



Universiteit
Leiden
The Netherlands

Computational, biochemical, and NMR-driven structural studies on histone variant H2A.B

Zhang, H.

Citation

Zhang, H. (2020, August 25). *Computational, biochemical, and NMR-driven structural studies on histone variant H2A.B*. Retrieved from <https://hdl.handle.net/1887/135944>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/135944>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/135944> holds various files of this Leiden University dissertation.

Author: Zhang, H.

Title: Computational, biochemical, and NMR-driven structural studies on histone variant H2A.B

Issue Date: 2020-08-25

Computational, biochemical, and NMR-driven structural studies on histone variant H2A.B

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 25 augustus 2020
klokke 10:00 uur

door
Heyi Zhang
geboren te Wuxi, China
in 1990

**Computational, biochemical, and NMR-driven structural studies
on histone variant H2A.B**

Doctoral thesis, Leiden University, 2020

Cover design: Heyi Zhang

ISBN number: 978-94-91546-17-4

©2020, Heyi Zhang

Printed by Boekdrukken.com, the Netherlands

Promotor:	Prof.dr. M. Ubbink
Co-promotor:	Dr. H. van Ingen
Promotiecommissie:	Prof.dr. H.S. Overkleeft (voorzitter)
	Prof.dr.ir. S.J.T. van Noort (secretaris)
	Dr. R.T. Dame
	Prof.dr. R. Boelens (Universiteit Utrecht)
	Prof.dr. T.K. Sixma (Nederlands Kanker Instituut)

The research described in this thesis was performed in the Macromolecular Biochemistry Department of the Leiden Institute of Chemistry in Leiden, the Netherlands and the NMR Spectroscopy Research Group of the Bijvoet Center for Biomolecular Research in Utrecht, the Netherlands.

To my grandfather (1928-2018)

Table of Contents

ABBREVIATIONS.....	1
CHAPTER 1. GENERAL INTRODUCTION	3
CHAPTER 2. ISOTOPE-LABELING STRATEGIES FOR SOLUTION NMR STUDIES OF MACROMOLECULAR ASSEMBLIES	23
CHAPTER 3. VARIANT H2A.B-H2B HISTONE DIMER IS MORE STABLE THAN THE CANONICAL DIMER	45
CHAPTER 4. STRUCTURE AND DYNAMICS OF THE H2A.B VARIANT NUCLEOSOME.....	77
CHAPTER 5. THE ELECTROSTATIC POTENTIAL OF THE NUCLEOSOME ACIDIC PATCH	109
CHAPTER 6. DISCUSSION	144
SUMMARY	156
SAMENVATTING.....	159
CURRICULUM VITAE	162
PUBLICATIONS.....	162

Abbreviations

AEP	Asparaginyl endopeptidases
AFM	Atomic force microscopy
AP	Acidic patch
ATP	Adenosine triphosphate
Bp	Base pairs
BME	β -mercapto-ethanol
BMRB	Biological magnetic resonance data bank
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP sequencing
CPMG	Carr-Purcell-Meiboom-Gill sequence
CS	Chemical shift
CSI	Chemical shift indices
CSP	Chemical shift perturbation
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPL	Expressed protein ligation
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein
HAC	Acetic acid
HDAC	Histone deacetylase
HSQC	Heteronuclear single quantum coherence spectroscopy
IPTG	Isopropyl β -D-1-thiogalactopyranosid
LB	Lysogeny broth
LEGO	Label, express and generate oligomers
MD	Molecular dynamics
MES	2-ethanesulfonic acid
MNase	Micrococcal nuclease
MQ	Multiple quantum
mRNA	Messenger RNA
NCP	Nucleosome core particle
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser effect spectroscopy

NTD	N-terminal domain
PCR	Polymerase chain reaction
PDB	Protein data bank
Pi	Inorganic phosphate
PMSF	Phenylmethylsulfonyl fluoride
PTM	Post translational modification
PTS	Protein trans-splicing
R2	Transverse relaxation rate
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RDC	Residual dipolar coupling
SAXS	Small angle X-ray scattering
SAIL	Stereo-array isotope libeling
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SHL	Super helix location
SrtA	Sortase A
S/N	Signal to noise ratio
T _m	Melting temperature
TROSY	Transverse relaxation optimized spectroscopy
TSA	Thermal shift assay
tc	Rotational correlation time
wt	Wild type
Xi	Inactive X chromosome

Chapter 1. General introduction

The nucleosome, a DNA-protein complex, is well known today as the basic unit of chromatin. However, the discovery and characterization of it took more than a hundred years of unveiling each building block of the nucleosome step by step. The existence of histones, the protein components of the nucleosome, was first recognized in 1884 by Kossel¹. It was not until the 1960s that the five histone types were observed², now known as H1, H2A, H2B, H3 and H4. DNA, on the other hand, was first isolated by Miescher in 1869, and its double-helix structure was revealed by Watson and Cricks more than eighty years later in 1953³. With the identification and characterization of both protein and DNA components available, studies emerged in the 1970s to understand the fundamental basis of chromatin structure⁴⁻⁷. In 1974, a complex of eight histone molecules and about 200 DNA base pairs was proposed by Kornberg as the repeating unit of the chromatin structure⁸. Finally, twenty-three years later in 1997, the first high resolution crystal structure of the isolated nucleosome core particle was solved by Luger *et al.*⁹ Two copies of each of the four core histones H2A, H2B, H3, and H4 form a histone octamer, around which wraps 147 bp of DNA to form the basic unit of chromatin: the nucleosome (Figure 1.1a). These nucleosomes can be further coiled into thicker chromatin fibers, which are further compacted and folded into the typical X-shaped chromosomes during metaphase. This packaging of DNA is dynamic and modulates nuclear processes. At lowest level, the presence of nucleosomes forms a roadblock, limiting DNA accessibility for polymerases, transcription factors, and other DNA binding proteins. