands that mediate interactions with H2B in the H2A-H2B dimer (Figure 2C). Importantly, the C-terminal β-strand, which forms a 'docking' sheet with H4 in the nucleosome crystal structure,[44] is also observed. This region of H2A, the docking sequence (residues 97-116), binds to H3-H4 and thus stabilizes the histone octamer. In our dipolar ssNMR spectra, resonances for several stretches in this region, including part of the 3_{10}-helix observed in the crystal, are missing, suggesting increased mobility. In addition, this region has, compared to the histone-

fold region, slightly decreased order parameters, as predicted from the chemical shifts (Figure 2C).[45]

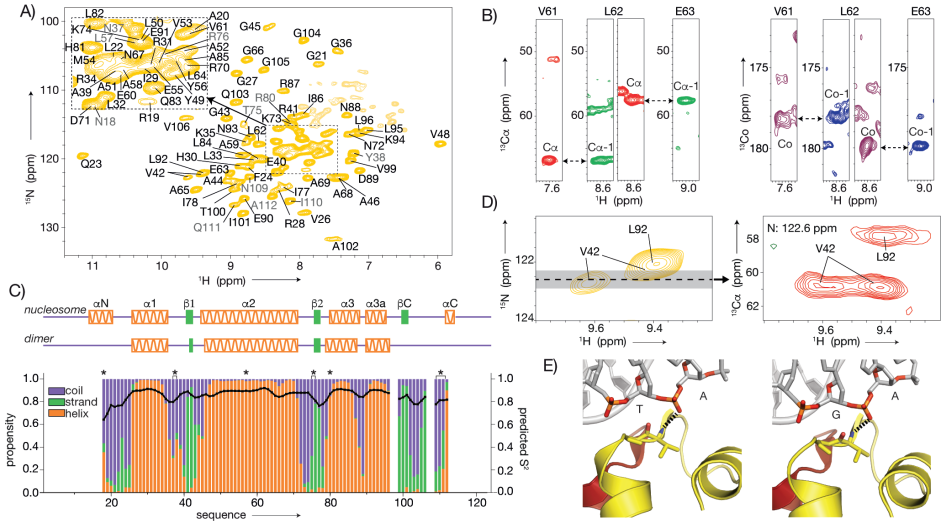


Figure 2: Resonance assignment and secondary structure of H2A in sedimented nucleosomes. A) 2D NH spectrum with (tentative) assignments indicated in (grey) black. Side chains resonances in light colors. B) Representative strips illustrating the sequential backbone assignment based on 3D CANH and CA(CO)NH (red and green) or CONH and CO(CA)NH (magenta and blue) spectra. C) Secondary structure propensities (colored bars) and predicted S^2 values (black line) based on assigned backbone chemical shifts. Secondary structure in the nucleosome crystal structure (PDBid 2PYO[44]) and isolated H2A-H2B dimer (unpublished results) shown at the top. Asterisks indicate tentative assignments. D, E) Peak doubling of V42 observed in the 2D NH (left) and 3D CANH (right) spectra (panel D), correlates to the asymmetric environment in 601-nucleosomes (PDBid 3LZ0[46]). Dashed lines indicate hydrogen-bonds (panel E).

Interestingly, residues 15-17 that are part of the N-terminal α-helix of the crystal structure are absent in both dipolar and J-based spectra and residues 18-20 exhibit low helical propensity and low predicted order parameters, suggesting the αN-helix is not a stably folded structure (Figure 2C). Comparison of the chemical shifts of H2A in sedimented nucleosomes to those of the H2A-H2B dimer in solution shows significant changes for regions that interact with either other histones or DNA

(Supporting Information, Figure S5). As an example of the sensitivity to the chemical environment afforded by ^{1}H detection, we note the peak doubling observed for V42 (Figure 2D). Because the 601-DNA is not palindromic, this residue is hydrogen-bonded to either a TpA or GpA base pair step in the two copies of H2A in the nucleosome (Figure 2E).[46]

We next used the assigned amide backbone resonances as sensitive reporters on nucleosome-protein interactions. We chose to study the interaction with the N-terminal domain of the viral LANA protein, for which a crystal structure of the complex with the nucleosome is available.[6,32] LANA binds to the acidic patch, a negatively charged region on the nucleosome surface formed by acidic residues from H2A and H2B[1,5]. A 20-residue peptide derived from the LANA N-terminus binds tightly (K_D 0.16 μM) to nucleosomes, in agreement with previous studies using longer fragments[47] (Supporting Information, Figure S6). The LANA peptide (2 kDa) was added in five-fold molar excess to a dilute solution of H2A-labeled nucleosomes from the same batch as the sample used for assignment, followed by sedimentation into the rotor. Comparison of the remaining supernatant and pure LANA peptide solution by NMR suggested that the peptide was successfully co-sedimented with the nucleosome (Supporting Information, Figure S6). The dipolar ^{1}H-^{15}N correlation spectrum of the nucleosome:LANA complex, measured under identical conditions to the spectrum of free nucleosomes, shows perturbations specifically for a subset of resonances (Figure 3A and Supporting Information, Figure S7). Importantly, there is no peak doubling upon binding, indicating that LANA is bound to both H2A copies in the nucleosome whereas only one copy is bound in the crystal structure.[6] Transfer of H2A assignments to the LANA-bound state was aided by making use of ^{13}Cα chemical shifts from a 3D CANH spectrum. Lack of significant changes in Cα shifts indicate there is little to no structural change in H2A upon binding and underscores the benefit of ^{1}H-detection. Thus, an unambiguous comparison of the ^{15}N/^{1}H chemical shifts in free and LANA-bound states could be made for the majority of H2A residues (Figure 3B).

Notably, resonances from the acidic patch residues E55, E63, E90 and E91 show clear chemical shift changes. In addition, a number of residues with L92 in Figure 3A given as example, display significant peak broadening in the LANA-bound state, either reflecting increased local ^{1}H density or local structural or dynamic disorder as result of binding. Displaying the observed chemical shift perturbations (CSP) on the crystal structure reveals a striking co-localization of the changes with the interaction surface of LANA seen in the crystal (Figure 3C).[6] The highest CSPs are observed for the residues in the center of the acidic patch, which contacts the LANA arginine anchor.

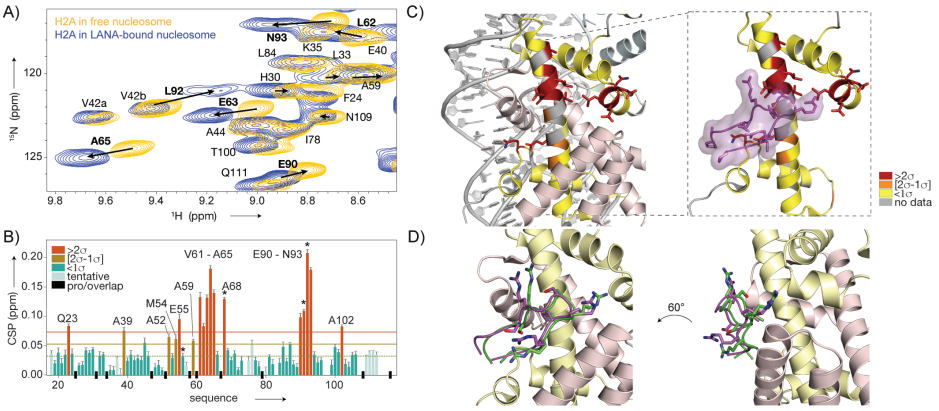


Figure 3. Mapping of binding site of LANA on the nucleosome surface by ssNMR. A) Overlay of 2D NH spectra of H2A in sedimented nucleosomes (yellow) and nucleosome-LANA complexes (blue). Residues with significant chemical shift changes labelled in bold, peak displacement indicated with arrows. B) Weighted chemical shift perturbation (CSP) per residue, color coding in the Figure. The 10% trimmed mean (dashed line) plus one (orange line) or two (red line) standard deviations (σ) are indicated. Residues with significant peak broadening in the bound state are labelled with asterisks. C) CSPs color coded on the structure of H2A. Residues with high CSPs co-localize with the binding site of LANA (PDBid 1ZLA). D) NMR-data driven model of the LANA-nucleosome complex, showing the best-scoring water-refined solution (green). This model corresponds within 1.5 Å backbone RMSD to the crystal structure (magenta).

We next used our ssNMR data and published mutagenesis data on the LANA-nucleosome interaction[6] to generate a molecular model of the complex using HADDOCK.[48] The best scoring solution in terms of energetics and agreement with experimental data closely resembles the binding mode observed in the crystal, illustrating the potential of this approach to reveal the architecture of nucleosome-protein complexes (Figure 3D and Supporting Information, Figure S8). We note that for this approach to work well, the structures of the two interacting molecules must be known and the conformational changes upon binding must be limited.[49,50]

Discussion

We have shown the first high-resolution solid-state NMR study on a non-mobile region of the nucleosome and demonstrate the accurate mapping of binding sites on the nucleosome surface, the main goal of this study. High-quality spectra of histone proteins in sedimented nucleosomes allowed us to obtain near-complete assignments of histone H2A. Our data indicate that the H2A N- and C-terminal regions, including the αN element and docking domain, have increased flexibility compared to the core, and that its N- and C-terminal tail are highly dynamic, yet DNA-bound. It is noteworthy that these regions are rich in post-translational modifications, suggesting that the structural flexibility may be exploited by the enzymes that install, read or remove these epigenetic marks. In addition, this is the first high-resolution study of histone structure and nucleosome-binding in the crowded, solvated environment of the sediment, thereby mimicking the cellular conditions. Next to these biochemical insights, our approach provides a starting point for the more detailed analysis of the structure and dynamics of nucleosome-complexes using the entire toolbox of biomolecular ssNMR. While the solution NMR methyl-TROSY approach[51] is superior in terms of spectral sensitivity and resolution,[9,10] the approach described here allows the observation of all non-proline residues in the interface and poses fewer labeling requirements. Even though fractional deuteration and considerable measurement time are required, a major advantage is ssNMR's lack of an intrinsic size

limit, making the approach extendable to studies of larger complexes, including larger chromatin substrates such as nucleosomal arrays. We anticipate that in particular for dynamic complexes, for interactions with smaller proteins that cannot easily be captured by cryo-EM or crystallography, or when plasticity of histones is thought to play a role[52], the use of sedimented nucleosome-protein complexes will be an effective and complementary method in the structural biology of chromatin function.

References

1 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

2 Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285-294, doi:10.1016/S0092-8674(00)81958-3 (1999).

3 Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* **13**, 436-447, doi:10.1038/nrm3382 (2012).

4 McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem Rev* **115**, 2255-2273, doi:10.1021/cr500373h (2015).

5 McGinty, R. K. & Tan, S. Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol* **37**, 54-61, doi:10.1016/j.sbi.2015.11.014 (2016).

6 Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**, 856-861, doi:10.1126/science.1120541 (2006).

7 Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562-566, doi:10.1038/nature09321 (2010).

8 Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 A resolution. *Science* **334**, 977-982, doi:10.1126/science.1210915 (2011).

9 Kato, H. *et al.* Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A* **108**, 12283-12288, doi:10.1073/pnas.1105848108 (2011).

10 van Nuland, R. *et al.* Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin* **6**, 12, doi:10.1186/1756-8935-6-12 (2013).

11 Kato, H. *et al.* A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**, 1110-1113, doi:10.1126/science.1235532 (2013).

12 McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514**, 591-596, doi:10.1038/nature13890 (2014).

13 Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366-369, doi:10.1038/nature14495 (2015).

14 Wilson, M. D. *et al.* The structural basis of modified nucleosome recognition by 53BP1. *Nature* **536**, 100-103, doi:10.1038/nature18951 (2016).

15 Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440-445, doi:10.1038/nature22036 (2017).

16 Farnung, L., Vos, S. M., Wigge, C. & Cramer, P. Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature* **550**, 539-542, doi:10.1038/nature24046 (2017).

17 Kitevski-LeBlanc, J. *et al.* The RNF168 paralog RNF169 defines a new class of ubiquitylated histone reader involved in the response to DNA damage. *Elife* **6**, doi:10.7554/eLife.23872 (2017).

18 Hu, Q., Botuyan, M. V., Cui, G., Zhao, D. & Mer, G. Mechanisms of Ubiquitin-Nucleosome Recognition and Regulation of 53BP1 Chromatin Recruitment by RNF168/169 and RAD18. *Mol Cell* **66**, 473-487 e479, doi:10.1016/j.molcel.2017.04.009 (2017).

19 Luca, S. *et al.* The conformation of neurotensin bound to its G protein-coupled receptor. *Proceedings of the National Academy of Sciences* **100**, 10706-10711, doi:10.1073/pnas.1834523100 (2003).
20 Bertini, I. *et al.* Solid-state NMR of proteins sedimented by ultracentrifugation. *Proceedings of the National Academy of Sciences* **108**, 10396-10399, doi:10.1073/pnas.1103854108 (2011).
21 Gardiennet, C. *et al.* A Sedimented Sample of a 59 kDa Dodecameric Helicase Yields High-Resolution Solid-State NMR Spectra. *Angew. Chem.-Int. Edit.* **51**, 7855-7858, doi:10.1002/anie.201200779 (2012).
22 Mainz, A. *et al.* NMR Spectroscopy of Soluble Protein Complexes at One Mega-Dalton and Beyond. *Angew. Chem.-Int. Edit.* **52**, 8746-8751, doi:10.1002/anie.201301215 (2013).
23 Lamley, J. M. *et al.* Solid-State NMR of a Protein in a Precipitated Complex with a Full-Length Antibody. *Journal of the American Chemical Society* **136**, 16800-16806, doi:10.1021/ja5069992 (2014).
24 Barbet-Massin, E., Huang, C.-T., Daebel, V., Hsu, S.-T. D. & Reif, B. Site-Specific Solid-State NMR Studies of "Trigger Factor" in Complex with the Large Ribosomal Subunit 50S. *Angewandte Chemie International Edition* **54**, 4367-4369, doi:10.1002/anie.201409393 (2015).
25 Dannatt, H. R. W. *et al.* Weak and Transient Protein Interactions Determined by Solid-State NMR. *Angewandte Chemie International Edition* **55**, 6638-6641, doi:10.1002/anie.201511609 (2016).
26 Hansen, J. C., Ausio, J., Stanik, V. H. & Van Holde, K. E. Homogeneous reconstituted oligonucleosomes, evidence for salt-dependent folding in the absence of histone H1. *Biochemistry* **28**, 9129-9136, doi:10.1021/bi00449a026 (2002).
27 Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847, doi:10.1126/science.1124000 (2006).
28 Bertin, A., Mangenot, S., Renouard, M., Durand, D. & Livolant, F. Structure and phase diagram of nucleosome core particles aggregated by multivalent cations. *Biophys J* **93**, 3652-3663, doi:10.1529/biophysj.107.108365 (2007).
29 Korolev, N. *et al.* Electrostatic origin of salt-induced nucleosome array compaction. *Biophys J* **99**, 1896-1905, doi:10.1016/j.bpj.2010.07.017 (2010).
30 Maeshima, K. *et al.* Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *The EMBO Journal* **35**, 1115-1132, doi:10.15252/embj.201592660 (2016).
31 Gao, M. *et al.* Histone H3 and H4 N-terminal tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy. *J Am Chem Soc* **135**, 15278-15281, doi:10.1021/ja407526s (2013).
32 Chodaparambil, J. V. *et al.* A charged and contoured surface on the nucleosome regulates chromatin compaction. *Nat Struct Mol Biol* **14**, 1105-1107, doi:10.1038/nsmb1334 (2007).
33 Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19-42, doi:10.1006/jmbi.1997.1494 (1998).
34 Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **375**, 23-44, doi:10.1016/S0076-6879(03)75002-2 (2004).
35 Mance, D. *et al.* An Efficient Labelling Approach to Harness Backbone and Side-Chain Protons in 1H-Detected Solid-State NMR Spectroscopy. *Angewandte Chemie International Edition* **54**, 15799-15803, doi:10.1002/anie.201509170 (2015).

36 Andronesi, O. C. *et al.* Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *Journal of the American Chemical Society* **127**, 12965-12974, doi:10.1021/ja0530164 (2005).
37 Ward, M. E. *et al.* Proton detection for signal enhancement in solid-state NMR experiments on mobile species in membrane proteins. *Journal of Biomolecular NMR* **63**, 375-388, doi:10.1007/s10858-015-9997-5 (2015).
38 Zhou, B. R. *et al.* Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J Mol Biol* **421**, 30-37, doi:10.1016/j.jmb.2012.04.032 (2012).
39 Stutzer, A. *et al.* 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, doi:10.1016/j.molcel.2015.12.015 (2016).
40 Kato, H., Gruschus, J., Ghirlando, R., Tjandra, N. & Bai, Y. Characterization of the N-terminal tail domain of histone H3 in condensed nucleosome arrays by hydrogen exchange and NMR. *J Am Chem Soc* **131**, 15104-15105, doi:10.1021/ja9070078 (2009).
41 Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *Journal of Molecular Biology* **319**, 1097-1113, doi:10.1016/s0022-2836(02)00386-8 (2002).
42 Zhou, D. H. *et al.* Solid-state NMR analysis of membrane proteins and protein aggregates by proton detected spectroscopy. *Journal of Biomolecular NMR* **54**, 291-305, doi:10.1007/s10858-012-9672-z (2012).
43 Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR* **56**, 227-241, doi:10.1007/s10858-013-9741-y (2013).
44 Clapier, C. R. *et al.* Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **71**, 1-7, doi:10.1002/prot.21720 (2008).
45 Berjanskii, M. V. & Wishart, D. S. A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. *Journal of the American Chemical Society* **127**, 14970-14971, doi:10.1021/ja054842f (2005).
46 Vasudevan, D., Chua, E. Y. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1-10, doi:10.1016/j.jmb.2010.08.039 (2010).
47 Beauchemin, C., Moerke, N. J., Faloon, P. & Kaye, K. M. Assay Development and High-Throughput Screening for Inhibitors of Kaposi's Sarcoma-Associated Herpesvirus N-Terminal Latency-Associated Nuclear Antigen Binding to Nucleosomes. *J Biomol Screen* **19**, 947-958, doi:10.1177/1087057114520973 (2014).
48 van Dijk, A. D., Boelens, R. & Bonvin, A. M. Data-driven docking for the study of biomolecular complexes. *FEBS J* **272**, 293-312, doi:10.1111/j.1742-4658.2004.04473.x (2005).
49 Bonvin, A. M. J. J. Flexible protein-protein docking. *Curr Opin Struct Biol* **16**, 194-200, doi:10.1016/j.sbi.2006.02.002 (2006).
50 Lensink, M. F., Velankar, S. & Wodak, S. J. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* **85**, 359-377, doi:10.1002/prot.25215 (2017).
51 Ollerenshaw, J. E., Tugarinov, V. & Kay, L. E. Methyl TROSY: explanation and experimental verification. *Magn Reson Chem* **41**, 843-852, doi:10.1002/mrc.1256 (2003).
52 Sinha, K. K., Gross, J. D. & Narlikar, G. J. Distortion of histone octamer core promotes nucleosome mobilization by a chromatin remodeler. *Science* **355**, doi:10.1126/science.aaa3761 (2017).
53 Speranzini, V., Pilotto, S., Sixma, T. K. & Mattevi, A. Touch, act and go: landing and operating on nucleosomes. *EMBO J* **35**, 376-388, doi:10.15252/embj.201593377 (2016).

54 Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**, 227-241, doi:10.1007/s10858-013-9741-y (2013).
55 van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725, doi:10.1016/j.jmb.2015.09.014 (2016).
56 Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **125**, 10420-10428, doi:10.1021/ja030153x (2003).

Supplementary information

Experimental procedures - Nucleosome sample production

Histone production

Escherichia coli (*E. coli*) Rosetta 2 (DE3) was used to express either *Drosophila melanogaster* (*Dm.*) H2A, H3 and H4 cloned in a pET-21 vector, or a codon optimized gene for *Dm.* H2B cloned in a His-tag-depleted pET-28. Unlabeled histones were produced in Lysogeny Broth[1] (LB), labelled histones for proton detection in deuterated minimal medium (M9 – 100% D_2O, 2g/L ^{1}H-^{13}C glucose, 1g/L $^{15}NH_4Cl$). Expression was induced upon reaching an OD of 1.5, by addition of 0.5mM IPTG and ensured for 4 to 6 hours.

Histones were extracted and purified according to Dyer et al.[2] with additional anion exchange chromatography to remove DNA contamination if necessary.[3] Pure individual histones were stored at -80 °C in either IEX buffer (7 M urea, 50 mM $NaPO_4$ buffer pH 7.5, 150 mM NaCl, 1 mM EDTA) supplemented with 1 mM lysine or unfolding buffer (6 M GdnCl, 50 mM Tris-HCl pH 7.5, 100 mM NaCl) at concentrations between 3 and 5 mg/mL. In the case of H3, all buffers were supplemented with 1 mM 2-mercaptoethanol (2-ME).

DNA production

A pUC19 plasmid containing 12 repeats of a 167bp 601 Widom sequence was amplified in *E. coli* DH5α and extracted by alkaline lysis followed by isopropanol and ethanol precipitations as described in Dyer et al.[2] The pellet was dissolved in TE buffer (20 mM Tris-HCl pH 7.5, 5mM EDTA, 100mM NaCl) and purified by anion exchange chromatography. The purified plasmid was restricted with ScaI (ThermoFisher) in home-made buffer for 36 hours at room temperature, with three additions of 700 units of restriction enzyme per 20mg of plasmid. The restricted fragment was purified by anion exchange chromatography.

Histone octamer refolding and reconstitution of mononucleosomes

Histone octamers were refolded according to Dyer et al[2] with the following modification. After preparation of an initial mix of histones (with isotope-labeling as in Table S1 and 5% excess of H2A/H2B) in denaturing conditions based on absorbance and theoretical extinction coefficients, equimolarity of this mix was ensured using an SDS-PAGE based approach, where integration of the bands was compared with intensity ratios determined from preliminarily refolded pure nucleosomes (see Supplementary Figure S1A and S1B). The refolded and purified histone octamers were used together with 167bp 601 DNA to reconstitute mononucleosomes according to Dyer et al[2]. The preparation was directly dialyzed to PK buffer (10 mM PO_4 pH 6.5, 10 mM KCl) for 12 hours with two buffer changes, concentrated to 8 mg/mL and stored at 4 °C until use. Efficiency of the reconstitution was assessed by native and SDS PAGE (Supplementary Figure 1C).

Table S1 : Isotope-labelling scheme of the nucleosome samples used in this study

	H2A - ^{1}H detection	H3 - ^{1}H detection	H3 - ^{13}C detection
H2A	Frac. ^{2}D[a], U-^{13}C, ^{15}N	Natural abundance	Natural abundance
H2B	Natural abundance	Natural abundance	Natural abundance
H3	Natural abundance	Frac. ^{2}D, U-^{13}C, ^{15}N	U-^{13}C, ^{15}N
H4	Natural abundance	Natural abundance	Natural abundance

[a] Frac. ^{2}D refers to fractionally deuterated, i.e. produced in D_2O with protonated glucose.

Mononucleosome sedimentation

An ultracentrifuge funnel device similar to the ones described in references[4,5] loaded with a 3.2mm thick walled Zirconia rotor or 1.3 mm Zirconia rotor (Bruker) for ^{13}C-detected and ^{1}H-detected experiments, respectively. The funnel was filled with 2.2 mg

of mononucleosome in PK buffer. Magnesium was added in the form of an equal volume of PK buffer supplemented with 4 mM $MgCl_2$, for a final concentration of 2 mM Mg^{2+}. The device was loaded in a swinging bucket ultracentrifuge and sedimented for 18 hours at 4 °C. The supernatant was then removed by pipetting, the rotor recovered and its top cleared before closing. Sedimentation efficiency was assessed by UV absorbance measurement from the sedimentation supernatant to be between 94 and 99%.

Interaction of LANA 2-22 with the nucleosome

Peptide preparation

A peptide corresponding to residues 2-22 from the Kaposi's sarcoma herpesvirus Latency Associated Nuclear Antigen (LANA, Uniprot ID: E5LC01; sequence: APPGMRLRSGRSTGAPLTRGS) was ordered from Biomatik as TFA salt with N-terminal acetylation, C-terminal amidation and with >95% purity. The peptide was dissolved in mQ water and pH was adjusted to 6.7 using diluted KOH. The peptide stock concentration was measured using ^{1}H NMR to be 2.99 mM. Briefly, an accurate volume of peptide solution was diluted in PK buffer, including 7% D2O and a known concentration of trimethylsilylpropanoic acid (TSP) as reference. A 1D ^{1}H spectrum was measured with a recycle delay of 60 s, ensuring proper relaxation of the TSP signals. The concentration of peptide in this NMR sample was determined by integration of the isolated leucine methyl peaks compared to the TSP methyl peak.

Microscale thermophoresis

Unlabeled mononucleosomes were fluorescently tagged using the Red-NHS Monolith NT kit (Nano Temper) and were titrated against increasing concentrations of LANA 2-22 in TKB buffer (10 mM Tris-HCl pH 7.5, 100 mM KCl, 0.1 mg/ml BSA fraction V). A serial dilution of LANA was mixed with a final concentration of 135 nM mononucleosomes, and the thermophoresis was followed in a Monolith NT.115 (NanoTemper Technologies) using hydrophobic capillaries. The LANA concentration-dependent change in thermophoresis was fitted using an in-house Matlab script in order to extract the binding constant (Supplementary Figure S6)[6]

Peptide co-sedimentation

Mononucleosomes (2.2 mg in 240 µL PK buffer, final concentration : 21 µM) were loaded in the funnel device, supplemented with a 5-molar excess of LANA (0.016 µL from stock, final concentration : 96.5 µM) and incubated at room temperature for 20 min. Magnesium addition, sedimentation and rotor preparation were made as described above. Peptide co-sedimentation was assessed by solution-state NMR (Supplementary Figure S6). A 1D ^{1}H spectrum of the sedimentation supernatant, supplemented with 7% D_2O, was recorded and compared with a 1D ^{1}H spectrum of pure LANA at the same concentration of the original sedimented mix (96.5 µM).

NMR Spectroscopy

Solution-state NMR

All solution-state spectra of LANA peptide and mononucleosomes were acquired on a Bruker 14.09 T (600 MHz ^{1}H Larmor frequency) spectrometer equipped with an Avance III console and either a 5 mm CPTCI or TXI probe.

Solid-state NMR

The ^{13}C detected ssNMR experiments were carried out at 14 kHz MAS rate on a 22.31 T (950 MHz ^{1}H Larmor frequency) Bruker Advance III HD spectrometer equipped with a 3.2mm $^{1}H/^{13}C/^{15}N$ triple-resonance probe at 278K effective temperatures. The ^{1}H to ^{13}C CP transfers were carried out with a 50%-100% ramp of 86/49.9 kHz on the $^{1}H/^{13}C$-channel. The ^{13}C -^{13}C mixing time in PARIS spin diffusion experiments were set to 30ms and 120ms favor either intra- or inter-residue correlations. The SPINAL decoupling scheme[7] was applied at 83.3 kHz on ^{1}H channel during both ^{13}C evolution periods. The carrier of $^{13}C/^{1}H$ was set to 56/5.5 ppm with spectral widths for indirect/direct ^{13}C-dimension to 123.8/402.4 ppm. Acquisition times were set to 3.7 and 10 ms for indirect and direct dimensions, respectively.

All ^{1}H-detected ssNMR experiments were carried out at 60 kHz MAS rate and 308 K effective temperature on an 18.8 T (800 MHz ^{1}H Larmor frequency) Bruker Avance III HD spectrometer equipped with a 1.3 mm $^{1}H/X/Y$ triple-resonance MAS probe.

The temperature was calibrated by an external DSS sample and monitored during measurements through the water signal position.

The 2D J-based HSQC spectra were recorded with a decoupled HSQC pulse program[8] with 5.4 ms INEPT transfer times. During ^{15}N evolution time, a single hard 180° pulse on ^{1}H was used to refocus the J_{NH}-coupling, while ^{15}N-decoupling during ^{1}H detection was achieved using the WALTZ-16 sequence.[9] For the 2D dipolar-based HN spectra, the PISSARRO decoupling[10] sequence was employed in both dimensions.

The 3D CANH, CA(CO)NH, CONH and CA(CO)NH experiments were implemented as described previously.[11] Transfer and acquisition parameters are listed in the Tables 1 and 2, respectively. The PISSARRO decoupling sequence was used for ^{1}H decoupling during the ^{13}C and ^{15}N evolution periods, and for ^{15}N-decoupling during ^{1}H detection. The measurement times of CANH and CONH were ~5 days, while the CA(CO)NH and CO(CA)NH cost ~1 week. All data sets were recorded in blocks of 1-day length. The drift of external magnetic field was compensated by monitoring the H_2O ^{1}H signal through 1D measurements between each block. Data points in both indirect dimensions were extended twice using the HMSist software package.[12] The data sets were subsequently processed by NMRpipe[13] and analyzed using Sparky[14].

Assignments

Assignment of H2A was performed manually using standard sequential backbone assignment approach. In total 77 resonances in the dipolar-based spectra could unambiguously be assigned based on the set of 3D CANH/CA(CO)NH and CONH/CO(CA)NH. An additional set of 11 resonances for which part of the connectivity information is missing due to overlap or low signal-to-noise of cross peaks were tentatively assigned. In total, this corresponds to an assignment completeness for the core region between residues 18-112 of 93% (or 80% considering only unambiguous assignment). Assigned backbone chemical shifts of

H2A in sedimented nucleosomes were validated by and are deposited in the Biological Magnetic Resonance Bank (BMRB) under accession number 27346.

Chemical shift perturbation analysis

Average, weighted chemical shift perturbations (CSPs) were calculated using Eq. 1:

$$CSP = \frac{1}{2}\sqrt{(\Delta\delta_H)^2 + \left({}^{\Delta\delta_N}/_{6.51}\right)^2} \quad [1]$$

where $\Delta\delta_{H,N}$ is the chemical shift change in ppm along the ^{1}H or ^{15}N dimension. The typical error in the experimental peak position, in either ^{1}H or ^{15}N dimension, $\sigma_{H,N}$, was used to calculate the error in the CSP using standard error-propagation techniques. The typical error in the peak position was estimated using Eq. 2:

$$\sigma_{H,N} = \frac{1}{M}\sum_{i=1}^{M}\frac{lw_{H,N}}{SN} \quad [2]$$

where $lw_{H,N}$ is the linewidth in ppm along the ^{1}H or ^{15}N dimension, SN is the signal-to-noise ratio, and the summation runs over M=54 cross peaks, from both apo and bound-state spectra, that could reliably be integrated. The resulting values were 0.0029 ppm for σ_H and 0.052 ppm for σ_N.

All figures representing structures were made using PyMol [15]

Data-driven docking of LANA on the nucleosome surface

The structure of the peptide was extracted from the crystal structure (PDBid 1ZLA[16]) and docked on the *Drosophila melanogaster* nucleosome structure (PDBid 2PYO[17]) using HADDOCK[18]. Active residues for the peptide were determined according to the mutagenesis study present in Barbera et al[16] – namely res. 5-6 and 8-11 – while passive residues were res. 4 and 12-14. Active residues for the nucleosome were defined as solvent-exposed residues with chemical shift perturbation larger than 2 standard deviations from the 10% trimmed mean (res. 23, 55, 63, 64, 68, 90-93, see Main Text Figure 3B). Passive residues were determined automatically by HADDOCK as surface neighbors of active residues. The default peptide-docking procedure was followed without random 50% exclusion of restraints. All solutions were ranked according to

their HADDOCK score, which is a weighted sum of van der Waals, electrostatic, desolvation, binding and restraint energies, and the buried surface area. Out of 200 water-refined solutions, 98% were clustered into a single cluster according to fraction of common contacts (FCC).

Table S2. Dipolar-transfer ssNMR parameters for ^{1}H detected experiments.

Transfer steps	Contact time(μs)	Pulses
H->C'	3500	H:96.8 kHz (90-100 % ramp on H)/C:42 kHz
H->Cα	3500	H:102 kHz (90-100 % ramp on H)/C:42 kHz
H->N	1400	H:114.1kHz (80-100 % ramp on H)/N:40.4kHz
C'->Cα	4000	27.7 kHz (100-80 % ramp on Cα)
Cα->C'	4000	28.7kHz (100-80 % ramp on C')
C'->N	6000	N:21kHz/C': 42 kHz (100-90 % ramp on C')
Cα->N	6000	N:21kHz/Cα:42kHz (100-80 % ramp on Cα)
N->H^N	750	H:104.6 kHz (100-80 % ramp on H)/N:40.4 kHz

Table S3. Acquisition parameters of ^{1}H detected ssNMR experiments

Spectrum	Nucleus	Spectral width (Hz)		Maximum acquisition time (ms)	
		H2A	H3	H2A	H3
2D NH	N	3244	3244	20	20
	H^N	16025	16025	20	20
3D CANH	Cα	5836	5635	2.5	2.5
	N	2838	2514	6.0	6.0
	H^N	16025	16025	20.0	20.0
3D CONH	CO	2415		4.0	
	N	2838		6.0	
	H^N	16025		20.0	
3D CA(CO)NH	Cα	5836		3.0	
	N	2838		6.0	
	H^N	16025		20.0	
3D CO(CA)NH	CO	2415		4.0	
	N	2838		6.0	
	H^N	16025		20.0	

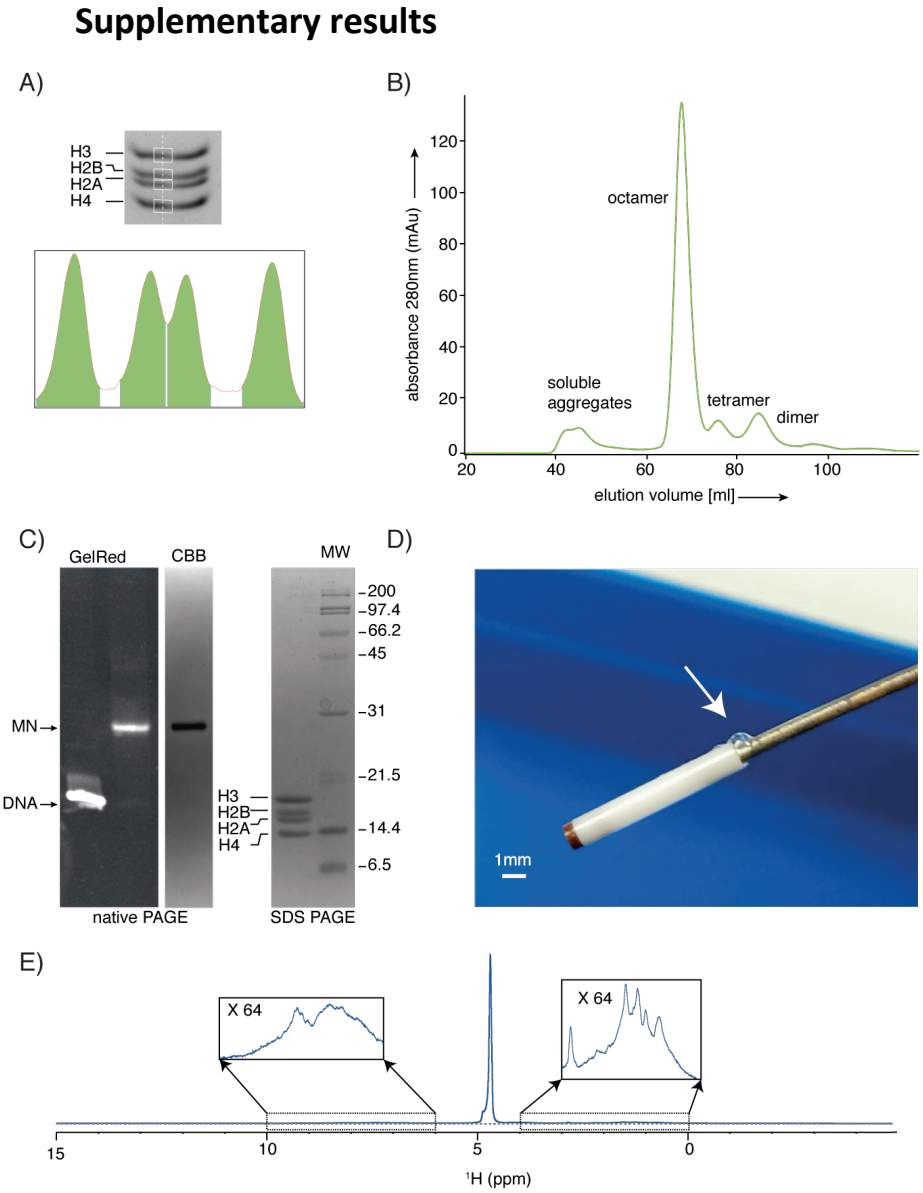


Figure S1. Nucleosome production and sample characterization. A) SDS-PAGE-based equimolarity adjustment of the histone refolding mix. The ratio of integrated band intensities was compared with the optimal ratio derived from a previous nucleosome sample. Here, the gel after adjustment is shown. B) Representative gel filtration chromatogram from the purification of refolded and isotope-labeled octamers. C) Analysis of nucleosome

reconstitution with (left) native PAGE post-stained with GelRed and Coomassie Brilliant Blue (CBB) and (right) CBB stained SDS-PAGE (right). Equivalent sample quality was obtained for each preparation. MW = molecular weight marker; MN = mononucleosome, position of free 601-DNA template is indicated. D) Picture of the 1.3 mm rotor filled with sedimented nucleosomes. The arrow points to a droplet of the viscous transparent gel formed through sedimentation, here visible during the removal of excess of sample. E) 1D ^{1}H spectrum of sedimented nucleosomes. The dominant signal at ~4.7 ppm is from the water in the sample. The two boxes zoom in the amide and aliphatic proton regions. The high relative intensity of the water signal confirms the hydrated nature of the sample

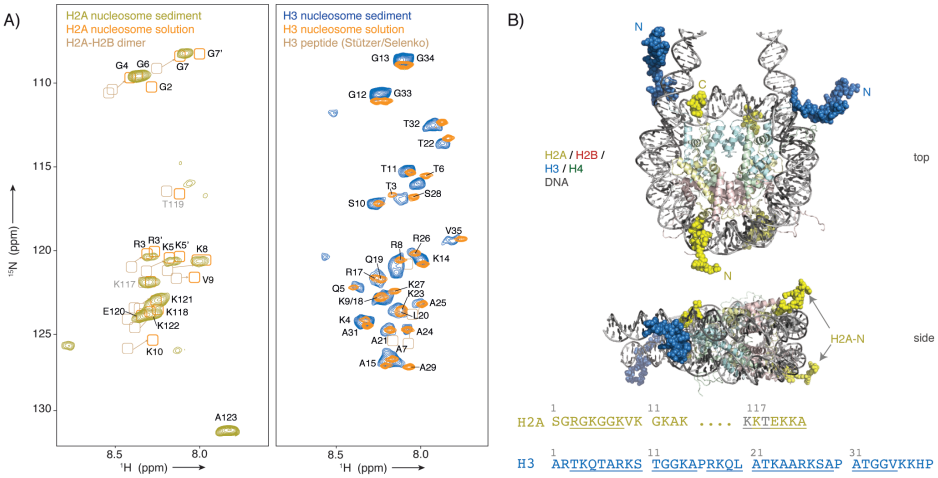


Figure S2. The H2A and H3 tails have similar properties in both solution and sediment. A) Left panel: 2D NH HSQC spectra of H2A from sedimented nucleosomes with chemical shifts from H2A tails in nucleosomes[19] and H2A-H2B dimer highlighted as boxes. Significant peak displacements from the isolated dimer to the nucleosome most likely result from DNA binding. Notably, the 601-DNA sequence is highly asymmetric around the H2A N-terminal but not around the C-terminal tail, matching the observed peak doubling pattern. Right panel: overlay of 2D NH HSQC spectra of H3 within nucleosomes either in solution (blue) and sediment (color) with chemical shifts from a H3 tail peptide [20] highlighted as boxes. B) Highlighting of H2A and H3 tails on the crystal structure, with the different positions of the H2A N-terminal is emphasized. Underlined residues in the sequence of the *Dm.* H2A and H3 tails are observed in the sediment.

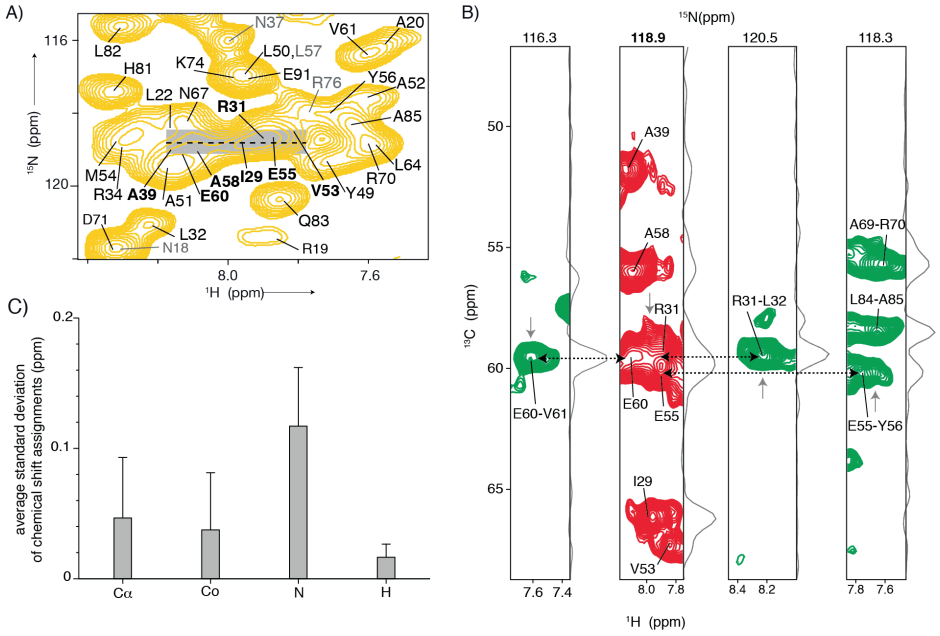


Figure S3. Assignment of resonances in the crowded region of the 2D. A) Zoom on the crowded region from the 2D NH correlation spectrum of H2A-labeled sedimented nucleosomes. Dashed line and gray box indicate the position and width of the strip @ 118.9 ppm displayed in panel B, highlighting the assignment and resolution of overlap for the resonances in bold. B) Strip from the 3D CANH spectrum (red) highlighting the resolved Cα cross peaks observed for I29, A38, V53 and A58 in the crowded region of panel A. The sequential connectivities on which the assignment of assignment of R31, E55 and E60 are shown in the strips from the CA(CO)NH (green). 1D slices on the ^{13}C dimension are displayed next to each strip, taken at the position of the gray arrows. C) Average standard deviation of the assigned ^{13}Cα, ^{13}Co, ^{15}N, and ^{1}H chemical shifts across the five spectra used for assignment, with error-sticks indicating the standard deviation over all assignments.

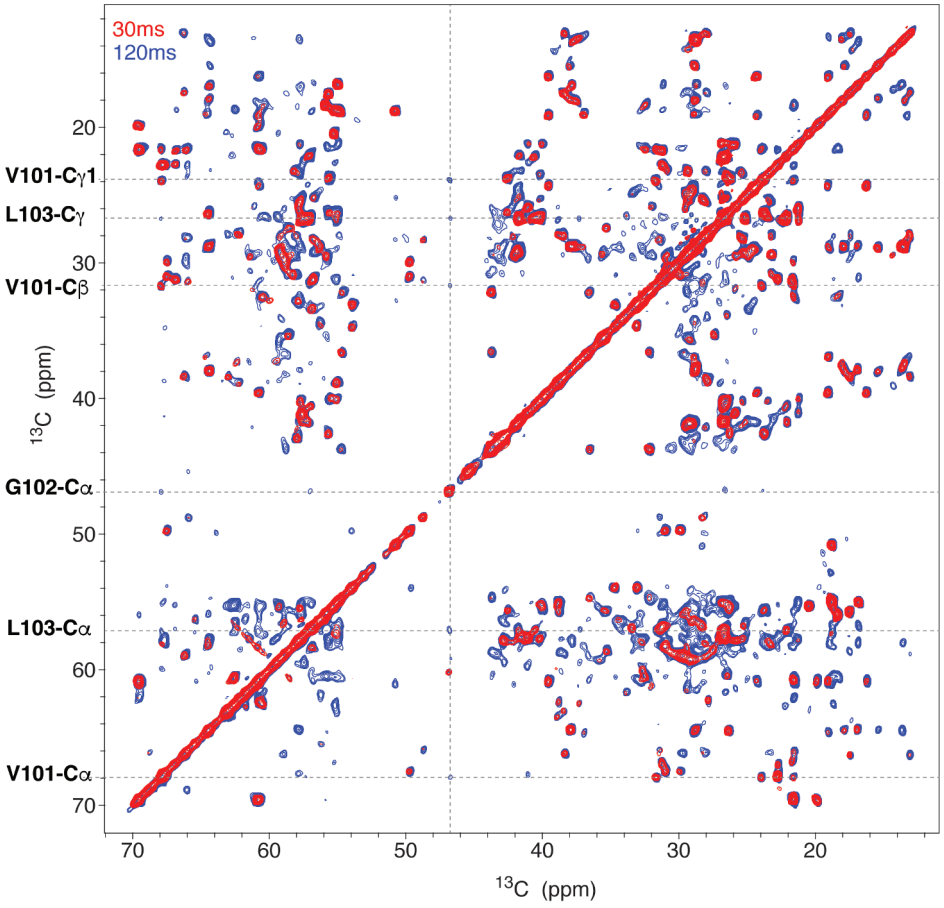


Figure S4. ^{13}C-detected NMR spectra of H3 in sedimented mononucleosomes.

Overlay of ^{13}C-detected 2D PARIS cross-correlation spectra measured with 30 (red) and 120 (blue) ms mixing time. The 30ms mixing time displays short range correlations – mostly intra-residue – while 120ms mixing time additionally reveals inter-residue correlations. Assignments of cross-correlations between G102, V101 and L103 are highlighted as example.

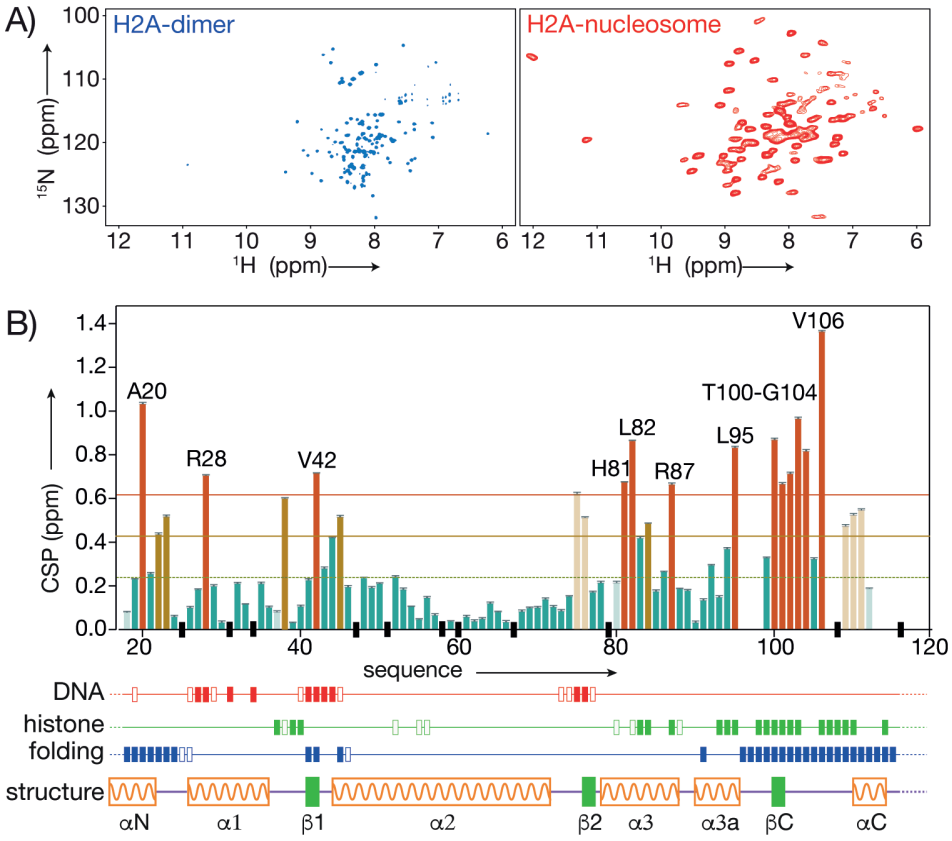


Figure S5. Chemical shift changes for H2A between dimer and nucleosome forms. A) 2D NH HSQC solution-state spectra of H2A in the H2A/H2B dimer (left, blue) and dipolar-based 2D NH ssNMR spectrum of H2A in sedimented nucleosomes (red). B) Chemical shift perturbations (CSPs) between the two spectra from panel A. Residues with CSPs deviating more than 1 (orange) or 2 (red) standard deviations from the 10% trimmed mean (dashed line) are labeled. Data for tentatively assigned residues shown in lighter colors. Black bars indicate prolines or missing data due to peak overlap. Location of changes in the chemical environment due to contacts to DNA or other histones, or due to changes in secondary structure are indicated below the plot.

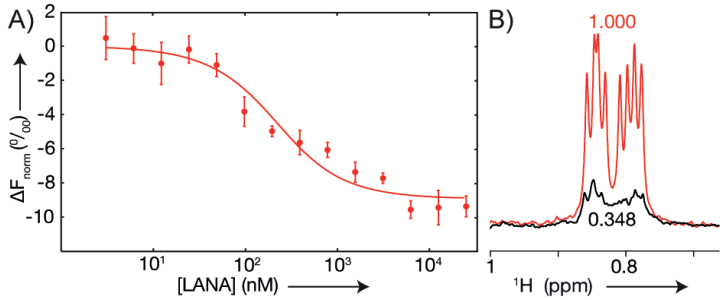


Figure S6: Nucleosome binding and co-sedimentation of the LANA peptide.

A) Microscale thermophoresis-derived binding curve of LANA$^{2\text{-}22}$ to fluorescently labelled nucleosomes. Error bars are the standard deviations from three independent titrations. The best-fit dissociation constant (K_D) is 160 nM. B) Overlay of the LANA L8 and L18 methyl peaks in 1D ^{1}H spectra of the LANA-nucleosome co-sedimentation supernatant (black) and of pure LANA at the same concentration as in the sedimentation mix (red). Since UV measurements indicate 6% of nucleosomes remained in the supernatant, the large decrease in peak volume suggests that LANA was largely co-sedimented with the nucleosome. Changes in peak patterns and line widths for LANA in the supernatant indicate that the remaining free and nucleosome-bound peptide are in fast-to-intermediate exchange, making it difficult to assess the precise molar ratio between LANA and the nucleosome in the sediment.

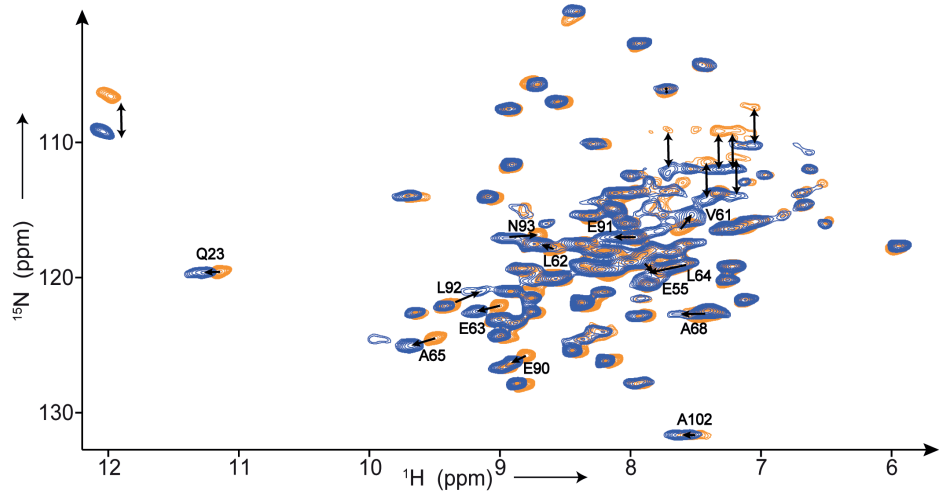


Figure S7: **LANA-binding induces specific chemical shift changes in H2A-labeled nucleosomes.** Overlay of the 2D NH spectra of H2A in free (yellow) and LANA-bound (blue) nucleosomes. Vertical two-way arrows indicate side chain resonances that are aliased to different ^{15}N chemical shifts due to different carrier positions.

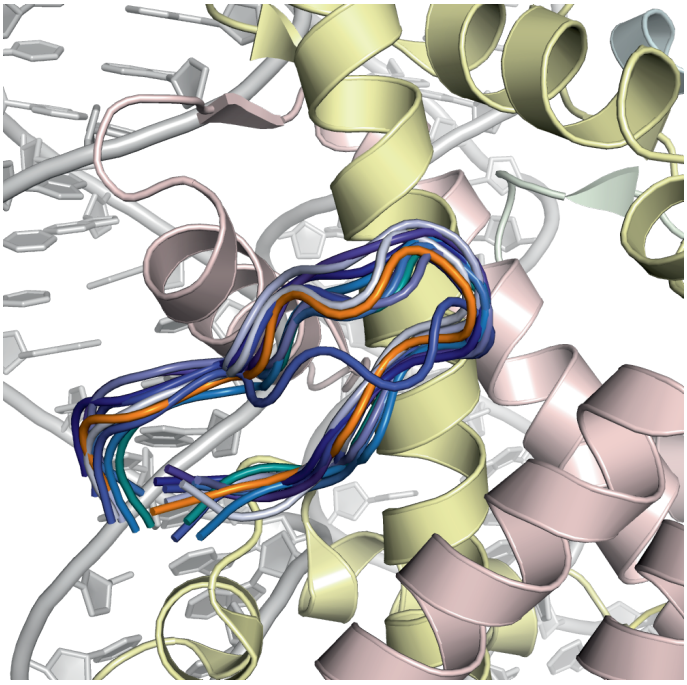


Figure S8: Top-10 structures of the LANA-nucleosome complex obtained via ssNMR- and mutagenesis-data-driven docking. Cartoon display of the 10 lowest-energy structures – aligned on 2PYO, colored in shades of blue, superimposed on the crystal structure 1ZLA in orange.

References

1 Bertani, G. STUDIES ON LYSOGENESIS I.: The Mode of Phage Liberation by Lysogenic Escherichia coli. *Journal of Bacteriology* **62**, 293-300 (1951).

2 Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **375**, 23-44, doi:10.1016/S0076-6879(03)75002-2 (2004).

3 Klinker, H., Haas, C., Harrer, N., Becker, P. B. & Mueller-Planitz, F. Rapid purification of recombinant histones. *PLoS One* **9**, e104029, doi:10.1371/journal.pone.0104029 (2014).

4 Gardiennet, C. *et al.* A sedimented sample of a 59 kDa dodecameric helicase yields high-resolution solid-state NMR spectra. *Angew Chem Int Ed Engl* **51**, 7855-7858, doi:10.1002/anie.201200779 (2012).

5 Bertini, I. *et al.* On the use of ultracentrifugal devices for sedimented solute NMR. *J Biomol NMR* **54**, 123-127, doi:10.1007/s10858-012-9657-y (2012).

6 Corbeski, I. *et al.* Microscale Thermophoresis Analysis of Chromatin Interactions. *Methods Mol Biol* **1837**, 177-197, doi:10.1007/978-1-4939-8675-0_11 (2018).

7 Fung, B. M., Khitrin, A. K. & Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids. *J Magn Reson* **142**, 97-101, doi:10.1006/jmre.1999.1896 (2000).

8 Chevelkov, V., Rehbein, K., Diehl, A. & Reif, B. Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew Chem Int Ed Engl* **45**, 3878-3881, doi:10.1002/anie.200600328 (2006).

9 Shaka, A. J., Keeler, J., Frenkiel, T. & Freeman, R. An improved sequence for broadband decoupling: WALTZ-16. *Journal of Magnetic Resonance (1969)* **52**, 335-338, doi:10.1016/0022-2364(83)90207-x (1983).

10 Weingarth, M., Trebosc, J., Amoureux, J. P., Bodenhausen, G. & Tekely, P. Efficiency at high spinning frequencies of heteronuclear decoupling methods designed to quench rotary resonance. *Solid State Nucl Magn Reson* **40**, 21-26, doi:10.1016/j.ssnmr.2011.03.004 (2011).

11 Zhou, D. H. *et al.* Solid-state NMR analysis of membrane proteins and protein aggregates by proton detected spectroscopy. *J Biomol NMR* **54**, 291-305, doi:10.1007/s10858-012-9672-z (2012).

12 Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR* **52**, 315-327, doi:10.1007/s10858-012-9611-z (2012).

13 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, doi:10.1007/bf00197809 (1995).

14 Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327, doi:10.1093/bioinformatics/btu830 (2015).

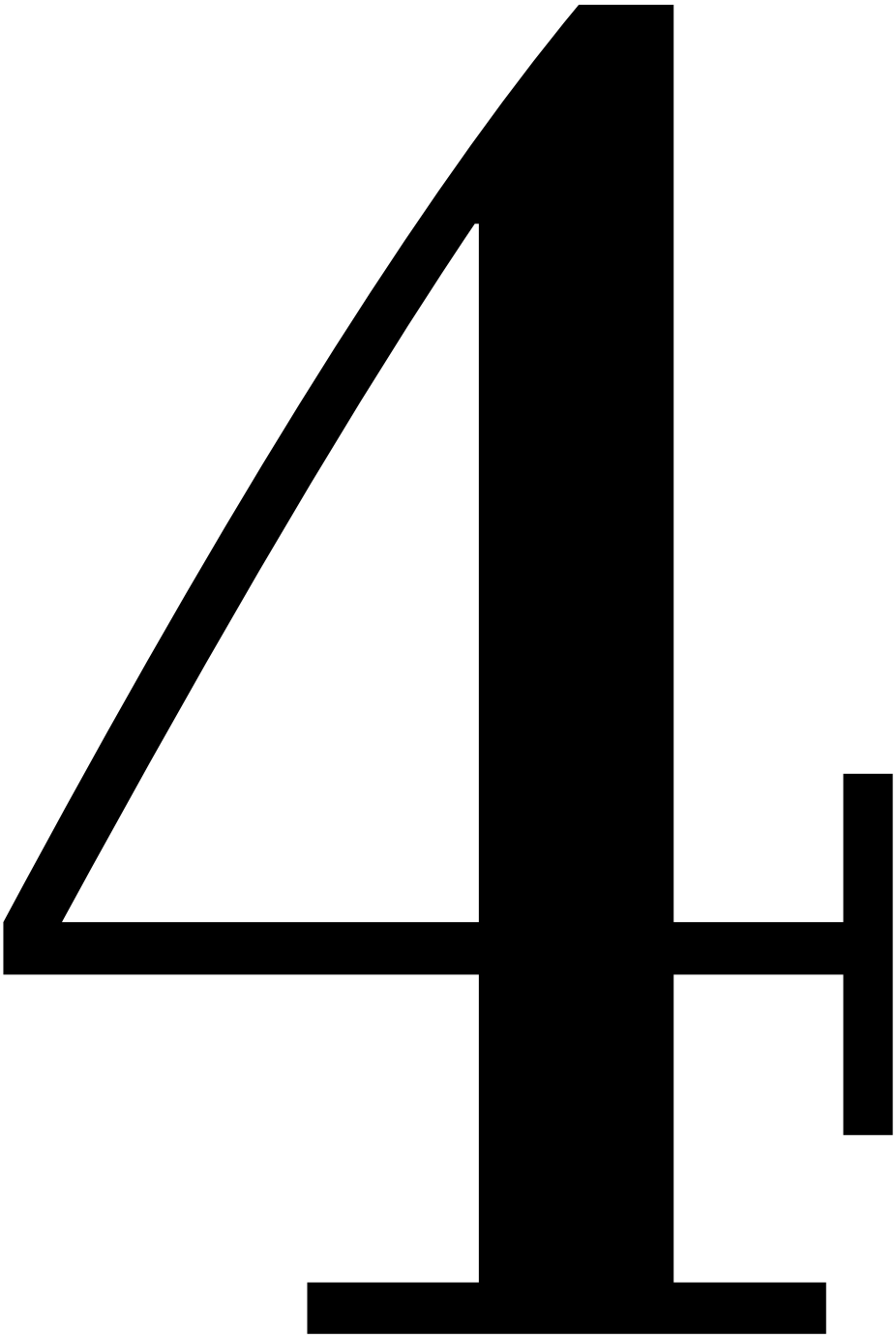
15 The PyMOL Molecular Graphics System, Version 2.0 Schr.dinger, LLC.

16 Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**, 856-861, doi:10.1126/science.1120541 (2006).

17 Clapier, C. R. *et al.* Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **71**, 1-7, doi:10.1002/prot.21720 (2008).

18 van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725, doi:10.1016/j.jmb.2015.09.014 (2016).

19 Zhou, B. R. *et al.* Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J Mol Biol* **421**, 30-37, doi:10.1016/j.jmb.2012.04.032 (2012).
20 Stutzer, A. *et al.* 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, doi:10.1016/j.molcel.2015.12.015 (2016).



Chapter 4

Characterization of the multifactorial Dot1p-nucleosome interaction

Contributions:
Dot1pD protein was produced by dr. Patrick Celie (Netherlands Cancer Institute, Amsterdam). Solid-state NMR experiments were performed by dr. ShengQi Xiang. A 15N-labelled ubiquitin sample was kindly provided by Siddarth Narasimhan.

Published as:
------ *in preparation* ------

Abstract

Post-translational methylation of lysine residues in nucleosomes is the result of a highly controlled interaction of a methyltransferase with the nucleosome in order to obtain the specific modification of a single target-residue. Recent work has shown that mammalian Dot1, Dot1L, responsible for methylation of lysine 79 in histone H3, relies on a network of interactions to impose a specific and active binding mode. Next to binding to the target lysine, these include interactions to ubiquitin attached to H2B, nucleosomal DNA, the H2A/H2B acidic patch and the H4 tail. Sequence and functional *in vivo* analyses suggest that yeast Dot1, Dot1p, relies on a significantly different binding mode. Here we characterized the Dot1-binding epitope within the nucleosome by NMR and detailed the nucleosome-Dot1p interaction. Our results indicate that the Dot1p core contains a minor ubiquitin-interaction site similar to that in mammals, but with very low affinity for ubiquitin, which may explain the reduced ubiquitin dependence of yeast Dot1. We further find that a lysine-rich region flanking the catalytic core has nanomolar affinity for nucleosomal DNA, while a longer construct binds with micromolar affinity, suggesting the presence of an auto-inhibition mechanism. Finally, our NMR analysis indicates the H3K79 loop has increased dynamics which may be crucial to allow entry of the target lysine into the binding cleft in Dot1. Collectively, our results provide new insights into the multi-factorial nucleosome interaction for yeast Dot1.

Introduction

Post-translational histone modifications are a key mechanism to regulate chromatin biology. This mechanism relies on specific interaction of proteins with nucleosomes to install, bind or remove these modifications[1]. While structural studies have revealed a great wealth of insight into the molecular mechanism for histone tail modifications, there is scarce data on how writer, reader and eraser proteins for modifications of the histone octamer core interact with the nucleosome surface. One of the most enigmatic core modifications is the methylation of K79 in H3 (H3K79me) by Dot1, a highly conserved enzyme[2,3]. While its precise molecular role is yet poorly understood, Dot1 is involved in essential nuclear processes such as regulation of gene expression and the DNA damage response[4,5,6]. Dot1 may also play a role in cellular ageing through accumulation of H3K79me, for which no demethylase has been identified[7]. In addition, H3K79me may serve as a memory marker of former nucleosome ubiquitination on H2B, a transient modification that is thought to open up chromatin[8,9]. Crucially, misregulation of human Dot1 activity can lead to cancer, most notably through MLL-based leukemia [10]. Thus, there is great interest to better understand how Dot1 engages with nucleosomes and specifically modifies H3K79 within a background of many accessible lysines on the nucleosome surface[11].

The yeast *Saccharomyces cerevisiae* has been used extensively as a model organism to study the molecular basis of H3K79me, resulting in the first identification of Dot1, Dot1p in yeast, as the enzyme responsible for mono-, di- and tri-methylation of H3K79[2]. Dot1p and the mammal homologue Dot1-like, Dot1L, feature a catalytic core flanked either on the N- or C-terminal end by long sequences without known folded domains (Figure 1A). These extensions contribute additional DNA binding motifs and, in the case of Dot1p, also to ubiquitin binding. A Lys-rich DNA-binding region is positioned C-terminal to the catalytic core in Dot1L, whereas it is in the N-terminal extension in Dot1p[12,13].

The crystal structures of the catalytic core revealed it to be a bipartite domain of a N-terminal helical bundle packed against the methyltransferase domain. As exception to

the family of lysine methyltransferases, the Dot1 catalytic domain does not belong to the SET-superfamily but rather resembles class one arginine methyltransferases[2,13,12]. The Dot1L and Dot1p cores have highly similar structures with backbone RMSD of 2.14 Å between the two proteins[12,13] (Figure 1B, C). The catalytic core relies on the co-factor SAM (S-adenosyl methionine) to donate a methyl group to the target lysine. Notably, the co-factor is partly buried in the catalytic core, requiring the target lysine to enter a hydrophobic channel to reach the acidic catalytic site containing the methyl donor[13] (Figure 1B). This explains the poor processivity observed for Dot1p, which releases the target lysine while replacing its co-factor in order to incrementally increase the methylation state[8,13]. The requirement for a highly accessible lysine starkly contrasts with the only partly exposed nature of H3K79 in the crystal structure of the nucleosome, where the lysine sidechain is positioned flat on the surface, covering a small hydrophobic pocket[14] (Figure 1D).

Ubiquitination of H2B on K123 (H2BK123Ub) on the nucleosome has been shown to stimulate activity of both Dot1p and Dot1L. For Dot1L the presence of H2BK123Ub is essential for enzymatic activity, while Dot1p can mono-methylate nucleosomes in absence of ubiquination in vivo[15,7,16,17]. Sequence analysis of Dot1p suggested the presence of a ubiquitin binding motif between residues 40 and 100[13], while deletion and pulldown studies showed that the residues 100 to 140 are critical for ubiquitin binding and removal thereof perturbs DNA- and nucleosome-binding properties of Dot1p[18] (Figure 1A).

During the course of this study, five independent studies reported high-resolution cryo-EM structures of the mammalian Dot1L-nucleosome complex, offering tremendous insights into the catalytic mechanism and the complex interaction at atomic detail[17,19-22] (Figure 1C,D). These studies all made use of H2BK123 ubiquitinated nucleosomes and showed that the C-terminal helix in the catalytic core of Dot1L forms a "hydrophobic cradle" together with a neighboring loop and binds directly to ubiquitin. Both helix and loop are conserved throughout metazoans, whereas the helix is partly present in the yeast enzyme. In addition, Dot1L is anchored

to the H2A/H2B acidic patch with two arginines close to the cradle (Figure 1C). Similarly, these arginines are only conserved in metazoans and are missing in the yeast Dot1p[21].

The interactions to ubiquitin and the acidic patch help to anchor and subsequently orient Dot1L, thereby promoting a conformation in which the active site is directed towards H3K79. Notably, all aforementioned publications report a "poised state" where Dot1L does not directly contact H3K79. In this state, the active site is distant from the target lysine and the hydrophobic channel remains unoccupied. In one of the cryo-EM studies, H3K79 was found to be mono methylated, further suggesting that the poised state may reflect a pre- or post-methylation conformation[20].

In their groundbreaking work, Worden et al.[21] also determined the structure of the active state of the Dot1L-nucleosome complex in which the H3K79 side chain is engaged. Their approach relied on the use of a non-modifiable lysine analogue, norleucine (Nle), in which the central nitrogen of the side-chain amino group is replaced by a carbon atom. In the active state the enzyme remains anchored to ubiquitin and the acidic patch as in the poised state, but additionally directly contacts the surface around the target lysine through two conserved hydrophobic loops (Figure 1D). The Nle79 side chain sits in the hydrophobic channel and reaches the active site of Dot1L. Additionally, in the active state, R17 from the H4 tail directly interacts with a conserved acidic pocket in Dot1L while H4R19 stabilizes the open conformation of the H3K79 loop (Figure 1D). This rationalizes previous findings that the H4 tail is required for full activity[18,4,16]. The DNA-binding Lys-rich C-terminal sequence of Dot1L is not resolved in either of the aforementioned structures, highlighting its dynamics or low specificity of binding[17]. Yet, deletion of this region results in a 60 to 200-fold increase for the K_m without affecting the k_{cat}[21].

These structural data have highlighted the complex and multifactorial nature of the Dot1L-nucleosome interaction and rationalized the dependence on H2B ubiqutination as anchor to orient the enzyme towards the target lysine. Given that there is no absolute requirement for nucleosome ubiquitination for Dot1p activity, the question arises as to what determines formation of the active complex for the yeast

protein. Together with the absence of the acidic patch-binding Arginine-anchor residues and the different relative position of the Lys-rich DNA-binding sequence, this suggest that Dot1p utilizes a significantly different nucleosomal binding mode than Dot1L while retaining the same catalytic mechanism.

Here, we aimed to characterize the molecular basis of the interaction of yeast Dot1p with the nucleosome using an NMR and biochemistry driven approach. Exploiting our recent sedimentation-driven solid-sate NMR approach to study nucleosome-protein interactions,[23] we here explore the interaction of a catalytically inactive Dot1p mutant with H3. Our NMR data strongly suggests that H3K79 in free nucleosomes has increased intrinsic dynamics on a microsecond time scale, which may be essential to allow for entry in the Dot1p substrate channel. We further find a strong perturbation of the H3 tail upon binding of Dot1p, possibly due to a direct interaction with the Dot1p core. We further show that the Lys-rich N-terminal region itself binds with low nanomolar affinity to both DNA and nucleosome, while Dot1p has significantly reduced affinity, suggesting the presence of an auto-inhibition mechanism. We further show that, in nucleosome-free context, ubiquitin has low affinity for Dot1p and may interact with the Dot1p core in a similar binding mode as Dot1L. Collectively, our results provide new insights into the nucleosome recognition by Dot1p and may better contextualize its use for human cancer research.

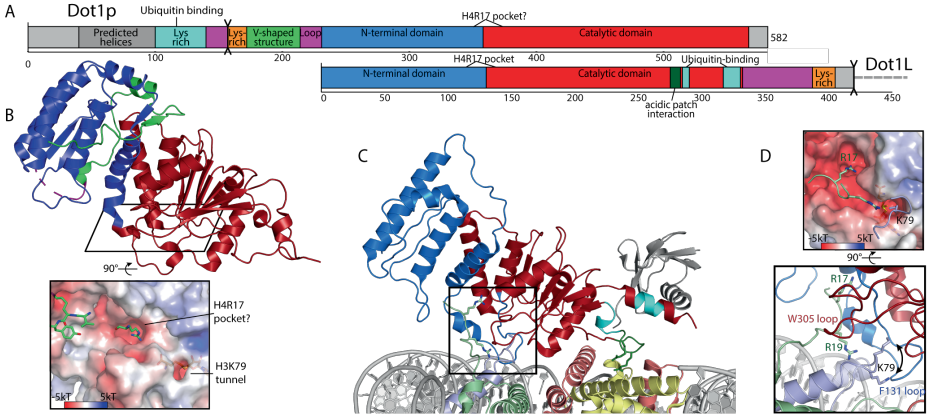


Figure 1. Architecture, structure and nucleosome interaction of Dot1. (A) Domain architecture of Dot1p and Dot1L. Flexible loops are colored in pink. Double arrows delimitate the constructs used in crystallographic, cryo-EM and the present studies. (B) Top panel: crystal structure of Dot1p (PDBid: 1U2Z [13]) with domains colored as in (A). Bottom panel: surface plot with color coding according to electrostatic potential of highlighted region in top panel; co-crystallized SAH (orange) and core-binding terminal residues of Dot1p (green) are shown as sticks. (C) Cryo-EM structure of active complex of Dot1L with ubiquitinated nucleosomes. Dot1L is colored as in (A), ubiquitin in dark grey, histones in light colors corresponding to the canonical coloring (PDBid : 6NJ9[21]). (D) Zoom on the active complex as indicated in (C). Top panel: electrostatic surface potential with interaction of H4 tail (green) and H3 Nle 79 (blue). Bottom panel : overlay of poised and active state Dot1L-H3K79 interaction with Dot1L colored as in A, showing the conformation change of the H3K79 loop upon formation of the active complex (PDBid : 6NOG[21]). Electrostatic potentials were calculated by the APBS program[24].

Results

Design of Dot1p constructs

We made use of two Dot1p constructs, one corresponding to the DNA-binding lysine-rich region (residues 157-178) and one corresponding to the catalytic core including the lysine rich region (158-582). The lysine-rich region was ordered as a peptide and will be denoted as DKR-peptide for Dot1p K-rich peptide. An inactive mutant G401R, previously shown to be unable to bind the substrate SAM,[8,2] was used to prevent methylation activity, as wild-type Dot1p co-purifies with SAM[13,12,20]. Removal of the first 157 residues was previously reported to yield efficient expression in *E. coli* and to result in similar activity and identical nucleosome binding behavior as the full-length protein[13]. This construct shall be referred to as Dot1pΔ further in the text.

The lysine-rich region binds DNA with nanomolar affinity

We first set out to determine the preferred binding substrate of the lysine-rich region using a fluoresceine-labeled DKR-peptide in fluorescence anisotropy-based binding assays. The peptide was titrated with histone H2A-H2B dimers, H3-H4 tetramers, nucleosomal DNA and nucleosomes. No significant binding could be observed with the dimer and tetramer, while a small increase in anisotropy was observed upon addition of DNA and a clear binding curve was obtained for titration with nucleosomes (Figure 2A,B). Fitting of this binding curve assuming a single binding site model resulted in a K_D of 27 nM for the nucleosome interaction (χ_{red} 1.1). To verify whether the small increase in anisotropy upon DNA addition was due to binding, a competition assay was performed in which DKR-peptide was first incubated with nucleosomes and subsequently free DNA was added. A clear reduction of anisotropy is observed (Figure 2C), and from the estimated inflection point a dissociation constant between 4 and 8 nM can be estimated[25]. We note that these are apparent affinities, subject to an avidity effect from the multitude of binding sites on DNA and nucleosome. Nevertheless, the high affinity for both DNA and

nucleosomes suggest that the Lys-rich region is the main driver of affinity of Dot1p for the nucleosome, effectively anchoring Dot1p to the nucleosome. It further suggests that inclusion of the lysine-rich region is required to obtain the high-affinity binding-mode required for successful co-sedimentation of Dot1p and the nucleosome.

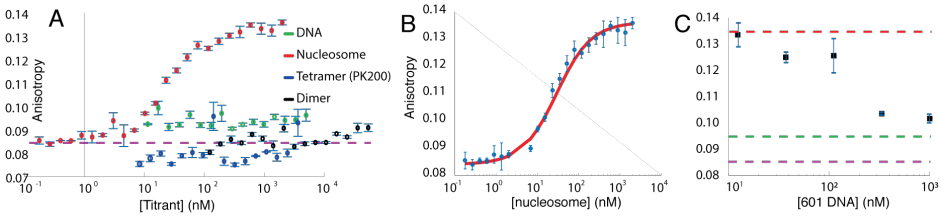


Figure 2. The DKR-peptide, corresponding to the Lys-rich region of Dot1p, binds with low-nanomolar affinity to DNA. Magenta dashed line represents the average anisotropy for free peptide. (A) Fluorescence anisotropy (FA) assay investigating the DKR-peptide binding to the different components of the nucleosome. (B) FA-derived binding curve for binding to nucleosome. Best-fit to a single binding site model yields a Kd of 27 nM (χ_{red} = 1.1). (C) Competition assay between DKR peptide binding to nucleosome and 601-DNA. Nucleosome-bound and DNA-bound average anisotropies are displayed as red and green dashed lines, respectively. All data points were measured in triplicate and error bars indicate one standard deviation.

The catalytic core construct is well-folded and binds tightly and specifically to nucleosomes

Since our Dot1pΔ construct is a truncated, inactive mutant protein, we first assessed folding of the protein by NMR. The 1D ^{1}H spectrum confirmed the folded nature of the protein, particularly through the observation of distinct upfield-shifted methyl signals characteristic of a hydrophobic core (Figure 3A). In addition, a thermal shift assay (TSA)[26] was conducted to assess the temperature stability of Dot1pΔ. Melting temperatures in the range of 48 to 51 °C were observed across a wide range of conditions, including the buffer suitable for our nucleosome NMR studies (Figure 3B).

To test and optimize binding conditions of this construct to unmodified nucleosomes, we performed electrophoretic mobility shift assays (EMSA) with reconstituted nucleosomes across multiple buffer conditions. Since the lysine-rich region in the construct likely binds nucleosomal DNA in a non-specific manner, we carefully assessed the results of the assay in order to minimize band smearing and to promote formation of distinct well-defined shifted bands, which typically correspond to formation of a specific complex. In low ionic strength conditions (I = 25 to 65 mM), well-defined singly and doubly shifted bands could be observed indicating the formation of a 1:1 and 2:1 Dot1p-nucleosome complex (Figure 3C). Specific complex formation is observed at pH 6.5 through 8.0 and seems insensitive to addition of $MgCl_2$ or identity of buffer species (Tris or PO_4) (Figure 3C). At higher excess of Dot1pΔ, the interaction seems dominated by non-specific binding. Across conditions, the optimal Dot1pΔ:nucleosome molar ratio appears to be 3:1, for a 0.5 μM nucleosome concentration. Additionally, even at a 5-molar excess of Dot1pΔ to nucleosome, the free form of the nucleosome can still be observed, suggesting a micromolar affinity of Dot1pΔ to the nucleosome. Together these results demonstrate that Dot1pΔ can bind nucleosomes in a specific manner and that this interaction is distinct from that of the DKR-peptide alone.

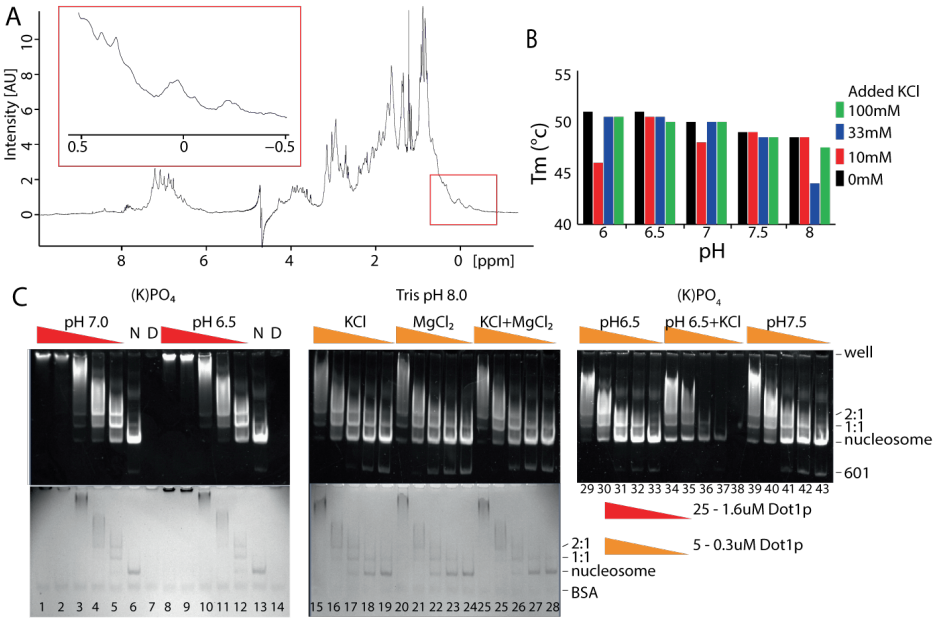


Figure 3. The Dot1pΔ construct is folded, stable in our conditions and can form a specific complex with the nucleosome. (A) ^{1}H 1D spectrum of Dot1pΔ. The insert highlights upfield shifted methyl signals characteristic of a hydrophobic core. (B) Melting temperatures for Dot1pΔ in 15 mM KPi buffer with different pH and added KCl concentrations as determined from TSA assays. (C) EMSA in native PAGE of nucleosome titrated with 2-fold serial dilutions of Dot1pΔ; reaction mixtures were incubated for 20 minutes in 15 mM KPi and 10mM KCl (I=33/41/49 mM at pH 6.5/7.0/7.5, respectively) or 25 mM Tris pH 8 (I=14 mM) with indicated supplemental KCl or $MgCl_2$ (50 and 2.5 mM, respectively). "N" and "D" lanes contain controls with pure nucleosome and pure Dot1pΔ, respectively. Gels were casted and run in 0.2x TBE. GelRed® (top) and Coomassie brilliant blue (bottom) successive stainings of the gels.

The H3K79 site is dynamic in the nucleosome

We next reconstituted nucleosomes containing isotope-labeled H3 in order to characterize the H3K79 binding site of Dot1p. These preparations were ultracentrifuged to create a dense sediment to be interrogated by ssNMR. Previously, we showed that H3 gives rise to high quality dipolar- and INEPT-based spectra,

probing rigid and flexible parts, respectively[23]. In particular, a dispersed peak pattern with overall narrow line widths is observed in the dipolar, cross-polarization (CP)-based spectrum that probes the rigid parts of the histone core, reflecting correct folding of H3, and by extension, of the nucleosome. The INEPT-based spectrum of the H3 N-terminal tail was previously assigned through comparison of the peak pattern with solution-state spectra[27]. Some additional INEPT signals could be confidently linked to signals appearing in the dipolar spectrum, based on matching peak positions (see below in Figure 5A, C). Here, we assigned nearly all backbone signals detected in the CP-based 2D ^{15}N-^{1}H spectrum through the combined use of proton-detected 3D (H)CANH/(H)CA(CO)NH, (HNCA)CBCA(CA)NH, (H)HNH and 4D (H)CACONH/(H)COCANH experiments (Supplementary Table 1). In total 85% (93% including tentative assignments) of the core residues (residues 40-135) were assigned (Figure 4A). The assigned backbone and sidechain chemical shifts were used to predict the secondary structure propensities using TALOS-N[28], confirming the canonical α1-α2-α3 histone fold of the core of H3 (Figure 4B). The additional N-terminal helix in the nucleosome core, αN, is also observed. Overall, the NMR data correspond well to the crystal structure, as was also observed for H2A[23] and H4[29]. Some deviation in the backbone structure may be present in the loop between αN and α1 that is involved in DNA-binding. While the predicted extended conformation matches the elongated nature of this loop in the crystal structure, the predicted backbone dihedral angles for residues 60 to 62 systematically deviate from that observed in the crystal structure (Figure 4E).

Akin to our observation with H2A[23], the backbone NH of resonances of several residues close to the DNA are either split or broadened (Figure 4C,D). G44, and S86 are of particular interest as these amino acids bind the minor groove of the DNA near superhelix locations (SHL) +/–0.5 for G44 and +/–2.5 for S86. Because of the non-palindromic nature of the 601 DNA[30,31], these residues face different DNA sequences in the minor grooves on the left and right halves of the nucleosome, each with slightly altered structure, rigidity and thus slightly changed histone-DNA interactions[32] .

Indeed, S86 faces a CpGpA trinucleotide at SHL +2.5 and TpTpC at –2.5 (Figure 4C). The splitting of G44 can be similarly correlated to one side of the nucleosome presenting a CpGpC and the other CpGpT to this residue (Figure 4D).

Strikingly, the K79 loop (res. 77 to 82) could merely be partly and tentatively assigned. This portion of the sequence showed an overall low- to no- signal intensity in the CP-based spectra and in particular no signal in any of the 4D CP-spectra (Figure 4B). This indicates that this loop may have increased flexibility or heterogeneity, which may be also somewhat reflected in the in the crystal structure by the elevated temperature factors observed for this region[11,12]. Such increased flexibility may be crucial to present the K79 target side chain to Dot1.

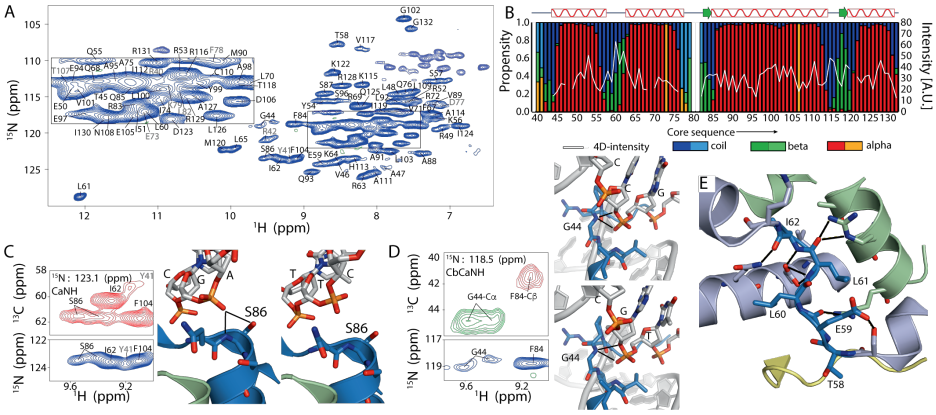


Figure 4. Assignment of the H3 folded core in mononucleosomes. (A) Dipolar 2D NH of H3 in mononucleosomes with assignments. Tentative assignments are displayed in grey. (B) Backbone-chemical shift based secondary structure of H3 as determined by TALOS-N. Tentative assignments are displayed in lighter colors. The signal intensity in the 4D COCANH is plotted in white. (C, D) Peak doubling of S86 (C) and G44 (D) observed in the 2D NH (blue) and 3D CaNH or CbCaNH (red) spectra correlates to the asymmetric environment in 601-nucleosomes (PDBid 3LZ0 [31]). (E) The extended conformation of loop 1 is stabilized by a strong hydrogen binding network involving the sidechains of α2-residues in the crystal. Black lines in C, D, E indicate hydrogen-bonds predicted by Pymol.

Co-sedimentation of Dot1pΔ reveals a strong impact on H3 N-terminal tail

Having established that H3 is well-folded in our sedimented nucleosomes and may present a flexible binding epitope, we next co-sedimented Dot1pΔ with H3-labeled, non-ubiquitinated nucleosomes. Native and SDS-PAGE of the various fractions from the co-sedimentation suggest that, similarly to the EMSA experiments, a mixture of specifically and non-specifically bound Dot1pΔ-nucleosome complex was obtained in the rotor (Figure 5E).

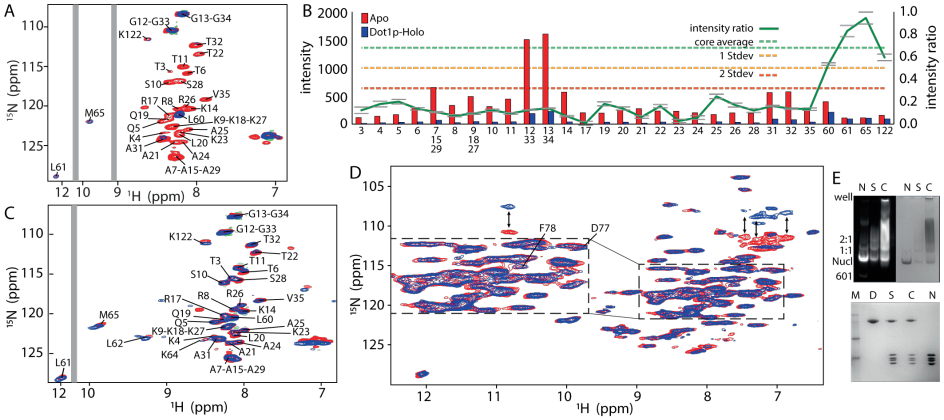


Figure 5. Dot1pΔ binds the nucleosome in a poised-like state and its binding involves the H3 tail. (A) Overlay of the INEPT-based spectra of H3 in free (red) and Dot1pΔ-co-sedimented nucleosomes (blue), measured at 50 kHz MAS. (B) Changes of signal intensities in the 50 kHz INEPT spectrum upon Dot1pΔ co-sedimentation, based on the tentative assignment of the bound state signals. Bar plot of intensities and scatter plot of intensity ratio with error bars indicating absolute error. (C) Overlay of the J-based spectra of H3 in free (red) and Dot1pΔ-cosedimented samples (blue), measured at 60 kHz MAS. (D) Overlay of the CP-based spectra of H3 in free (red) and Dot1pΔ-co-sedimented nucleosomes (blue), measured at 50 kHz MAS. (E) GelRed- and subsequently Coomassie-stained native PAGE and Coomassie-stained SDS-PAGE analysis of the Dot1pΔ-nucleosome co-sedimentation. Lanes show nucleosome (N), Dot1pΔ (D), supernatant after co-sedimentation (S) and the resuspended top part of the sediment in the rotor (C).

Dipolar-based and J-based ^{1}H-detected 2D NH correlation spectra were recorded at

50 kHz MAS on the co sedimented Dot1pΔ/nucleosome sample, followed by the same measurements at 60 kHz MAS. The dipolar-based spectra of the H3 core were again of high quality. Assignments were transferred from the free state at 60 kHz with the aid by a 3D CANH to resolve overlap in the 2D. Strikingly, no significant chemical shift or intensity differences could be seen in the dipolar spectra (Figure 5D). Based on the recent cryo-EM structures, the intermolecular interface on H3 can be expected to include residues 73 to 82 for the active state of the Dot1p-nucleosome complex, while in the poised state no direct contact would be made with H3. The observable resonances in the K79 binding loop exhibit low peak intensities in the co-sediment as in the apo state without significant chemical shift changes, indicating Dot1pΔ does not contact the target lysine. Since the gel-analysis unequivocally demonstrates the presence of Dot1pΔ-nucleosome complexes in the sample, we conclude Dot1p is either bound in the poised state or bound non-specifically.

Surprisingly, analysis of the J-based spectrum shows a dramatic intensity drop for all of the H3-tail resonances in the co-sediment (Figure 5A, B). A nearly uniform 80% decrease in intensity for the tail resonances is observed, while the few signals arising from the core remained. This strongly suggests that nucleosome binding of Dot1pΔ causes a large-scale change of the H3 tail conformational state. This can be due to a direct interaction or could be indirect, mediated by DNA binding. Since the H3 tail is known to be transiently bound to the nucleosomal DNA[33], a competitive effect from the Dot1p-DNA interaction could induce a subsequent change in H3. In either case, the intensity drop corroborates that Dopt1pΔ is bound to the nucleosome within the sediment. Notably, after increase of the MAS spinning speed to 60 kHz, the intensities and peak pattern of the J-based spectrum of the co-sediment resembled strongly the apo spectrum, suggesting that the increased sample temperature resulted in Dot1pΔ denaturation (Figure 5C). We thus conclude that while Dot1pΔ is bound to nucleosome in the sediment and that Dot1pΔ binding induces a conformational change in the H3-tail, the active state of the complex is at best only marginally populated in the sediment.

Dot1pΔ weakly interacts with ubiquitin

We next assessed whether the Dot1p core can interact in similar manner as Dot1L with ubiquitin. Even if non-essential for activity, nucleosome ubiquitination stimulates Dot1p activity[15] and increases affinity two-fold in the case of Dot1pΔ[17]. We thus titrated isotope-labeled ubiquitin with Dot1pΔ, without the N-terminal ubiquitin-binding element, and probed their interaction by NMR. Addition of the 51 kDa Dot1p resulted in minimal chemical shift or intensity changes for most residues with one notable exception (Figure 6A). A dramatic reduction in peak intensity effect together with significant chemical shift perturbation was observed for the resonance of L71, indicative of exchange-induced line broadening and thus pointing to a weak Dot1pΔ-Ub interaction (Figure 6A, C). The absence of significant overall peak intensity changes, despite formation of a large 59 kDa complex, suggests that at most 5 to 10% of bound state is formed considering the expected relaxation rates in free and Dot1pΔ-bound ubiquitin (Ub). This in turn indicates that the binding affinity is in the milli-molar range, with a K_D around 2 mM. Since such low affinity interactions typically occur in the fast-exchange regime in which also the relaxation rates of free and bound states are averaged, this also suggest that the Ub-Dot1p complex is not rigid, but flexibly linked. Together, these data strongly suggest that there is a specific but weak interaction between the Dot1p core and Ub. Notably, L71 is also part of the ubiquitin interface for Dot1L. As we cannot exclude that other mildly affected residues are also involved in the specific interaction, this signifies that the Dot1p interaction at least partly coincides with the hydrophobic patch that binds Dot1L (Figure 6B). We thus conclude that Dot1pΔ contains a ubiquitin-binding motif that can mediate a weak specific interaction with Ub. The low millimolar intrinsic affinity of Dot1pΔ for Ub may explain its reduced dependency on ubiquitination.

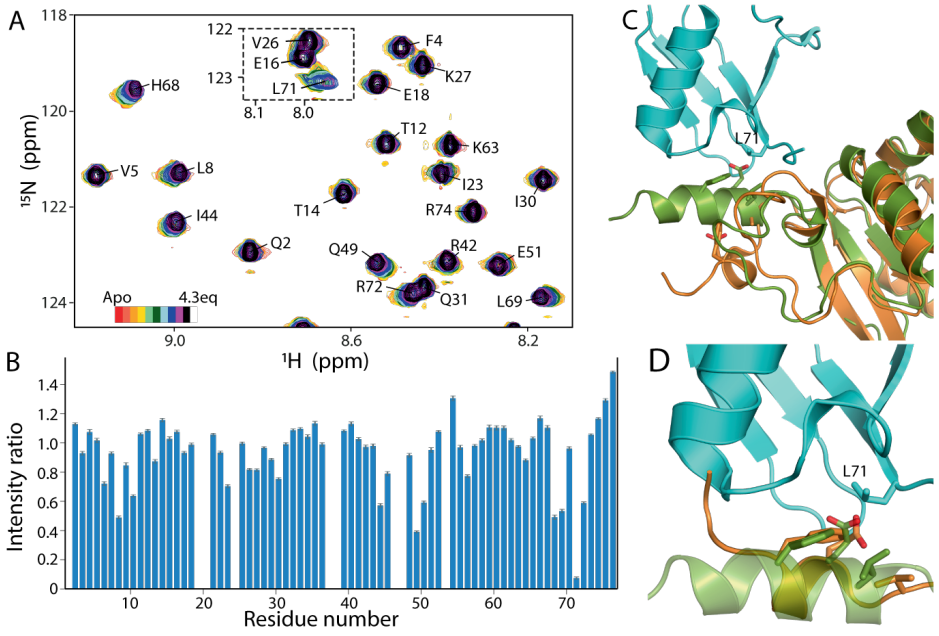


Figure 6. Dot1p binds weakly but specifically to ubiquitin. (A) Section of the 2D ^{15}N-H TROSY spectra recorded upon titration of labelled ubiquitin with Dot1pΔ. The inset highlights L71. Color coding indicated. (B) Ubiquitin peak intensity ratio between zero and four equivalents Dot1pΔ added. (C) Comparison of the ubiquitin-Dot1L interface with the Dot1p structure, with Ub in cyan, Dot1L in green and Dot1p in orange. Important residues for ubiquitin L71-binding described by Worden et al[21]. and corresponding residues from Dot1p are shown as sticks (PDBid: 6NJ9, 1U2Z). (D) Zoomed view from C with Dot1L and Dot1p aligned on the short C-terminal helix showing the conserved position of crucial sidechains L322, E323 and F326 in human, V559, E561 and F564 in yeast.

Discussion

The interaction of Dot1 with the nucleosome has been shown to rely on multiple interactions (namely DNA, the acidic patch, ubiquitin and the H4 tail) that together allow the formation of a highly specific complex[17,19-22]. Poor conservation of several of the Dot1 regions involved in these interactions together with functional differences in the dependency on H2B ubiquitination have pointed to significant differences in

binding mode between the metazoan and yeast Dot1. Here we studied the different elements of the nucleosome interaction of yeast Dot1p and attempt to integrate this into a yeast-specific binding mode.

In this study we aimed to exploit the atomic resolution of NMR to map the interaction of Dot1p on the nucleosome surface. Since the H3K79 loop is devoid of ILV residues that are required for solution NMR studies, we used the in-rotor-sedimentation approach that we previously developed for the study of nucleosome-protein complexes[23]. An underlying hope was that the sedimentation or the dense environment of the sediment would possibly promote the formation of a compact complex. Unfortunately, we did not observe the formation of the active state complex but rather obtained, most likely, the more loosely associated poised state. This could at least in part be due to the specific mutant used in this study. The lack of the SAM co-factor and replacement of the G401 with an arginine may have impeded binding of the basic H3K79 sidechain and prevented formation of the active state. In this regard the Norleucine replacement of H3K79 together with use of the wild-type enzyme would be preferred for future NMR studies.

Nevertheless, our NMR study revealed that the H3K79 loop has increased intrinsic flexibility. This is particularly relevant as H3K79 is too close to the nucleosome surface to reach the catalytic core of Dot1. While Worden et al demonstrated with Dot1L that the H3K79 loop undergoes a large conformational change in which the lysine ε-amino group is moved by 10 Å, it remained unclear whether Dot1L loosens H3K79 from the nucleosome surface or whether the stably held specific position of Dot1L allows it to trap a transiently open conformation of the H3K79[21]. The intrinsic flexibility observed here suggests that H3K79 may transiently be exposed and thus allow the conserved "W305" and "F131" loops in Dot1 to stabilize or extend this open state in synergy with the H4 tail.

In addition, a strong reduction in peak intensities for the H3 tail was observed upon Dot1pΔ binding, pointing to an involvement of H3 tail in the interaction. This could be due to reduced flexibility, dynamic exchange between free and bound states and/or increased proton density in the direct surroundings of the bound state, which can all

cause line broadening. Whether the H3 tail binds directly to the Dot1p core or is affected indirectly through the DNA binding of Dot1p remains to be established.

We showed that the Lys-rich region of Dot1p binds with nanomolar affinity to nucleosomes and DNA. One could argue that this characteristic poses a conundrum as it would likely anchor Dot1p tightly at an arbitrary position on the DNA, independently from proximity to H3K79. Moreover, unlike the situation in Dot1L where the Lys-rich region is flexibly linked to the catalytic core by a 55-residue linker, the Lys-rich region is connected to the Dot1p core by a short 5-residue linker. Thus, while the bound Dot1L Lys-rich region will still allow the core to scan the nucleosome surface for ubiquitin and the acidic patch and thereby form poised or active complexes, the Lys-region in Dot1p will heavily constrain the position of the catalytic core on the nucleosome surface. Following that reasoning, a high-affinity DNA interaction would be expected to strongly impede formation of a specific complex in which the catalytic core is properly oriented towards the target lysine.

A clue for a solution to this dilemma comes from the EMSA experiments. These consistently show that the affinity of Dot1pΔ for nucleosomes corresponds to K_D values in micro- rather than nanomolar range. We thus postulate that the Lys-region is bound to the folded core of Dot1p in the free state, thus reducing the affinity for the nucleosome. Of note, the crystal structure of Dot1p showed density for a few residues bound to the surface of the Dot1p core[13]. The authors could not determine whether these residues originated from an intramolecular contact with the N-terminus or from a crystal-lattice-induced contact with the C-terminus of the next Dot1p. Since these residues bind to a partly acidic surface, we speculate that the Lys-rich region is bound here. This auto-inhibition of DNA binding could then be relieved upon interaction with nucleosomes in a close to correct orientation (Figure 7A). In this orientation a specific component of the nucleosome, most likely the H3 or H4 tail, would compete off the Lys-rich region from the Dot1p surface and allow effective anchoring of Dot1p in a configuration compatible with H3K79 binding and methylation. In this reasoning, the Dot1p Lys-rich region acts as a hawser, mooring

the enzyme once in position, while the Dot1L Lys-rich region rather act as a hook on a long fishing line (Figure 7B).

An ubiquitin interaction motif was previously identified in the N-terminal region, corresponding to residues 100 to 140 of Dot1p[18]. Since this interaction would position ubiquitin N-terminal to the catalytic core and would link the two domains through a flexible linker, it is unlikely to fulfill the same orienting role as seen for Dot1L. Using a construct lacking the first 157 residues, we here showed that the Dot1p core interacts specifically but very weakly to ubiquitin, via a hydrophobic patch centered around L71. Since this interface partly overlaps with that of Dot1L and key Dot1 residues that mediate the interaction to L71 are conserved (L322, E323 and F326 in human, V559, E561 and F564 in yeast), we postulate that this mode of ubiquitin interaction is similar in Dot1p as in Dot1L. It will be very interesting to compare the ubiquitin affinities of Dot1L and Dot1p cores to see whether these correspond to their dependency on ubiquitin for successive methylation. In addition, the role of the N-terminal ubiquitin interaction motif remains to be established. Potentially, it could be important to recruit Dot1p to H2B ubiquitinated chromatin or act as a fishing hook through its DNA-binding activity[18].

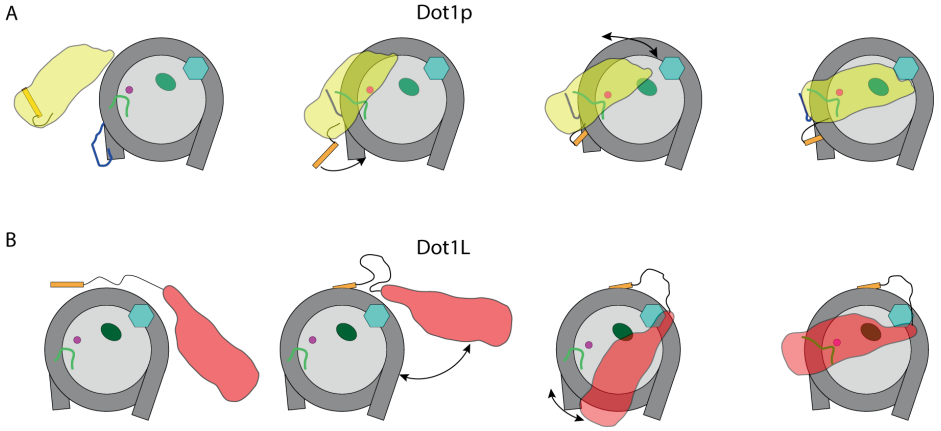


Figure 7. Model for Dot1pΔ and Dot1L interaction with the nucleosome. H3K79 and the H3/H4 tails are represented as a magenta circle, and blue/light green lines, respectively. (A) Dot1pΔ relies on preliminary interactions with the nucleosome, putatively to H3 tail, to deploy its N-terminal mooring Lys-rich region. Binding the nucleosomal DNA

would allow a stable poised state to be formed as due to the short linker, the catalytic core is effectively restrained on the nucleosome surface. The weak interaction with ubiquitin could promote the successive formation of the active state required for efficient di- and trimethylation. (B) Dot1L (red) uses the C-terminal Lys-rich motif (orange) to flexibly anchor itself to the nucleosome, allowing interactions of the catalytic core with the nucleosome acidic patch (dark green) and H2B-ubiquitin (cyan). This allows the formation of the poised state and subsequently the active state.

Conclusions

We here presented a detailed analysis of the nucleosome binding mode of Dot1p. Based on series of binding experiments we propose a Dot1p-specific binding mode that relies on the histone tails and the DNA-binding Lys-rich region to orient Dot1p in a methylation competent position on the nucleosome surface. We further showed that the Dot1p core has intrinsic, low affinity for ubiquitin and may use this additional interaction in a similar manner as Dot1L to stimulate formation of the active state. Finally, our NMR data revealed that the H3K79 loop is dynamic suggesting that the large conformation changes seen upon Dot1 binding are in part due to conformation selection. While several aspects of our binding model remain to be further validated, and others, such as the interaction with the acidic patch, are yet lacking, our data present novel insight into the yeast-specific molecular basis of Dot1p-mediated nucleosome methylation and constitutes a valuable starting point for further investigations.

Materials and methods

Nucleosome production

167bp 601 DNA, unlabeled and H3-labelled nucleosomes were produced as described previously[23] and stored in PK buffer (10 mM (K)PO_4 pH 6.5, 10 mM KCl). H2A/H2B dimers and H3/H4 tetramers were refolded at 2 M NaCl[34], and buffer exchanged to PK buffer for dimer, and PK with 200 mM KCl (PK200) for the tetramer.

Fluorescence anisotropy

The fluorescent DKR peptide, of sequence FAM-GEKKPLKKGRANKKNDRDSPSST-NH_2 corresponding to residues 157 to 178 of Dot1p (Uniprot iD: Q04089) and an extra N-terminal glycine, was ordered from Synpeptide as TFA salt. The powder was dissolved in miliQ water and pH was adjusted to 7.1 by adding diluted KOH. Peptide concentration was determined by NMR as described previously[23].

Peptide and nucleosome, dimer and 601 DNA stocks were diluted in PK buffer supplemented with 0.1 mg/ml BSA. The tetramer assay was made in PK200 with 0.1 mg/ml BSA. 5 nM of fluorescent peptide was mixed with a serial dilution of titrant, in triplicate, and imaged with a SpectraMax® I3 (Molecular Devices). In the competition assay, final concentrations of 500 nM nucleosome was mixed with 5 nM peptide, followed by the addition of a serial dilution of 601.

Dot1pΔ production

Dot1p Δ157 G401R (Dot1pΔ) was cloned in a pET16b and expressed in *E. coli* BL21(DE3). Expression was made in lysogeny broth (LB) at 19 °C and cells were lysed by sonication (4 min, 10 s on, 30 s off) in 25 mM HEPES pH 7.5, 200 mM NaCl, 2% glycerol, 1 mM TCEP and 5 μg/ml DNAseI. The cleared lysate was 2-step purified by cobalt affinity (Talon Superflow, GE Healthcare) followed by ion exchange (Resource S, GE Healthcare). Purified Dot1pΔ was checked for

oligomerization by size exclusion chromatography (Superdex 200 16/60) and stored in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP at –80 °C.

Thermal shift assay

An aliquot of Dot1pΔ was buffer exchanged to 10 mM HEPES pH 7.0, 25 mM KCl, 1mM DTT by concentration and dilution, and diluted 12.5-fold in an array of 15 mM potassium phosphate buffers with pH ranging from 6.0 to 8.0, supplemented with 0 to 100 mM KCl and 2x Sypro-Orange (sigma). Temperature titration and synchronous fluorescence measurement were made in a Biorad MyiQ® single color real-time PCR system coupled to a iCycler® thermal cycler.

EMSA

An aliquot of Dot1pΔ was buffer exchanged to PK-buffer supplemented with 0.5mM DTT. Serial dilutions were made in PK-buffer with 0.5mM DTT and 0.1 mg/ml BSA. Dot1p-nucleosome mix was then loaded in a 5% polyacrylamide gel casted in 0.2X TBE, pre-run on 0.2X TBE, and run for one hour at 120V at 4 °C.

Ubiquitin titration

An aliquot of ^{15}N-labelled ubiquitin (gift from Siddarth Narasimhan) and of Dot1pΔ were buffer exchanged to a 7.5% D_2O PK buffer supplemented with 250 μM DTT.

Dot1p co-sedimentation.

An aliquot of Dot1p was buffer exchanged to PK buffer with 250 μM DTT by concentration and dilution, then 3-step added to labelled nucleosomes in a 3-molar excess with 5 minutes waiting time between additions. The mix was supplemented with 2mM $MgCl_2$ and ultracentrifuged directly into a 1.3mm rotor via a funnel device for 37 hrs at 83 kG at a temperature of 4°C. The rotor was closed and directly inserted in the magnet.

NMR experiments

All solution state experiments were acquired on a Bruker 21.1 T (operating at 900 MHz ^{1}H Larmor frequency) equipped with an Avance III console and a CPTCI probe. Solid state NMR experiments were acquired on a 18.8 T (800 MHz Larmor frequency) equipped with 1.3 mm ^{1}H/X/Y triple-resonance MAS probe. A 3D CANH spectrum was recorded at 50 kHz and processed with NMRpipe. 3D CANH, CA(CO)NH and 4D versions of the CACONH and COCANH spectra for backbone walk were recorded at 60 kHz using a NUS (NUS) acquisition scheme and processed with NMRpipe[35] and SMILE[36]. Experimental setup of pulse sequences was according to our previous work[23]. The 3D CBCANH and HNH pulse sequences were implemented as described in the works of Xiang et al. and Zhang et al, respectively[37,38].

Table S1. Dipolar transfer parameters for solid state NMR ^{1}H-detected experiments

Transfer step	Contact time (μs)	Pulses
H->Cα	4500	Ramp 90 to 100
H->C'	4000	Ramp 90 to 100
H->N	1400	Ramp 80 to 100
Cα->C'	4000	Ramp 90 to 100
C'->Cα	4000	Ramp 90 to 100
Cα->N	5500	Ramp 90 to 100
N->H^{N}	750	Ramp 100 to 80
DREAM ^{13}C-^{13}C	6000	Ramp 80 to 100
RFDR ^{1}H-^{1}H	5300	

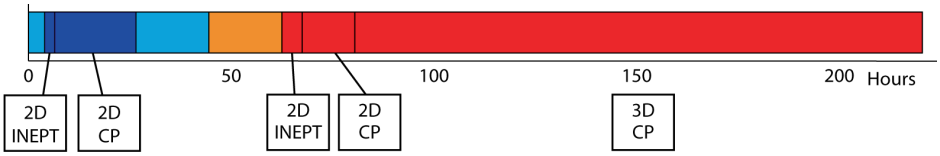


Figure S1. Order of acquisition for the spectra on H3-labelled nucleosomes cosedimented with Dot1pΔ. Blue and red colors denote 50kHz MAS and 60kHz MAS rates, respectively. Experiment setup times are represented in lighter colors

References

1 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).

2 van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745-756, doi:https://doi.org/10.1016/S0092-8674(02)00759-6 (2002).

3 Feng, Q. *et al.* Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain. *Current Biology* **12**, 1052-1058, doi:10.1016/s0960-9822(02)00901-6 (2002).

4 Fingerman, I. M., Li, H. C. & Briggs, S. D. A charge-based interaction between histone H4 and Dot1 is required for H3K79 methylation and telomere silencing: identification of a new trans-histone pathway. *Genes Dev* **21**, 2018-2029, doi:10.1101/gad.1560607 (2007).

5 Vlaming, H. & van Leeuwen, F. The upstreams and downstreams of H3K79 methylation by DOT1L. *Chromosoma* **125**, 593-605, doi:10.1007/s00412-015-0570-5 (2016).

6 Nguyen, A. T. & Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev* **25**, 1345-1358, doi:10.1101/gad.2057811 (2011).

7 De Vos, D. *et al.* Progressive methylation of ageing histones by Dot1 functions as a timer. *EMBO Rep* **12**, 956-962, doi:10.1038/embor.2011.131 (2011).

8 Frederiks, F. *et al.* Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. *Nat Struct Mol Biol* **15**, 550-557, doi:10.1038/nsmb.1432 (2008).

9 Debelouchina, G. T., Gerecht, K. & Muir, T. W. Ubiquitin utilizes an acidic surface patch to alter chromatin structure. *Nat Chem Biol* **13**, 105-110, doi:10.1038/nchembio.2235 (2017).

10 Krivtsov, A. V. *et al.* H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* **14**, 355-368, doi:10.1016/j.ccr.2008.10.001 (2008).

11 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

12 Min, J., Feng, Q., Li, Z., Zhang, Y. & Xu, R.-M. Structure of the Catalytic Domain of Human DOT1L, a Non-SET Domain Nucleosomal Histone Methyltransferase. *Cell* **112**, 711-723, doi:10.1016/s0092-8674(03)00114-4 (2003).

13 Sawada, K. *et al.* Structure of the conserved core of the yeast Dot1p, a nucleosomal histone H3 lysine 79 methyltransferase. *J Biol Chem* **279**, 43296-43306, doi:10.1074/jbc.M405902200 (2004).

14 Lu, X. *et al.* The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol* **15**, 1122-1124, doi:10.1038/nsmb.1489 (2008).

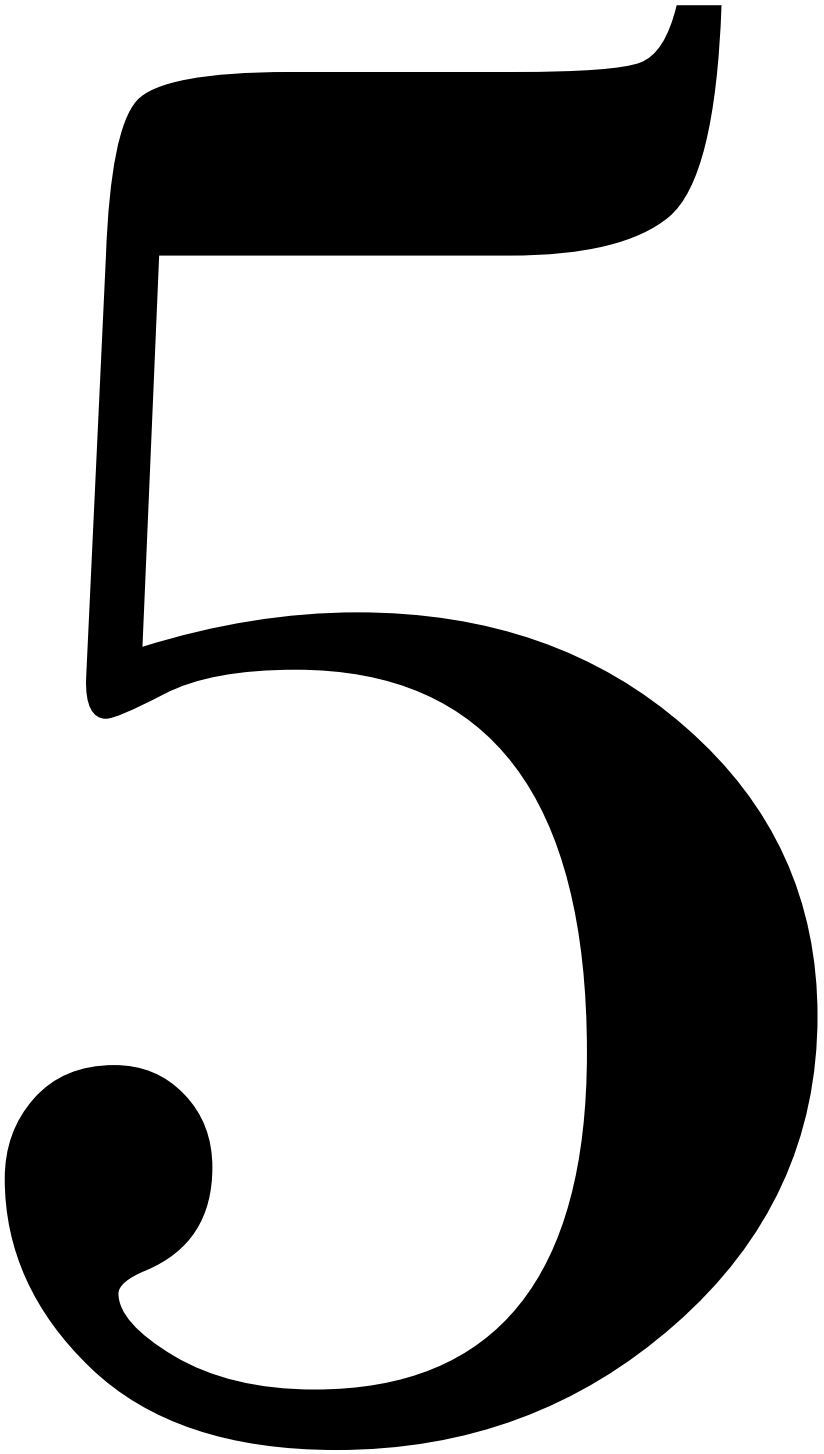
15 Shahbazian, M. D., Zhang, K. & Grunstein, M. Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. *Mol Cell* **19**, 271-277, doi:10.1016/j.molcel.2005.06.010 (2005).

16 McGinty, R. K. *et al.* Structure-activity analysis of semisynthetic nucleosomes: mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem Biol* **4**, 958-968, doi:10.1021/cb9002255 (2009).

17 Valencia-Sanchez, M. I. *et al.* Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. *Mol Cell* **74**, 1010-1019 e1016, doi:10.1016/j.molcel.2019.03.029 (2019).

18 Oh, S., Jeong, K., Kim, H., Kwon, C. S. & Lee, D. A lysine-rich region in Dot1p is crucial for direct interaction with H2B ubiquitylation and high level methylation of H3K79. *Biochem Biophys Res Commun* **399**, 512-517, doi:10.1016/j.bbrc.2010.07.100 (2010).
19 Yao, T. J., W.; Hu, Z.; Tan, Ming.; Cao, M.; Wang, Q.; Li, Y.; Yuan, G; Lei, M.; Huang, J. Structural basis of the crosstalk between histone H2B monoubiquitination and H3 lysine 79 on nucleosome. *Cell Research* **20**, 330-333, doi:10.1038/s41422-019-0146-7 (2019).
20 Anderson, C. J. *et al.* Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase. *Cell Rep* **26**, 1681-1690 e1685, doi:10.1016/j.celrep.2019.01.058 (2019).
21 Worden, E. J., Hoffmann, N. A., Hicks, C. W. & Wolberger, C. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**, 1490-1501 e1412, doi:10.1016/j.cell.2019.02.002 (2019).
22 Jang, S. *et al.* Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. *Genes Dev* **33**, 620-625, doi:10.1101/gad.323790.118 (2019).
23 Xiang, S. *et al.* Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **57**, 4571-4575, doi:10.1002/anie.201713158 (2018).
24 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041, doi:10.1073/pnas.181342398 (2001).
25 Cheng, Y.-C. & Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochemical Pharmacology* **22**, 3099-3108, doi:10.1016/0006-2952(73)90196-2 (1973).
26 Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* **79**, 28 29 21-14, doi:10.1002/0471140864.ps2809s79 (2015).
27 Zhou, B. R. *et al.* Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J Mol Biol* **421**, 30-37, doi:10.1016/j.jmb.2012.04.032 (2012).
28 Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**, 227-241, doi:10.1007/s10858-013-9741-y (2013).
29 Shi, X. *et al.* Structure and Dynamics in the Nucleosome Revealed by Solid-State NMR. *Angew Chem Int Ed Engl* **57**, 9734-9738, doi:10.1002/anie.201804707 (2018).
30 Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19-42, doi:10.1006/jmbi.1997.1494 (1998).
31 Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1-10, doi:10.1016/j.jmb.2010.08.039 (2010).
32 Chua, E. Y., Vasudevan, D., Davey, G. E., Wu, B. & Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res* **40**, 6338-6352, doi:10.1093/nar/gks261 (2012).
33 Stutzer, A. *et al.* 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, doi:10.1016/j.molcel.2015.12.015 (2016).
34 Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Chromatin and Chromatin Remodeling Enzymes, Pt A* **375**, 23-44 (2004).
35 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, doi:10.1007/bf00197809 (1995).

36 Ying, J., Delaglio, F., Torchia, D. A. & Bax, A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR* **68**, 101-118, doi:10.1007/s10858-016-0072-7 (2017).
37 Xiang, S., Biernat, J., Mandelkow, E., Becker, S. & Linser, R. Backbone assignment for minimal protein amounts of low structural homogeneity in the absence of deuteration. *Chem Commun (Camb)* **52**, 4002-4005, doi:10.1039/c5cc09160h (2016).
38 Zhang, R., Nishiyama, Y., Sun, P. & Ramamoorthy, A. Phase cycling schemes for finite-pulse-RFDR MAS solid state NMR experiments. *J Magn Reson* **252**, 55-66, doi:10.1016/j.jmr.2014.12.010 (2015).



Chapter 5

On the sedimentation of nucleosomes and nucleosome-protein complexes for solidstate NMR studies

Contributions:
Marco M. R. M. Hendrix (TU Eindhoven) performed the SAXS measurements. Yi Zhang (UC Denver) produced and purified the PHD2 finger from CHD4, dr. ShengQi Xiang performed the solid-state NMR experiments. dr. Stuart C. Howes performed the EM measurements.

Published as:
------ *in preparation* ------

Abstract

We had recently adapted sedimentation for the study of the nucleosome and interactions therewith by solid state NMR. However, nucleosomes are long known to be able to interact with each other in condensed conditions, forming superstructures that would hinder the general applicability of our method. We used NMR, SAXS and SEM to characterize the dense phase generated in our method. We show that the nucleosomes in our sediment remain isolated and hydrated despite the compaction. Additionally, we used our method to study the low affinity interaction between the PHD2 finger of CHD4 to the unmodified H3 tail. We show that the cosedimentation approach is relevant for low affinity binding studies, and that even if the bound state is invisible, the binding site can still be identified.

Introduction

The nucleosome is the central interaction platform for a multitude of proteins and protein complexes to bind and regulate the function of chromatin[1]. The precise timing and specificity of these interactions are critical for the proper function of a cell. One of the key questions is thus to find out how proteins bind to nucleosomes and to what extent their binding and function is affected by higher-order chromatin structure.

We recently introduced an NMR based method to study nucleosome-protein interactions in a dense sediment. Using ^{1}H-detected solid-state NMR (ssNMR), this approach allows to observe nearly all ^{13}C-^{1}H and ^{15}N-^{1}H atom pairs in an isotope-labelled histone without requirement for selective labeling as in the methyl-TROSY solution NMR approach. It thus offers more NMR observables per histone, and more complete mapping and characterization of nucleosome-protein complexes, in particular when combined with distance- and dynamics information (Chapter 3, Chapter 4). Additionally, since the method is based on ssNMR, it can be used for the study of larger assemblies like whole chromatin fibers or nucleosomal arrays[2,3].

In this method, nucleosome and binding partner are mixed and then subjected to ultracentrifugation to create a sediment of the nucleosome complex directly in the sample container, the NMR rotor. Sedimentation has long been used to create samples suitable for solid state NMR as it is a safe, fast and easy-to-use technique, as opposed to precipitation, lyophilization or crystallization that are potentially damaging and/or cumbersome. Moreover, a series of previous studies showed that sedimentation of soluble proteins results in high-quality samples for solid-state NMR studies[4-6]. Additionally, within the context of sedimentation velocity analytical ultracentrifugation (SV-AUC), sedimentation has long been used for the study of nucleosomes and nucleosomal arrays, attesting to the non-damaging nature of the technique[7].

We here aim to characterize in detail the packing of nucleosomes in the sediment. Since sedimentation creates a very dense sample phase with tightly packed particles, it may promote formation of nucleosome-nucleosome interactions that affect the

availability of the nucleosome surface for protein interactions and thus interfere with the study of native nucleosome-protein interactions. This is particularly relevant since nucleosomes are well known to interact with each other, capable of forming various forms of short and long stacks both *in vivo*[8-10] and *in vitro* [11-16]. Notably, isolated nucleosomes are able to stack into columns in highly concentrated solutions[17-25]. The exact type of higher-order structure formed and exact type of interactions responsible depend on many factors, such as DNA linker length, salt types and concentrations, *etc.* [19,23,26].

In addition, we aim explore the requirements on binding affinity and binding epitope within the nucleosome of the co-sedimentation technique. In chapter 3, we focused on a nanomolar interaction to the nucleosome core surface that due to its high-affinity allowed efficient co-sedimentation. Due to rigid nature of the nucleosome core the interaction could be studies through dipolar-based experiments. However, many proteins bind nucleosomes with much lower affinities, with dissociation constants (K_D) in the micromolar range and many bind the highly flexible histone tails rather than the nucleosome core. We here tested our method on the second PHD finger (PHD2) or chromatin remodeller CHD4, part of the NuRD complex that is involved in DNA repair and cell cycle progression[27]. CHD4 contains a paired PHD finger domain of which PHD2 binds with micromolar-range affinity (4 +/-1 μM) to non-modified H3 tail peptides[28,29] and initiates binding to nucleosomes[30]. Binding is strongly impeded in the nucleosomal context due to restricted availability of the histone tails in the nucleosome, thus further decreasing the effective affinity in a salt-dependent manner [30,31]. The low affinity, yet specific binding to unmodified H3 tails together with its small size and the salt dependent affinity, make the PHD2 interaction an interesting test case to investigate the general applicability of our co-sedimentation approach. In this case, it could provide detailed structural information on the impact of the nucleosomal environment on the PHD2-tail interaction.

In this work, we combined ssNMR, small-angle X-ray scattering (SAXS) and scanning electron microscopy (SEM) to characterize the nucleosome packing in our sample

preparation method and report the case study of CHD4 PHD2 finger-nucleosome interaction as a worst-case scenario test of the method.

Results

Nucleosomes are tightly packed in the sediment

As a first measure of the degree of nucleosome packing created by ultracentrifugation of nucleosomes into the NMR rotor, we assessed the nucleosome concentration within the rotor for four recent sample preparations (Table 1). Briefly, the sedimentation process is started from a dilute solution of nucleosomes with or without binding partner, typically 500 µL at 4 mg/mL, placed in a custom made device[32]. This is then centrifuged at 83,000 G for 24 hours at 4 °C. As seen from Table 1, the homogenized supernatant after sedimentation retains, with one exception, only 2-5% of the initial UV absorbance, indicating a near-quantitative sedimentation. For sample 4 a much higher nucleosome concentration in the supernatant was observed, but this can be rationalized by the also much higher starting mass, see below. Upon removal of the sediment from the very top of the rotor to make room for placement of the cap, a dense and transparent viscous droplet was formed in all cases. This indicates the rotor is filled with a dense solution rather than a precipitate. The final nucleosome mass in the rotor is estimated to be 1.44 – 1.59 mg, resulting in concentrations in the range of 480 to 530 mg/mL or 2.3 to 2.5 mM. This value is similar to the in-rotor concentration reported by Shi et al[2] using Mg^{2+}-induced precipitation of nucleosomes. Notably, starting from higher nucleosome mass in the sedimentation mix, as for sample 4, does result in significant increase in in-rotor nucleosome concentration, indicating the observed values are close to the limiting concentration. Assuming the nucleosomes to be homogeneously distributed through the volume of the packed rotor and approximating the volume of one nucleosome to be of 420 nm^3 [33], the observed nucleosome concentration corresponds to a packing ratio of ~61%. These concentrations and packing ratios of the nucleosome sediment are close to, but slightly lower than those found in nucleosome crystals. Inspection of crystallography

parameters from four nucleosome crystal structures (PDBid's 2pyo, 1kx5, 1aoi, 3lz0)[34-37] show that the typical concentrations are ~2.9 mM, with packing coefficients of ~67% and a solvent content of ~54%. Assuming the nucleosome to be a flat disk, the tightest packing theoretically achievable is formed by a homogeneous hexagonal arrangement of columns of stacked disks, which would correspond to 3.1 mM or 655 mg/mL nucleosome concentration and a packing ratio of 78%. These simple considerations indicate that the sediment is highly dense with a packing ratio close to 80% of the maximum, suggesting that a significant amount of ordering and nucleosome-nucleosome interactions may occur.

Table 1. Estimated in-rotor concentration of nucleosome sediments

Nucleosome mass[a] (mg) in:	sample 1 (H3 assignment)	sample 2 (H2A assignment)	sample 3 (LANA complex)	sample 4 (H3 assignment)
- initial solution	1.98	1.90	1.90	2.76
- supernatant	0.06	*n.d.*	0.11	1.03
- cap clearing[d]	0.38	*0.35*[b]	0.35	0.11[e]
- closed rotor	1.54	*1.45*[b]	1.44	1.59
final nucleosome concentration in rotor (mg/mL)[c]	514	484[b]	481	529

[a] Nucleosome mass determined from absorbance measurements at 260 nm assuming all absorbance originates from nucleosomal DNA.

[b] Estimated mass and concentration assuming 95% sedimentation efficiency.

[c] Calculated using an internal volume of 3 uL for the 1.3 mm rotor.

[d] Estimated assuming the nucleosome distribution to be homogeneous in the rotor, for a cleared space of 0.73µl

[e] Measured by diluting the cleared material in buffer and measuring absorbance

Nucleosomes are stably folded and remain hydrated in the sediment.

The 1D single-pulse ^{1}H NMR spectrum of the sediment is dominated by an intense water signal, indicating the nucleosome sediment is highly hydrated. Comparison of these spectra throughout the measurements reported in Chapter 3[38] shows that the water signal remains prominent over time, despite the exposure to 34 days of high-speed magic-angle-spinning (MAS) at effective temperature of 37 °C. The intensity at peak maximum decreases by 20% over this time while the line width increases by 40%. While these spectra are not suitable for integration due to the very intense nature of the water signal, its variability in line shape and long T_1 relaxation time, we can nonetheless safely conclude that the sample remained strongly hydrated throughout the measurements.

Furthermore, the 2D CP-based NH correlation spectra recorded at the beginning and the end of the measurements (five months after rotor closing) do not show large changes and are both of high quality, showing a well-resolved and well-dispersed spectrum (Figure 1C). Together with the presence of a single set of resonances without peak doubling, other than for residues close to the non-palindromic 601-DNA, indicates that the histone is well-folded and has a homogeneous chemical environment throughout the sediment that is constant throughout the measurements. Small chemical shift changes are observed for residues A44, G45, G66 and I78, that are in the vicinity of buried waters or salt ion in the crystal structure[35,39].

Comparison of J-based spectra of H3-labeled mononucleosomes show that the H3 N-terminal tail has a highly similar chemical environment in solution and sediment[38]. Since the H3 tail has been shown to bind to nucleosomal DNA in solution[40], this suggests that within the sediment the H3 tail binds similarly to the proximal nucleosomal DNA or, if mediating inter-nucleosome interactions, binds DNA in a non-specific manner. Together these data demonstrate the nucleosomes in the sediment remain well-folded and hydrated through the measurements, but do not show clear evidence for direct nucleosome-nucleosome contacts.

To further investigate the packing within the sediment, we recovered the contents of the rotor from a sample. To our surprise, the sediment had formed a translucid hard

paste that was delicately hammered out of the rotor (figure 1A). This suggests that some dehydration had taken place without negatively affecting nucleosome folding, as evidenced by the high-quality NMR spectra, or without strongly affecting the abundance of water in the 1D spectrum. We speculate that the gradual increase in water line width may correlate with the transition of the gel-like solution to a paste-like material.

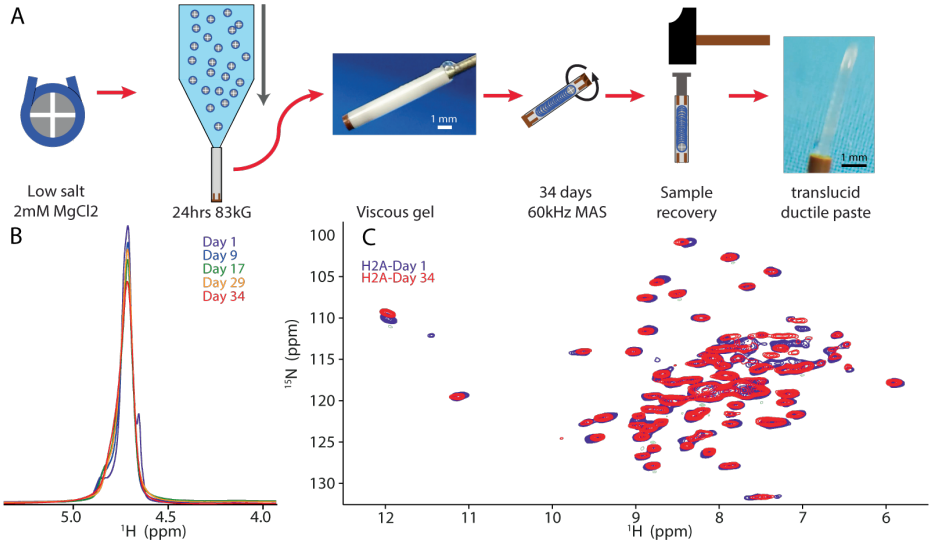


Figure 1. Nucleosomes are stably hydrated in the sediment. A: Sample sedimentation and recovery. The nucleosome solution was ultracentrifuged directly into the 1.3 mm rotor, leading to nucleosome sedimentation. The sample was collected after 34 days of MAS at an effective temperature of 37 °C in a 5-week period with intermittent storage at 4 °C. The caps of the rotor ended up being sealed by the sample and broke off during the opening, forcing us to hammer the sample and the rest of the caps out. B: Overlay of the 1D one-pulse ^{1}H spectrum of the sample acquired throughout the ssNMR measurements, showing the highly dominant water signal. C: Overlay of the 2D ^{1}H-detected NH correlation spectrum acquired at the beginning (blue) and end (red) of the NMR measurements.

Nucleosome sediment lacks long-range ordering

To investigate the packing and ordering of nucleosomes in the recovered sediment, we first collected SAXS data on a mononucleosome solution to further confirm

proper folding of nucleosome (Figure 2A). The scattering spectrum of the nucleosome solution can be properly fitted to the form factor of a cylinder with main dimension of 56.9 Å and a maximum particle diameter of 133 Å (Figure 2B), which correspond to the average thickness and end-to-end length of a nucleosome with 10 bp of linker DNA, respectively[14]. The curve corresponds well with previous reports SAXS experiments on soluble nucleosomes with linker DNA both qualitatively and quantitatively[41].

The recovered sediment paste (Figure 1A) shows a strikingly different scattering diagram than soluble nucleosomes (Figure 2C). We observed three broad peaks with Q-values that are multiples of the first. This regular pattern suggests a laminar organization with a main characteristic distance of ~74 Å. The very broad appearance of the scattering peaks, which is far beyond the experimental resolution, either reflects a heterogeneous distribution of the characteristic distance across the sample, or indicate the organization is only regular over a short distance. In this latter case, the length scale of the regular structure can be estimated to be ~180 Å based on the full-width at half-maximum (FWHM) of the peak.

While this scattering profile could reflect a "columnar phase" organization[22], comparison to literature data shows that our result is very distinct from the previous reports on condensed nucleosome core particles (Figure 2D). In an isotropic columnar phase, in which the columns are not highly ordered with respect to each other, only few broad scattering peaks are detected[19]. (Black trace in Figure 2D). The first peak was assigned to the intercolumnar distance, followed by peaks relating to the form factor of the column and stacking distance between the core particles composing the column. In a highly ordered columnar phase, such as obtained from Mg^{2+}-induced precipitation of nucleosome core particles, the previously broad signals are replaced by highly resolved signals at the same Q-values[23]. (blue trace in Figure 2D). Increase in DNA linker length increases the intercolumnar distance[19,21] (dashed line Figure 2D). Extrapolation of the literature data suggests that for our nucleosome preparations, that contain a 167 bp DNA sequence, the first scattering peak would be expected at Q ~0.06 $Å^{-1}$ for a columnar phase ordering. Since this first peak is at 0.085

$Å^{-1}$ for the sediment paste (Red trace in Figure 2D), we conclude that the nucleosomes are not mainly arranged in a columnar order.

Taken together, the SAXS and NMR data suggest that the organization of the nucleosomes is highly heterogeneous, while the direct environment of the nucleosome surface is homogeneous. Some short length structures cannot be excluded, e.g. the data are consistent with the presence of small stacks of up to three nucleosomes, accounting for a length of ~180 Å and inter-particle distance of 74 Å, without a significant preference in relative orientation between the stacks. Alternatively, the nucleosome packing in the sediment has no short-range structure. The nucleosomes remain dispersed without surface-based contacts between them. They have highly heterogeneous relative orientations, but are nonetheless in a closely compacted state with a mean nucleosome-nucleosome pattern length of 74 Å. In either case, the presence of long columnar phases or strong ordering can be safely excluded.

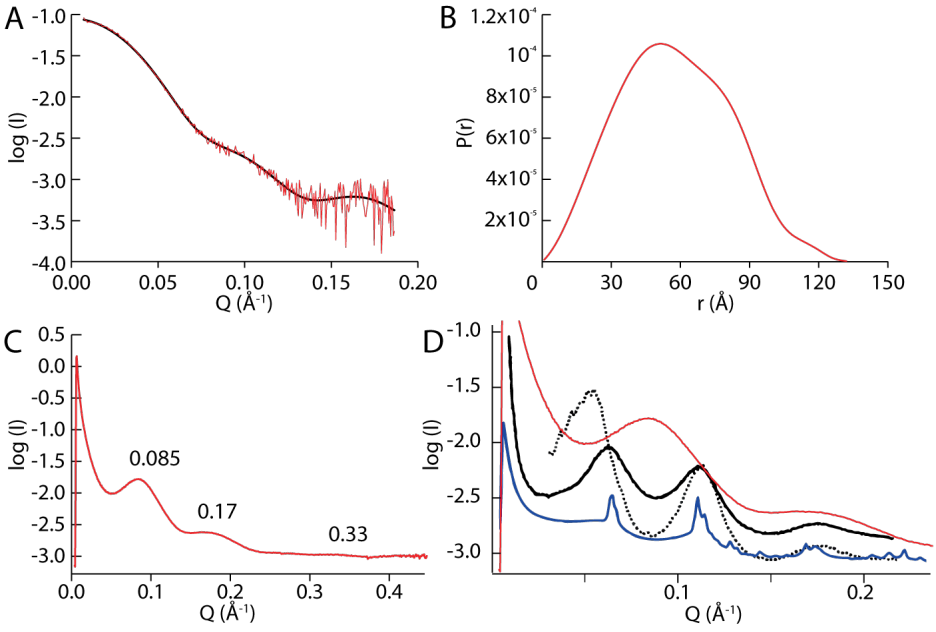


Figure 2. Small angle X-ray scattering of the mononucleosome. A. SAXS on soluble nucleosomes in 25 mM ionic strength buffer. The buffer-substracted scattering profile of soluble mononucleosomes was fitted in Dammif[42] to a flat monodisperse particle function. B. distance distribution calculated by Dammif. C. X-ray scattering of the recovered sediment

paste formed from monunucleosomes sedimented in 30 mM ionic strength buffer, including 2mM $MgCl_2$. Medium- and small-angle data were merged. Three broad peaks are annotated and reflect a characteristic distance of 73.6 Å ($2\pi/Q_{max}$). D. Comparison of SAXS traces with literature data. Blue: 20mM Mg^{2+}-precipitated nucleosome core particles forming an orthorhombic structure[23]. Black: mononucleosomes with 155bp ± 7 bp DNA in columnar phase[21]. Dashed: mononucleosomes with 170 ± 15 bp nucleosomal DNA in columnar phase[21]. Red: sediment of mononucleosome with 167 bp DNA (our work, same curve as in C).

To gain further insight on the structure occurring in our samples, a piece of the sediment paste was subjected to scanning electron-microscopy (SEM). The shearing induced by the sample recovery and cutting, alongside with air exposure, typically affect the sample surface (Figure 3A). To circumvent this, we used focused-ion-beam milling to expose a preserved region. However, our current attempts at sample preparation and mounting did not yet allow to reach a sufficient resolution to assess the nucleosome packing in the sediment. (Figure 3B). Native- and SDS-PAGE analysis of the resuspended sediment reveal nonetheless that despite exposure to X-rays, several rounds of sample manipulation and storage at 4 °C for several months, a large part of the nucleosome preparation remained intact (Figure 3C).

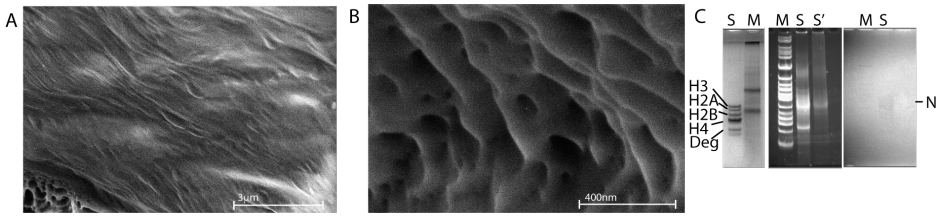


Figure 3. SEM and PAGE analysis of the sediment. A: view of the surface of the sediment. B: cutting-exposed surface of the sediment. C: SDS (left) and native PAGE (right) of the resuspended sediment. S, S' and M designate sediment, diluted sediment and molecular weight marker, respectively. Nucleosomal band (N) and histones are highlighted in native and SDS PAGE, respectively. A small portion of tail-less histones are obtained, resulting from a potential protease contamination after sample recovery (Deg).

Optimization of PHD2 co-sedimentation.

Having established that the nucleosome sediment in our studies is not strongly ordered and is thus likely to only minimally interfere with protein binding, we next turned to investigate the co-sedimentation process in more detail using the CHD4 PHD2 domain as a stringent test on the general applicability. Before attempting co-sedimentation, we first assessed binding of PHD2 to nucleosomes under various conditions by solution NMR by titrating H3-labeled mononucleosomes with PHD2. At low salt (25 mM ionic strength) little to no spectral changes were observed after addition of up to 1 molar equivalent of PHD2 finger for the H3 tail N-terminal residues (data not shown), in-line with previous reports[30]. At high salt (125 mM ionic strength) binding was visible as a peak intensity decrease for specific residues in the H3 N-terminal region (Figure 4A). Notably, no significant chemical shift perturbations or new signals were observed, suggesting the bound state chemical shift are invisible in solution NMR. Upon addition of 10 molar equivalents (315 μM), peak intensity dropped to ~50% relative intensity for the N-terminal tail residues, with no intensity change for the C-terminal half of the tail (Figure 4C). Residues T3, K4, T6 and A7 show the largest intensity reduction, which, when fitted to a single binding site model, yields a K_D of 168 ± 8 μM.

To investigate the origin of the increase in binding affinity upon increase in ionic strength, we compared H3 tail spectra at 25 and 125 mM added KCl. A doubling in peak intensity is observed for residues 19 to 21 and 23 to 29 (Figure 4B, D), thereby reaching intensities equivalent to that of the ten N-terminal residues. Small chemical shift perturbations are observed for several of these residues (see for instance K27 and S28 in Figure 4B), which when compared to the data of Stützer et al[40] signify a slight shift towards to DNA-free state. This indicates that primarily this intermediate region in the H3 tail becomes more flexible, and more accessible upon addition of salt. Within the PHD2 binding site, only T6 shows a significant shift towards the DNA-free state, whereas other residues are unaffected by addition of salt. It should be noted that the resonance of A1 and R2 are not observed and that thus effects for the very of end the tail cannot be probed directly. Nevertheless, these data suggest

that increase in salt level does not result in a large-scale increase in tail availability for the residues composing the PHD2 binding site. We conclude that at 125 mM KCl the tail is still predominantly bound to the DNA, in various conformations and to various DNA sites that are in fast exchange with each other, and that the increase in binding affinity originates rather from a weakened microscopic tail-DNA interaction. The failure to observe the PHD2-bound state of the H3-tail in these solution NMR experiments suggests that the bound tail may be significantly rigidified and thus might be better probed via ssNMR.

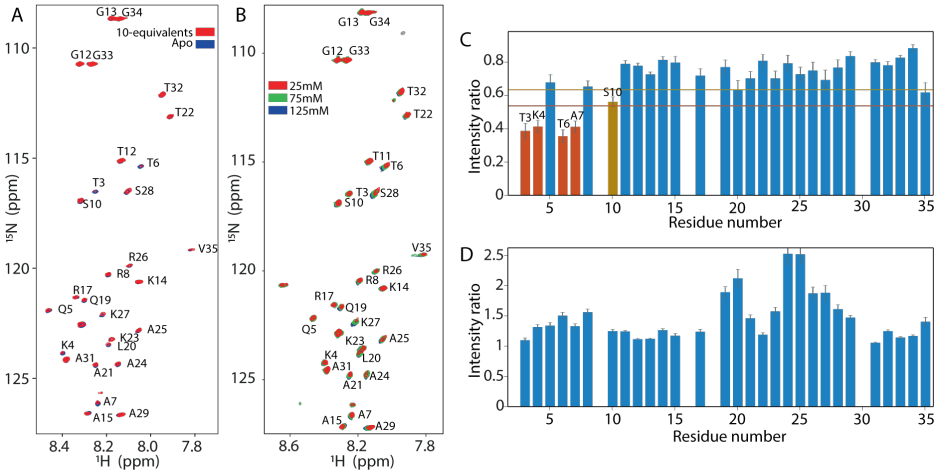


Figure 4. PHD2 binds the H3 tail in nucleosomes at high salt conditions. A. Titration of H3 tail with PHD2 at 125 mM ionic strength. B. KCl titration of the H3 tail, from 25 mM to 125 mM ionic strength. Color coding indicated. C. Peak intensity ratio of H3 tail resonances between free and bound nucleosomes. Addition of 10 equivalents of PHD2 results in large intensity decrease for the N-terminal residues of the tail that comprise the PHD2 binding site. One- and two standard deviation intensity differences from a 10%-trimmed average are displayed in orange and red, respectively. D. Peak intensity ratio of H3 tail resonances between spectra recorded in 25 and 125 mM KCl. Higher ratios mean increased intensity at higher salt.

Co-sedimentation of PHD2-nucleosome complex

Having established that PHD2-binding leads to detectable NMR signal changes in the H3-tail at high salt, the protein and nucleosomes were mixed at 125 mM KCl in a 20:1

CHD4 to H3 tail molar ratio, and then directly ultracentrifuged into the rotor for 24 hours. Again, a gel-like droplet was recovered while clearing space for rotor closure. Both J- and CP-based spectra were recorded at 50 kHz MAS and compared to spectra of H3-labeled nucleosome without binding partner that were recorded at 25 mM KCl. The J-based spectrum, probing the flexible parts of H3, displays a similar residue-specific drop of peak intensity as observed in solution (Figure 5A, B). Even if the lack of resolution in these spectra hinders interpretation somewhat, it can clearly be seen that resonances for the first ten tail resides show significantly decreased peak intensities, down to 30-50% of the original intensity. Again, no new peaks corresponding to the bound state can be observed, suggesting rigidification of the H3 tail in the bound state. The CP-based spectrum of the co-sediment, probing the rigid parts of H3 in the nucleosome, is well-resolved and well-dispersed and corresponds well to the CP-spectra of unbound nucleosomes (Figure 5A). Careful examination of the spectra unfortunately does not reveal resolvable changes in the peak pattern. Observation of the bound state may be hindered by peak overlap, line broadening due to increased local proton density and/or lower transfer efficiencies when the bound state is not fully rigid but has intermediate dynamics. We conclude that the despite the low affinity the PHD2 finger can successfully be cosedimented with the nucleosome and that despite its flexible binding site, specific binding of the PHD2 finger to the H3-tail can be demonstrated using the sediment ssNMR approach. Unfortunately, the PHD2-bound state is not directly observable, preventing further detailed structural characterization of the bound H3-tail conformation.

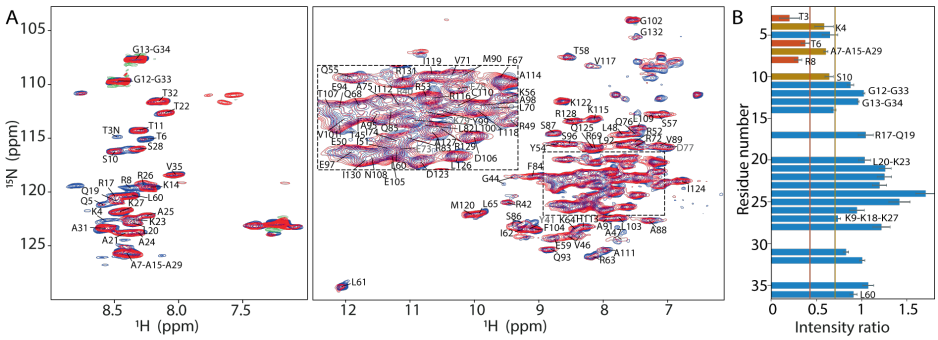


Figure 5. PHD2 finger co-sediments with the nucleosome and has the same effect on H3 tail as in solution. A. J-based (left panel) and CP-based (right panel) spectra of H3 co-sedimented with PHD2 finger in high salt, overlayed with the spectra of the apo nucleosome in low salt, displayed in red and blue, respectively. B. Intensity differences between the J-based spectra in A.

Discussion

We here characterized in some detail the nucleosome sediment that is central to our ssNMR investigation of nucleosome-protein interactions. We find that the sedimentation is robust and reproducible. The nucleosome concentration in the sediment approaches that observed in a crystal, allowing the sensitivity required for ssNMR. Nucleosomes remain well-folded, hydrated in the sediment over the course of several weeks of MAS. Microscopically, nucleosomes are exposed to homogeneous environment, while arranged heterogeneously on a macroscopic scale. In our view, the sediment corresponds to a highly concentrated solution in which nucleosomes are forced in close proximity, close enough to prevent overall tumbling, but distant enough to prevent long-range ordering (Figure 6A). This balance most likely results from both electrostatic repulsion caused by the nucleosomal DNA and stabilizing (transient) inter-nucleosome interactions. Fragai et al[43] report that sedimentation of highly charged proteins typically results in low packing ratios in sedimentation, while higher-than-crystalline concentrations can be achieved for proteins with low overall charge. Despite the high overall net negative charge of the nucleosome, we find

packing ratios ~80% of that in nucleosome crystals, pointing to a contribution of attractive interactions to counteract the electrostatic repulsion. Even if starting from low-salt conditions in which electrostatic repulsion is detrimental to nucleosome stacking[25], sedimentation will also concentrate the positive ion sphere associated with a nucleosome. Together with the asymmetric charge distribution of nucleosomes, negative DNA and positive histones, this promotes attractive electrostatic nucleosome interaction via stacking[44]. Thus, high nucleosome concentrations within sediment can be achieved.

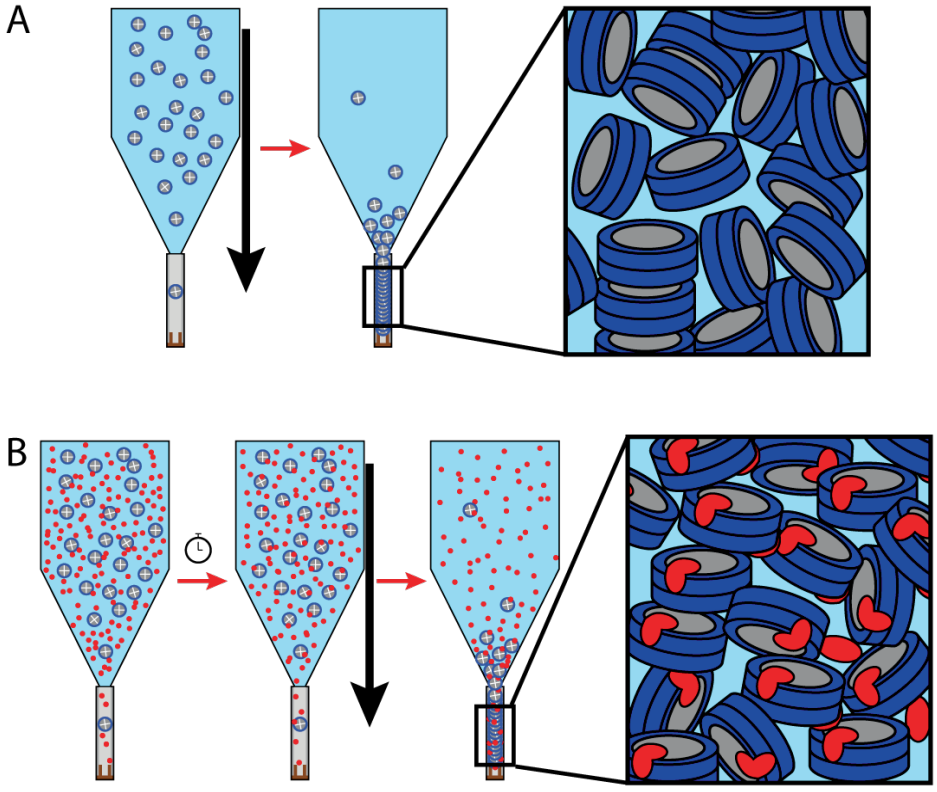


Figure 6. The ultracentrifugation brings the nucleosome in a dense, heterogeneous sediment. The black arrow indicates ultracentrifugation. A. the case of sedimentation of pure mononucleosomes. B. the case of co-sedimentation. The saturated nucleosome:binding-partner mixture is incubated for typically 20 minutes before proceeding

to ultracentrifugation, ensuring binding to reach an equilibrium. In the case of a small binding partner, the majority of the sediment is occupied by nucleosomes, free, 1:1 and 2:1 bound.

However, several unknowns remain. The effects of the very fast MAS on the sample organization are still elusive. The spinning speeds attained during MAS are so high that the centrifuge effect generates a solvent-based pressure reaching 96 atm near the rotor walls[45], nearly four times higher than the strongest osmotic pressures investigated by Mangenot et al[21] in a study of nucleosome self-association. Additionally, nucleosomes may spontaneously align when exposed to a magnetic field and retain such alignment upon forming the columnar phase[21]. Such magnetic alignment or even reorientation of the nucleosomes most likely does not occur in our conditions, as the sample was constantly spinning at 3.6 million RPM (60 kHz) at the magic angle relative to the magnetic field.

Several studies have investigated dense phases of nucleosomes. These either relied on reversible multivalent ion induced precipitation of into columnar phases[18,23], or high salt concentrations and high osmotic pressures to slowly condense nucleosomes to increasingly ordered phases, ranging from isotropically arranged sets of columns to dense orthorhombic or hexagonal networks of columns[19,21]. Our approach differs in a number of ways that in retrospect have helped to avoid formation of a strongly ordered sediment. First, we use mononucleosomes containing 10 bp of additional linker DNA, adding more net negative charge. Second, the Mg^{2+} concentration used in our study is below the minimum required to precipitate nucleosomes, and in addition, the use of K^+ instead of the harder Na^+ monovalent salt disfavors precipitation[24,26]. Finally, since ultracentrifugation is relatively fast, taking a day instead of months as in the osmotic pressure-based approach, it also impedes the formation of large-scale ordering.

As the sediment is devoid of strong ordering, we suspect that the sedimentation of an already formed, sufficiently stable nucleosome-protein complex would not be impeded by (transient) nucleosome-nucleosome contacts (Figure 6B). Indeed, we succeeded here to cosediment a protein that only weakly binds the nucleosome.

Similarly, our previous experience with the LANA peptide (Chapter 3) and Dot1p (Chapter 4) showed that complex formation did not result in loss in spectral quality, suggesting that the nucleosome packing within the sediment is not strongly perturbed. As an ultimate test of the applicability of the sedimentation method, we studied the low affinity interaction between the H3 N-terminal tail and the CHD4 PHD2 finger. The H3 tail has been shown to electrostatically interact with the nucleosomal DNA[30,31,37,40], yet is strongly disordered and flexible enough to be observable in J-based experiment both in solution and in condensed conditions [3,38,46]. Proteins that bind to the H3 tail must thus compete with the nucleosomal DNA in order to bind, which is also demonstrated here for the PHD2 finger. Surprisingly, solution NMR titration experiments revealed that despite a low affinity to the H3 tail, the binding of PHD2 occurred in the slow exchange regime, usually associated with long occupancy times and thus high affinities. This suggests that either also the association rate is slow, or that the competitive effect masks a higher affinity. Even though PHD2 binds the very tip of the long H3 tail and can thus be expected to be flexibly linked to the nucleosome core, signals from the bound state could not be observed. We thus resorted to ssNMR, posing several challenges to system. The first, successful co-sedimentation of the small and weakly binding PHD2 domain, was achieved by starting the co-sedimentation from a large molar excess of PHD2. Despite this excess, only partially saturated H3 tail was obtained as evidenced by the specific peak intensity drop for the PHD2 binding site. It is likely that the long complex life time helped to achieve co-sedimentation of the complex. The second challenge is the residue-specific interpretation of the changes in the NMR spectrum. Since the linewidths in ssNMR studies are broader than in solution (Chapter 4 for comparison of H3 tail spectra, this work), not all overlap in the spectrum of the disordered H3 tail could be resolved and thus only a partial mapping of the changes could be achieved. The third challenge is caused by the blind spot of ssNMR. Regions that have intermediate dynamics are neither observable in J-based nor in dipolar-based spectra. For binding to the H3 tail, this is particularly relevant, since binding can be expected to reduce the flexibility and thus may push the system into the unobservable intermediate regime, as is likely the

case for the PHD2 finger. In addition, like in solution NMR, line broadening due to exchange or proton density can limit the observation of regions involved in binding. The exact reason for the apparent rigidification and long life time of the PHD2-H3 tail complex in the nucleosomal context remains to be established. A putative dual interaction of PHD2 to both the tail and the nucleosomal DNA can be excluded as PHD2 does not appear to interact with DNA (Yi and Kutateladze, personal communication).

In conclusion, we examined here the general applicability of the co-sedimentation method for nucleosome binding studies. The sedimentation procedure results in robust and reproducible samples, in which nucleosomes are densely packed in a disordered arrangement. This suggests that there are no specific nucleosome-nucleosome interactions which would oppose co-sedimentation with a binding partner. As demonstrated here for the CHD4 PHD2 domain, co-sedimentation can successfully be carried out in high salt conditions with weakly binding proteins without compromising spectral quality. Even if the use of ssNMR does not guarantee observation of the bound state, our sedimentation method can be used to detect the binding site indirectly in a nonetheless quasi residue-specific way.

Materials and methods

Sample production

Recombinant *Drosophila melanogaster* histones, histone octamers, 167bp 601 DNA and mononucleosome were produced according to Dyer et al[47]. With the modifications described in chapter 3[38], dialysed and subsequently stored in PK buffer (10 mM (K)PO_4 pH 6.5 supplemented with 10 mM KCl). The sediment sample (sample 2 in Table 1) was used and characterized in chapter 3. Preparation of samples 1 to 3 are described in chapter 3, sample 4 is described in chapter 4.

The PHD2 finger domain from CHD4 was produced as described in Musselman et al[28], then dialyzed to PK buffer and lyophilized.

Solution state NMR experiments

All solution state NMR experiments were performed on a Bruker 21.1 T magnet (operating at 900 MHz ^{1}H Larmor frequency) equipped with an Avance III console and a CPTCI probe, at a temperature of 298K.

A 36 μM nucleosome sample containing 10% of D_2O-based PK buffer, with 0.01% NaN2 and 0.1X Roche cOmplete EDTA-free protease inhibitors was loaded in a 3mM tube. Salt addition was performed by diluting the nucleosome sample 1:1 with the above buffer supplemented with 100 and 150 mM of KCl (from a 4 M stock), successively and concentrating back to the original volume.

Lyophilized PHD2 finger was dissolved in 10% D_2O, containing 0.01% NaN_2 and 0.1X Roche cOmplete EDTA-free protease inhibitors and if relevant 100mM KCl, and maintained on ice.

2D ^{15}N-^{1}H TROSY HSQC spectra were recorded throughout the titration.

Solid state NMR

1.9mg of high salt nucleosome was mixed with a 10:1 molar ratio of high salt PHD2 (corresponding to a 20:1 molar ratio to H3 tail) and incubated in the sedimentation device for 10 minutes, followed by the addition of a final concentration of 2 mM

MgCl2 and sedimentation for 24 hours. The solid-state experiments were performed in a Bruker 18.8 T magnet (800 MHz Larmor frequency) equipped with 1.3 mm ^{1}H/X/Y triple-resonance MAS probe at 50 KHz MAS. The 2D J-based and CP-based spectra were recorded identically to those described in Chapter 4.

SAXS

A solution of 6 μM mononucleosome in PK buffer, and the nucleosome sediment in open air, were loaded in 2mm quartz capillaries (Hilgenberg GMBH) sealed with wax. The SAXS measurements were made on a Ganeshas 300XL system for 14400s. Data analysis was made using the ATSAS suite[48]. Backgrounds were PK buffer and an empty section of the capillary for soluble nucleosome and sediment samples, respectively.

Scanning Electron Microscopy

A portion of the sediment was cut off with a scalpel blade and either mounted in a sandwich of cupper grids or taped on the sample holder. Microscopy was performed in a Thermofisher Scientific Aquilos dual-beam SEM. Imaging was attempted with acceleration voltages between 0.8 and 5 kV using both backscattered and secondary detectors.

References

1 Fasci, D., van Ingen, H., Scheltema, R. A. & Heck, A. J. R. Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei. *Mol Cell Proteomics* **17**, 2018-2033, doi:10.1074/mcp.RA118.000924 (2018).

2 Shi, X. *et al.* Structure and Dynamics in the Nucleosome Revealed by Solid-State NMR. *Angew Chem Int Ed Engl* **57**, 9734-9738, doi:10.1002/anie.201804707 (2018).

3 Gao, M. *et al.* Histone H3 and H4 N-terminal tails in nucleosome arrays at cellular concentrations probed by magic angle spinning NMR spectroscopy. *J Am Chem Soc* **135**, 15278-15281, doi:10.1021/ja407526s (2013).

4 Mainz, A. *et al.* The chaperone alphaB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nat Struct Mol Biol* **22**, 898-905, doi:10.1038/nsmb.3108 (2015).

5 Bertini, I. *et al.* Solid-state NMR of proteins sedimented by ultracentrifugation. *Proc Natl Acad Sci U S A* **108**, 10396-10399, doi:10.1073/pnas.1103854108 (2011).

6 Gardiennet, C. *et al.* Solid-state NMR chemical-shift perturbations indicate domain reorientation of the DnaG primase in the primosome of Helicobacter pylori. *J Biomol NMR* **64**, 189-195, doi:10.1007/s10858-016-0018-0 (2016).

7 Gilbert, N. & Allan, J. Distinctive higher-order chromatin structure at mammalian centromeres. *Proc Natl Acad Sci U S A* **98**, 11949-11954, doi:10.1073/pnas.211322798 (2001).

8 Hsieh, T. H. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108-119, doi:10.1016/j.cell.2015.05.048 (2015).

9 Ricci, M. A., Manzo, C., Garcia-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145-1158, doi:10.1016/j.cell.2015.01.054 (2015).

10 Maeshima, K., Imai, R., Tamura, S. & Nozaki, T. Chromatin as dynamic 10-nm fibers. *Chromosoma* **123**, 225-237, doi:10.1007/s00412-014-0460-2 (2014).

11 Raspaud, E., Chaperon, I., Leforestier, A. & Livolant, F. Spermine-induced aggregation of DNA, nucleosome, and chromatin. *Biophys J* **77**, 1547-1555, doi:10.1016/S0006-3495(99)77002-5 (1999).

12 de Frutos, M., Raspaud, E., Leforestier, A. & Livolant, F. Aggregation of Nucleosomes by Divalent Cations. *Biophysical Journal* **81**, 1127-1132, doi:10.1016/s0006-3495(01)75769-4 (2001).

13 Robinson, P. J., Fairall, L., Huynh, V. A. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A* **103**, 6506-6511, doi:10.1073/pnas.0601212103 (2006).

14 Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138-141, doi:10.1038/nature03686 (2005).

15 Song, F. *et al.* Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380, doi:10.1126/science.1251413 (2014).

16 Kaczmarczyk, A. *et al.* Single-molecule force spectroscopy on histone H4 tail-cross-linked chromatin reveals fiber folding. *J Biol Chem* **292**, 17506-17513, doi:10.1074/jbc.M117.791830 (2017).

17 Eltsov, M. *et al.* Nucleosome conformational variability in solution and in interphase nuclei evidenced by cryo-electron microscopy of vitreous sections. *Nucleic Acids Res* **46**, 9189-9200, doi:10.1093/nar/gky670 (2018).

18 Bertin, A., Mangenot, S., Renouard, M., Durand, D. & Livolant, F. Structure and phase diagram of nucleosome core particles aggregated by multivalent cations. *Biophys J* **93**, 3652-3663, doi:10.1529/biophysj.107.108365 (2007).
19 Livolant, F. *et al.* Are liquid crystalline properties of nucleosomes involved in chromosome structure and dynamics? *Philos Trans A Math Phys Eng Sci* **364**, 2615-2633, doi:10.1098/rsta.2006.1843 (2006).
20 Mangenot, S., Leforestier, A., Durand, D. & Livolant, F. Phase diagram of nucleosome core particles. *J Mol Biol* **333**, 907-916, doi:10.1016/j.jmb.2003.09.015 (2003).
21 Mangenot, S., Leforestier, A., Durand, D. & Livolant, F. X-Ray Diffraction Characterization of the Dense Phases Formed by Nucleosome Core Particles. *Biophysical Journal* **84**, 2570-2584, doi:10.1016/s0006-3495(03)75062-0 (2003).
22 Leforestier, A. & Livolant, F. Liquid crystalline ordering of nucleosome core particles under macromolecular crowding conditions: evidence for a discotic columnar hexagonal phase. *Biophysical Journal* **73**, 1771-1776, doi:10.1016/s0006-3495(97)78207-9 (1997).
23 Berezhnoy, N. V. *et al.* The Influence of Ionic Environment and Histone Tails on Columnar Order of Nucleosome Core Particles. *Biophys J* **110**, 1720-1731, doi:10.1016/j.bpj.2016.03.016 (2016).
24 Allahverdi, A. *et al.* The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res* **39**, 1680-1691, doi:10.1093/nar/gkq900 (2011).
25 Korolev, N., Allahverdi, A., Lyubartsev, A. P. & Nordenskiöld, L. The polyelectrolyte properties of chromatin. *Soft Matter* **8**, doi:10.1039/c2sm25662b (2012).
26 Allahverdi, A., Chen, Q., Korolev, N. & Nordenskiold, L. Chromatin compaction under mixed salt conditions: opposite effects of sodium and potassium ions on nucleosome array folding. *Sci Rep* **5**, 8512, doi:10.1038/srep08512 (2015).
27 O'Shaughnessy, A. & Hendrich, B. CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. *Biochem Soc Trans* **41**, 777-782, doi:10.1042/BST20130027 (2013).
28 Musselman, C. A. *et al.* Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* **423**, 179-187, doi:10.1042/BJ20090870 (2009).
29 Mansfield, R. E. *et al.* Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem* **286**, 11779-11791, doi:10.1074/jbc.M110.208207 (2011).
30 Gatchalian, J. *et al.* Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nat Commun* **8**, 1489, doi:10.1038/s41467-017-01598-x (2017).
31 Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **7**, doi:10.7554/eLife.31481 (2018).
32 Bertini, I. *et al.* On the use of ultracentrifugal devices for sedimented solute NMR. *J Biomol NMR* **54**, 123-127, doi:10.1007/s10858-012-9657-y (2012).
33 van Vugt, J. J. *et al.* Multiple aspects of ATP-dependent nucleosome translocation by RSC and Mi-2 are directed by the underlying DNA sequence. *PLoS One* **4**, e6345, doi:10.1371/journal.pone.0006345 (2009).
34 Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* **403**, 1-10, doi:10.1016/j.jmb.2010.08.039 (2010).
35 Clapier, C. R. *et al.* Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **71**, 1-7, doi:10.1002/prot.21720 (2008).

36 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).
37 Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *Journal of Molecular Biology* **319**, 1097-1113, doi:10.1016/s0022-2836(02)00386-8 (2002).
38 Xiang, S. *et al.* Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **57**, 4571-4575, doi:10.1002/anie.201713158 (2018).
39 Materese, C. K., Savelyev, A. & Papoian, G. A. Counterion atmosphere and hydration patterns near a nucleosome core particle. *J Am Chem Soc* **131**, 15005-15013, doi:10.1021/ja905376q (2009).
40 Stutzer, A. *et al.* 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, doi:10.1016/j.molcel.2015.12.015 (2016).
41 Huang, Y. C. *et al.* The effect of linker DNA on the structure and interaction of nucleosome core particles. *Soft Matter* **14**, 9096-9106, doi:10.1039/c8sm00998h (2018).
42 Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* **42**, 342-346, doi:10.1107/S0021889809000338 (2009).
43 Fragai, M., Luchinat, C., Parigi, G. & Ravera, E. Practical considerations over spectral quality in solid state NMR spectroscopy of soluble proteins. *J Biomol NMR* **57**, 155-166, doi:10.1007/s10858-013-9776-0 (2013).
44 Gebala, M., Johnson, S., Narlikar, G. & Herschlag, D. Ion counting demonstrates a high electrostatic potential of the nucleosome. *bioRxiv*, 514471, doi:10.1101/514471 (2019).
45 Elbayed, K., Dillmann, B., Raya, J., Piotto, M. & Engelke, F. Field modulation effects induced by sample spinning: application to high-resolution magic angle spinning NMR. *J Magn Reson* **174**, 2-26, doi:10.1016/j.jmr.2004.11.017 (2005).
46 Zhou, B. R. *et al.* Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J Mol Biol* **421**, 30-37, doi:10.1016/j.jmb.2012.04.032 (2012).
47 Dyer, P. N. E., Raji S.; White, Cyndy L; Bao, Yunhe; Chakravarthy, Srinivas; Muthurajan, Uma M.; Luger, Karolin;. Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Methods in Enzymology* **Volume 375**, 23-44, doi:10.1016/S0076-6879(03)75002-2 (2003).
48 Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225, doi:10.1107/S1600576717007786 (2017).

6

Chapter 6

General discussion and perspectives

The aim of this work was the development of new methods and approaches to facilitate the characterization of nucleosome-protein interactions by NMR spectroscopy. Magnetic resonance has a strong fascination potential for me due to its unique "insider's" perspective. It allows to observe and study events on a very wide variety of scales, from the infinitely small in classical NMR setups, to the scale of whole organisms via MRI, passing through multiple intermediate scales and time-scales in-between. All rely on that same basic principle that a magnetic field gives a given type of atom a specific frequency, and push the concept to sometimes very elegant and/or complex tricks to exploit, through synergistic use of pulse sequences, hardware and sample preparation, the basic principle to its maximum.

Getting more probes through sample preparation: the wonders hidden in small molecules

Chapter 2 describes the surprise that revealed itself while preparing a new methyl-TROSY site-specific probe. This small molecule, with high symmetry, revealed an unexpected, deceptive spectrum. Its analysis revealed the spectrum to be both deceptively simple and deceptively complicated. All effects seen in this spectrum were only due to scalar couplings within that highly symmetrical system, where all ^{1}H and ^{13}C were chemically equivalent. Because of a sudden wild idea, we applied band-selective decoupling on one of the two triplets (Figure 1A, B). For chemically inequivalent spins, selective decoupling on one spin allows to remove its scalar influence, and thus only observe the couplings between the other unaffected spins. In this system, we showed that this effect also holds true for the different spin states of chemically equivalent spins. As the protons 1-bond coupled to the alpha-state carbon were decoupled, the 4-bound proton-proton coupling due to the inequivalence between alpha and beta-state carbon-bound protons was suppressed and brought the system back to magnetic equivalence.

While the selective decoupling trick resulted in a very sharp line, it is unlikely to result in further line sharpening in large-scale systems, as observed with TROSY. In fact, addition of glycerol already increased the intrinsic line width so much that the 3-bond

multiplet could not be resolved (Figure 1C, D). A compound with one single methyl being labelled and the other two deuterated would probably be a better option for methyl-TROSY studies of epigenetic methyllysine interactions.

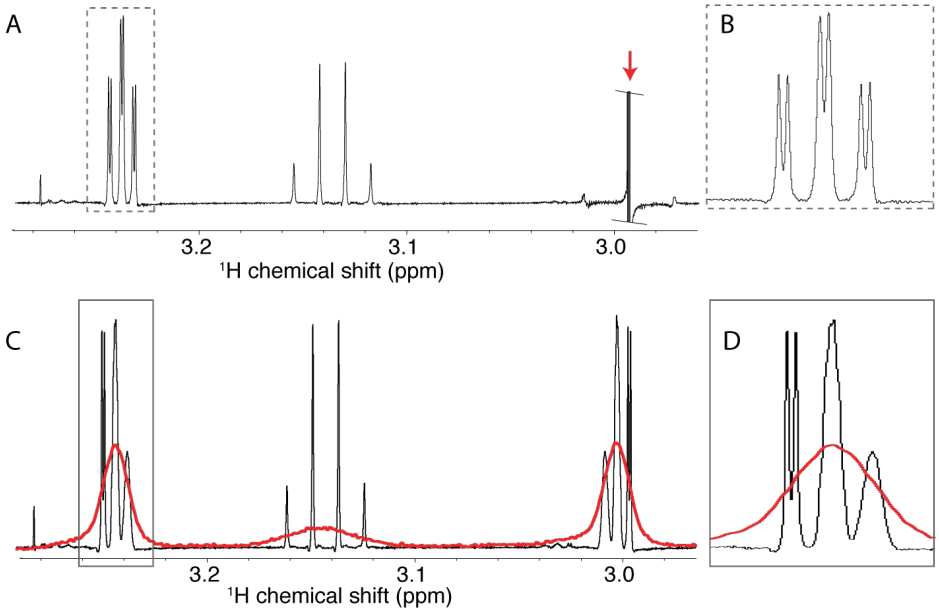


Figure 1. Magnetic inequivalence of an isotope-labelled trimethyl lysine analogue. (A). Homonuclear decoupling on the upfield component (red arrow) restores magnetic equivalence. (B). enlargement of the region highlighted in (A). (C). Overlay of the 1D ^{1}H spectra of the isotope-labelled methyl lysine analogue dissolved in deuterated water (black) and glycerol (red). (D) Enlargement of the region highlighted in (C).

Getting more probes through hardware

The state-of-the-art technique of methyl-TROSY yields very high-resolution spectra, with good sensitivity, and already allowed mapping, structural and dynamic characterization of the nucleosome and nucleosome-protein interactions[1-3]. This approach allows the use of a broad variety of quantitative techniques and pulse sequences to answer biological questions while additionally bringing high sensitivity. It however suffers from the requirement of extensive deuteration from all the actors of the experiment to avoid facing the loss of signal upon interaction. The use of the

very specific ILV-labelling also reduces the number of observables. While this strongly reduces overlap, which can be critical in the case of large molecules and thus aids specificity, it may cause a lack of probes in the region of interest, hindering proper characterization.

In chapters 3 to 5, we aimed to develop, characterize and use a solid-state NMR approach. Thanks to very-fast-Magic-Angle-Spinning, the need for deuteration was strongly reduced, allowing to use only fractional deuteration to get high-resolution spectra. We showed that the nucleosome in our preparations were well-folded through chemical shift analysis. Notably, DNA sequence-specific effects could be observed in our spectra that matched precisely with expectations from the crystal structure. Then, in a proof-of-concept experiment, we obtained experimental data on the binding interface of the LANA 2-22 peptide with the acidic patch, viewed from H2A, and could use it to generate a good quality model using HADDOCK.

Characterizing the method: What did we really see?

However, a few questions subsisted. Nucleosomes have been shown to stack together through an H4 tail to acidic patch interaction[4-6]. Additionally, in our setup, numerous nucleosomes were forced in a tight packing. Was our "apo" form the spectrum of an acidic patch bound to H4 tail – which would then be replaced by the strong binder that is LANA – or of a free nucleosome? To assess that we resorted to SAXS, which showed us that no strong, regular structure was formed. Together with the observation that the signals from the surface of H2A were not split, or particularly broader than residues from within the core, we showed that the nucleosome remained isolated throughout the preparative sedimentation and measurements. Additionally, over a month of magic-angle-spinning did not induce sample denaturation and showed limited water loss.

LANA, being a small peptide with high affinity for the nucleosome, represented a sort of ideal case for our technique. Similarly, it was part of the first nucleosome-bound crystal structure[7], four years before the first co-crystallized nucleosome-protein interaction[8]. The second question raised was thus how far we could push our approach

in the aim of answering biological questions. We thus attempted to use our technique on the study of two different, more challenging interactions.

Pushing the limits of the method: the PHD2 case

The PHD2 domain of CHD4, a low-affinity unmodified-H3-tail-binder, was chosen to challenge the range affinities required for successful co-sedimentation and the range of binding epitopes for successful observation. That interaction had already been thoroughly characterized through NMR and biochemistry, using labelled PHD2 finger and either unlabeled peptides or nucleosome[9-13]. An NMR structure of the PHD2 finger bound to an H3-tail peptide had been already solved, and the dual salt dependence of the interaction was already investigated. Indeed, ionic strength is detrimental to the interaction of PHD2 with H3 tail peptides yet is essential in nucleosomal context to allow tail accessibility. This latter effect would be akin to charge-altering post-translational modifications that hinder the interactions between H3 tail and the DNA[9,10].

Using this model, we could establish that the co-sediment still reflects the solution-state case. Both sediment and solution-state titration counterintuitively showed the interaction to be in slow exchange while with low binding affinity. This suggests that either both association and dissociation kinetics are slow, or that the apparent affinity is reduced by competition for the H3 tail, or a combination thereof. However, contrary to the case of LANA, the bound state of the histone could not be observed. This could be related to a strong contrast of the mode of interaction: the interaction with LANA mostly involves contacts with the sidechains of the acidic patch on the nucleosome core. On the other hand, the PHD2 finger mainly binds to H3 tail through contacts directly involving the backbone of the tail (Figure 2)[7,12]. The resulting high ^{1}H density in direct proximity of the NH group, as well as partial rigidification, could explain the signal loss in the J-based spectrum and lack of new signal in the dipolar-based spectrum. Deuteration of PHD2 could potentially prove helpful for further experiments.

In CHD4, there are two PHD fingers that are interconnected by a flexible acidic linker. The linker has been shown to strongly increase the affinity of the PHD fingers to the tail in nucleosomal context[11]. There, both fingers bind to essentially the same region of H3 tail independently and with high affinity. As the interaction had been only studied from the perspective of the labelled PHDs-linker, additional insight from the tail in this high affinity context would be an interesting follow-up.

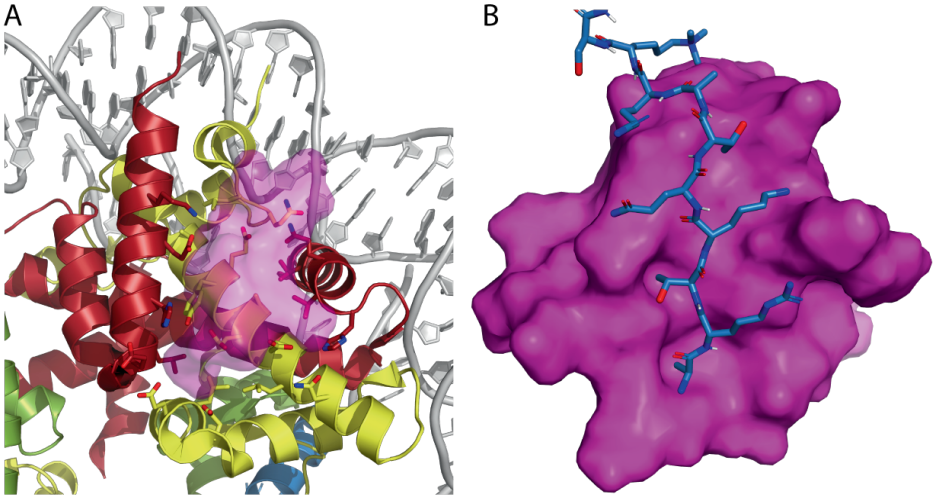


Figure 2. Comparison of LANA and CHD4-PHD2 interactions with the nucleosome. A. LANA (pink surface) contacts the acidic patch mostly via histone-sidechain-based contacts (PDBid: 1ZLA). B. PHD2 finger (pink surface) binds the H3 tail N-terminus (blue) via both sidechain and backbone contacts. (PDBid : 2L75)

A case study: the complex interaction of Dot1p with the nucleosome

To best judge the potential of a technique, one needs to see it in action. The investigation of the interaction between Dot1p and the nucleosome is described in Chapter 4. After careful optimization of binding conditions, Dot1p and nucleosome were successfully co-sedimented. We unfortunately could not detect binding of Dot1p to the nucleosome surface, indicating the formation of a poised rather than an active complex. However, a significant perturbation of the H3 tail NMR signals was observed, suggesting an either direct or indirect role in the binding. Through further

biochemical and NMR experiments, and backed by the wealth of information already published on both Dot1p and Dot1L, we elaborated a model for Dot1p interaction with the nucleosome.

Two key distinctions between the yeast and human homologues are the position of the lysine-rich DNA binding sequence, and the level of dependency on ubiquitin for activity (Figure 3A). Since Dot1p harbors its high-affinity DNA-binding sequence quasi-directly connected to the catalytic core, a direct interaction with the DNA would strongly favor the formation of non-productive complexes. This, alongside low-salt EMSA results, suggested that DNA binding occurs as a later event during complex formation with Dot1p. On the other hand, Dot1L most likely uses its DNA binding sequence, bound to a long linker, as a preliminary DNA-fishing hook. Similarly, the full length Dot1p contains a second lysine rich domain, absent in our construct, shown to bind both DNA and ubiquitin, positioned at a distance from the catalytic core[14]. One could postulate that the latter domain would serve as our termed "fishing hook", thus pre-positioning Dot1p near the nucleosome via ubiquitin or the DNA and allowing a position pre-scanning similar to that of Dot1L (Figure 3B). In this model, both enzymes would then moor on a pivot point from which they can more finely scan the nucleosome surface for the proper active position. While Dot1L uses the coupled acidic patch and ubiquitin interactions as pivot, Dot1p would then use its second DNA-binding sequence as pivot.

During the assignment of H3, we observed an overall drop in signal intensity for the region of the K79 loop, causing its assignment to be tentative. Two residues are even completely missing. This hinted an increased flexibility compared to the rest of H3. Notably, K79 methylation requires a large conformational change for which Worden et al[15] could not conclude whether it is binding-induced or the result of conformation selection, i.e. binding to a transient open conformation. We concluded that the either open or partly open K79 loop transient conformation would be immobilized and potentially enhanced by Dot1. Successive methylation renders the lysine sidechain increasingly hydrophobic, meaning that H3K79 would potentially lodge itself closer and tighter to the hydrophobic pocket on the surface of H4 underneath it[16]. In this

reasoning, the degressive processivity of Dot1 would at least partly arise from an increased unfavourability of the H3K79 loop open conformation. Dot1 would thus be required to withhold its position on the nucleosome long enough to capture that lesser transient state.

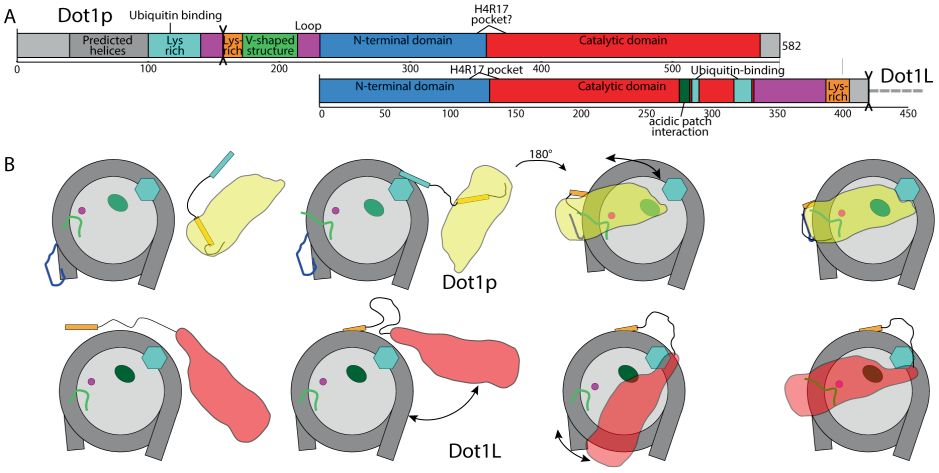


Figure 3. Comparison between Dot1p and Dot1L. (A). Domain architecture of Dot1p and Dot1L. unstructured loops are displayed in pink and termini in grey. (B). Putative model for full-length Dot1p interaction with the nucleosome. The folded core of Dot1p and dot1L are displayed in light yellow and red, respectively. Ubiquitin is displayed as cyan hexagon, the acidic patch as a green oval, K79 as a pink dot and the H3 and H4 tails as blue and light green lines, respectively.

General conclusions

As one of the most abundant biomolecular assemblies in the cell, the nucleosome is the key platform for an intricate network of interactions that control cell fate. Because of the wide scope of proteins that interact through various binding modes to the nucleosome, it is paramount to have matching set of experimental techniques to study these and dissect the underlying molecular mechanisms. Making use of several recent advances in the field, we set up a new approach for the determination of nucleosome-binding epitopes by ^{1}H-detected solid-state NMR of co-sedimented nucleosome-

protein complexes. It both poses significantly lower requirements in terms of deuteration and brings more observables than the current state-of-the-art methyl-TROSY approach. While the sedimentation approach is relatively time consuming in the assignment of NMR signals and does not readily allow titration experiments, it allows straightforward extraction of protein secondary structure directly from chemical shifts. This makes it particularly useful for the determination of the conformation and dynamics of nucleosome-bound proteins or histone variants. Coupled to protein folding programs like Rosetta and possibly including sparse NMR or other biophysical data, 3D structural models can in principle be obtained.

As a spectroscopic technique, characterization of binding events or protein structure and dynamics by NMR is intrinsically indirect. The strength of NMR relies on its intrinsic atomic resolution and its sensitivity to the slightest perturbations arising from lowly-populated, dynamic or low-affinity complexes. As also illustrated in this work, the combined use of NMR and other structural and biochemical techniques allows to advance our understanding. Adding up the different points of view provided by each technique on one complicated problematique is the way to better grasp its complexity.

References

1 Kitevski-LeBlanc, J. *et al.* The RNF168 paralog RNF169 defines a new class of ubiquitylated histone reader involved in the response to DNA damage. *Elife* **6**, doi:10.7554/eLife.23872 (2017).

2 Kato, H. *et al.* Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A* **108**, 12283-12288, doi:10.1073/pnas.1105848108 (2011).

3 Horn, V. *et al.* Structural basis of specific H2A K13/K15 ubiquitination by RNF168. *Nat Commun* **10**, 1751, doi:10.1038/s41467-019-09756-z (2019).

4 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

5 Song, F. *et al.* Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376-380, doi:10.1126/science.1251413 (2014).

6 Mangenot, S., Leforestier, A., Durand, D. & Livolant, F. Phase diagram of nucleosome core particles. *J Mol Biol* **333**, 907-916, doi:10.1016/j.jmb.2003.09.015 (2003).

7 Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**, 856-861, doi:10.1126/science.1120541 (2006).

8 Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562-566, doi:10.1038/nature09321 (2010).

9 Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **7**, doi:10.7554/eLife.31481 (2018).

10 Gatchalian, J. *et al.* Accessibility of the histone H3 tail in the nucleosome for binding of paired readers. *Nat Commun* **8**, 1489, doi:10.1038/s41467-017-01598-x (2017).

11 Musselman, C. A. *et al.* Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc Natl Acad Sci U S A* **109**, 787-792, doi:10.1073/pnas.1113655109 (2012).

12 Mansfield, R. E. *et al.* Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem* **286**, 11779-11791, doi:10.1074/jbc.M110.208207 (2011).

13 Musselman, C. A. *et al.* Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* **423**, 179-187, doi:10.1042/BJ20090870 (2009).

14 Oh, S., Jeong, K., Kim, H., Kwon, C. S. & Lee, D. A lysine-rich region in Dot1p is crucial for direct interaction with H2B ubiquitylation and high level methylation of H3K79. *Biochem Biophys Res Commun* **399**, 512-517, doi:10.1016/j.bbrc.2010.07.100 (2010).

15 Worden, E. J., Hoffmann, N. A., Hicks, C. W. & Wolberger, C. Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**, 1490-1501 e1412, doi:10.1016/j.cell.2019.02.002 (2019).

16 Lu, X. *et al.* The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol* **15**, 1122-1124, doi:10.1038/nsmb.1489 (2008).

English summary

The way a cell controls which part of its long DNA to access and read, which part not, and the way it maintains the integrity of the genome, are crucial for the proper functioning of an organism. These processes depend to a large extent on the packing of the DNA in nucleosomes that, through their presence or absence and types of epigenetic modifications they carry, allow genome maintenance and regulation. Conjointly, malfunctions in these mechanisms are detrimental to the cell and have been frequently correlated with cancer. A structural understanding of how these mechanisms operate on nucleosomes to ensure the proper functioning of a cell is thus crucial.

Recent efforts and developments in structural biology allowed an increasing level of understanding on the molecular basis of nucleosome function. In this work, we attempted to extend the toolbox of Nuclear Magnetic Resonance (NMR)-based approaches for the study of the nucleosome, with the long-term aim of helping with the biological understanding of chromatin function and epigenetics. The first part of **Chapter 1** reviews key aspects of nucleosome structure and function. The second part of **Chapter 1** introduces in layman terms the basic principles of NMR, a technique that offers structural information in a way that is fundamentally different from other prominent techniques as are crystallography and cryo-electron microscopy (cryo-EM). In **Chapter 2**, the remarkable 1D spectrum of a small molecule, designed to mimic a trimethyl lysine for methyl-TROSY NMR studies, is described in detail. The protons in its symmetrical and isotope-labelled trimethylammonium group are chemically equivalent, yet magnetically inequivalent due to the presence of $^{3}J_{CH}$ couplings. The small $^{4}J_{HH}$ couplings that are now active would normally perturb the spectral pattern in very complex way, but here the system remained effectively weakly coupled due to the large $^{1}J_{CH}$ coupling, resulting in a simple multiplet pattern. Subspectral analysis allowed us to intuitively describe and build up the deceptively complicated spectrum.

Additionally, we showed that the system can be returned to magnetic equivalence by selective homonuclear decoupling.

In **Chapters 3 to 5**, we developed, characterized and used a new approach for the study of the nucleosome making use of the well-established techniques of sedimentation coupled to ^{1}H-detected solid-state NMR.

Chapter 3 describes the proof-of-principle of the approach. High quality spectra were obtained for histones H2A and H3. H2A was then assigned and we de novo determined its secondary structure within the nucleosome, which was in agreement with previously reported crystal structures, confirming sample integrity. Co-sedimentation of a nucleosome-binding peptide allowed the observation of clear perturbations of specific H2A signals, allowing the determination of the binding site. Subsequent data-driven docking resulted in a structural model of the complex that was in high agreement with a previously reported co-crystal structure.

In **Chapter 4,** we applied our new approach to the study of Dot1p, the yeast H3K79 methyltransferase. To do so, histone H3 was assigned and characterized similarly to H2A in Chapter 3. Again, the secondary structure of H3 in the sediment was in agreement with the crystal structures. Through a divide-and-conquer approach and the combined use of biochemical and NMR methods, we found that the lysine-rich sequence directly N-terminal of the Dot1p core has a high affinity towards DNA. This suggested a nucleosome-anchoring role for this domain, mooring the enzyme in a close vicinity from H3K79. Additionally, the catalytic core of Dot1p has very low affinity for ubiquitin, in agreement with the low ubiquitin-dependency of K79 methylase activity in vivo compared to the human variant.

Chapter 5 describes the characterization of the nucleosome sediment as well as a case-study on the limitations of its use in the study of protein-nucleosome interactions. Using SAXS with NMR data from Chapter 3, we could conclude that the sediment, while highly dense, is devoid of long-range ordering of nucleosomes. We then attempted to co-sediment a low affinity, H3 tail-binding protein with the nucleosome, challenging the formation of a rigid, saturated protein nucleosome complex. While specific binding could be detected through the disappearance of the free state signals,

the bound state was unobservable. Using solution- and solid-state NMR, we showed that the PHD2 domain of CHD4 counterintuitively combines low affinity interaction and a slow exchange interaction.

In **Chapter 6**, the results from the thesis are further discussed and put in perspective along with suggestions for future works and prospects.

Nederlandse Samenvatting

Voor de goede werking van een organisme is het cruciaal dat elke cel nauwkeurig regelt welk deel van zijn lange DNA wordt afgelezen, en welk deel niet en tegelijkertijd zorg draagt dat DNA schade op tijd wordt hersteld. Deze processen zijn in grote mate alleen mogelijk doordat het DNA in nucleosomen verpakt is. Nucleosomen vormen een beschermende laag om het DNA die door hun aan- of afwezigheid en de epigenetische modificaties ze dragen, de instandhouding en regulatie van het genoom mogelijk maken. Storingen in deze mechanismen zijn zeer schadelijk voor de cel en worden vaak met kanker gecorreleerd. Een structureel begrip van de mechanismen die ingrijpen op nucleosomen is dus cruciaal.

Recente inspanningen en ontwikkelingen in de structurele biologie hebben voor een toenemend inzicht in de moleculaire basis van nucleosoom functie gezorgd. In dit proefschrift hebben we geprobeerd de toolbox van Nuclear Magnetic Resonance (NMR) voor de studie van het nucleosoom uit te breiden, met als lange termijn doel bij te dragen aan het biologische begrip chromatine functie en epigenetica. Het eerste deel van Hoofdstuk 1 bespreekt de belangrijkste aspecten van de structuur en functie van nucleosomen. Het tweede deel introduceert in leken-termen de basis-principes van NMR, een techniek die structurele informatie biedt die fundamenteel verschilt van andere prominente technieken zoals kristallografie en cryo-elektronenmicroscopie (cryo-EM).

In Hoofdstuk 2 wordt een het verrassend eenvoudige en tegelijkertijd verrassend gecompliceerde 1D spectrum van een kleine molecuul beschreven. Dit molecuul, gebaseerd een trimethyllysine, gaf een zeer ongebruikelijk spectraal patroon dat geheel door aan scalaire koppelingen is veroorzaakt. De protonen van de symmetrische en isotoop-gelabelde trimethylammoniumgroep zijn chemisch equivalent, maar magnetisch inequivalent vanwege de labeling geintroduceerde $^{3}J_{CH}$-koppelingen. De kleine, maar nu actieve, $^{4}J_{HH}$ koppelingen compliceren de multiplet-struktuur normaal gesproken, maar door de grote $^{1}J_{CH}$ koppeling blijft het systeem effectief echter in een

zwak koppelingsregime, wat resulteert in een onverwacht eenvoudig spectrum. Met subspectrale analyse is dit bedrieglijk gecompliceerde spectrum intuitief te beschrijven. Daarnaast lieten we zien hoe door selectieve homonucleaire ontkoppeling het systeem weer magnetisch equivalent is te maken.

Hoofdstukken 3 tot en met 5 beschrijven de ontwikkeling, gebruik en karaterisatie van een nieuwe benadering die we ontwikkeld hebben om de interactie tussen eiwitten en nucleosomen te bestuderen. Deze methode is gebaseerd op sedimentatie van nucleosomen en nucleosoom-eiwit complexen, gekoppeld aan ^{1}H-gedetecteerde vaste-stof NMR.

Hoofdstuk 3 beschrijft het proof-of-principle experiment van deze benadering. Sedimentatie van mononucleosomen die werden gereconstitueerd met één fractioneel-gedeutereerd en isotoop-gelabeld histon leverde ^{1}H-gedetecteerde spectra van hoge kwaliteit. De NMR spectra van histon H2A konden vervolgens worden toegekend en de secundaire structuur van H2A bepaald. Deze was in overeenstemming met de structuur gezien in kristalstructuren, wat de integriteit van het monster bevestigde. Spectra opgenomen na co-sedimentatie van een peptide dat sterk aan nucleosomen bindt, lieten duidelijke verstoringen van specifieke H2A-signalen zien, waardoor bepaling van de bindingsplaats van het peptide mogelijk was. Het daarop gebaseerde structureel model van het complex was in hoge overeenstemming met een eerder gerapporteerde co-kristalstructuur.

In Hoofdstuk 4 hebben we onze nieuwe aanpak toegepast op de studie van Dot1p, de H3K79-methyltransferase uit gist. Allereerst werden de NMR spectra van histon H3 toegekend en zijn structuur gekenmerkt op dezelfde manier als voor H2A in hoofdstuk 3. Over het algemeen was de secundaire H3-structuur in overeenstemming met de kristalstructuur. Met een 'verdeel en heers'-benadering en gecombineerd gebruik van biochemische en NMR-methoden, vonden we dat het lysine-rijk N-terminale gebied van Dot1p met hoge affiniteit aan DNA kan binden. Waarschijnlijk fungeert het domein daarmee als een anker om het enzym op korte afstand van H3K79 te lokaliseren. Verder vonden we dat ubiquitine slechts met een zeer lage

affiniteit aan Dot1p bindt, wat overeenkomt met de lagere ubiquitine-afhankelijkheid van Dot1p *in vivo*, in vergelijking met de menselijke homoloog.

Hoofdstuk 5 beschrijft de verdere karakterisering van het nucleosoom-sediment en van de vereisten voor succesvolle co-sedimentatie. Met behulp van SAXS en NMR data konden we concluderen dat het sediment, hoewel zeer dicht, verstoken is van hogere-orde structuren gevormd door specifieke nucleosoom-nucleosoom interacties. We hebben vervolgens onze methode geprobeerd met een eiwit dat met lage affiniteit aan de ongemodificeerde H3-staart bindt. Dit systeem maakte het lastig om een rigide en verzadigd eiwit-nucleosoom-complex te vormen en vormde daarmee een belangrijke test van de algemene bruikbaarheid van de methode. Hoewel specifieke binding kon worden gedetecteerd door het verdwijnen van de signalen van de vrije toestand, was de gebonden toestand niet waarneembaar. Met behulp van vloeistof en vaste-stof NMR toonden we aan dat het PHD2-domein of CHD4 op contra-intuïtieve wijze een interactie met lage affiniteit combineert met een langzame dissociatie van het complex.

In Hoofdstuk 6 worden de resultaten van het proefschrift verder besproken en in perspectief geplaatst, samen met suggesties voor toekomstige werk.

Aknowledgements

This work bears my name, but a lot of people were involved, directly or indirectly, with the process of reaching the state of this printed book.

First of all, I naturally want to thank my supervisor and co-promotor, Dr. Hugo van Ingen.

Hugo, thank you for initially looking over my initial oddity and blunders, choosing to welcome me in your then brand new team. Thank you for the climate you helped installing in the workplaces we went to, for the oversight, for the attentive listening to my ideas, hypotheses and discussions, for patiently correcting and explaining me when they were crazy and for recognizing the few times they weren't. Thank you for your dynamism and for your "It is what it is" way of looking at results and data.

This Work was spread over two laboratories, as Hugo's group migrated to Utrecht. I would thus like to thank Prof.dr. Marcellus Ubbink and Prof.dr. Marc Baldus for giving me the access to their labs and facilities, and thank you to Prof.dr Baldus again for becoming my promotor after the move to Utrecht.

Thank you to the nucleosome group: Velten, senior student that (sometimes painfully) paved the way, his rich humor and culture sometimes hidden under a grumpy face. Ivan, highly concentrated person, whose focused mask could fortunately be cracked open a few times. Clara, the calm(ing) figure now starting the biggest of all experiments. Heyi, the sheer madness mixed with deep philosophy, and Vincenzo, the new hope, sanely mad, soon to take over the group once the above are gone.

Speaking of lab, I really need to mention the shadow workers, those whom you notice most when they go missing (somehow, all machines are secretly waiting for them to go on holidays in order to synchronously break down, and all stocks are emptied during the same period): the technicians. Anneloes and Monika, "mothers" of the Leiden lab, later joined by Patrick, and complemented by Karthick and Fons for the NMR; Mark, Johan and Andrei in Utrecht. Thanks for the overall effect you had on the daily mood, intrinsically, through managing to keep the lab in working order, and for sharing all those little hints/tricks that one only learns through experience.

Last but definitely not least for the labs: all the colleagues and/or friends from both universities, former and present. It was a nice mix between professionalism and post-work craziness. I enjoyed the various discussions over a cup of coffee or a glass of beer. Thank you for the fodder to my curiosity, the quick answers to my "quick" questions, the quick questions that I got to more or less quickly answer back (sorry for when it turned into lectures), the various (board) games, the drinks, the barbecues, the two road trips, the… Thus, thanks for multiple occasions to get the necessary defocusing from work. Additional thanks go to my collaborators for their involvement in the research.

Good luck with the rest of their studies to the bachelor and master students that worked with me for several months.

Velten (again) and Raffaella, early housemates, thank with the help for adapting to the new country, city and lifestyle. Thank you for the culture shocks that helped opening my field of view, and thank you for having tried cracking open my hard shell of shyness.

Thanks to the friends at the ING (international neighbouring group, not the bank) for those enjoyable regular Wednesday social events.

A big thank you to my family and friends, for the good times and the mental support especially in the hardest moments, notably involving the times when experiments didn't want to work, when data analysis was stalling, or during close-to-deadline writing moments.

And finally, thank you to the various people that, throughout my life, in a way or another eventually led me to end up where I am today. Thank you to the various teachers, professors, friends and acquaintances that kept feeding and encouraging my curiosity.

Curriculum Vitae

Ulric B. le Paige was born on April 14th 1991 in Liège, Belgium. He obtained a bachelor's degree in biology followed by a master's degree in "Biochemistry, Molecular and Cellular Biology" in the Université de Liège. He completed his master's thesis under supervision of Professors Moreno Galleni and Christian Damblon, focusing on the study of the metallo β-lactamase CAU-1 by enzyme kinetics and NMR. In 2015, he joined the group of Dr. Hugo van Ingen in Leiden University, the Netherlands, and later Utrecht University to focus on the study of the nucleosome by NMR. The results from this are presented in this work.

List of publications

- **U. B. le Paige***, B. Smits*, P. 't Hart, F. Lefeber, N. L. Martin & H. van Ingen (2017). Weak Coupling Between Magnetically Inequivalent Spins: The Deceptively Simple, Complicated Spectrum of a (13)C-labeled Trimethylated Amine. *J Magn Reson* 278, 96-103
- S. Xiang*, **U. B. le Paige***, V. Horn, K. Houben, M. Baldus & H. van Ingen (2018). Site-Specific Studies of Nucleosome interactions by Solid-State NMR Spectroscopy. Angew Chem Int Ed Engl. 57(17), 4571-4575
- V. Horn*, M. Uckelman*, H. Zhang, J. Eerland, I. Aarsman, **U. B. le Paige**, C. Davidovitch, T. Sixma & H. van Ingen (2019). Structural Basis of Specific H2A K13/15 Ubiquitination by RNF 168. Nat Commun. 10, 1751
- I. Corbeski*, V. Horn*, R. A. van der Valk, **U. B. le Paige**, R. T. Dame & H. van Ingen (2018). Microscale Thermophoresis Analysis of Chromatin Interactions. Methods Mol Biol. 1837, 177-197.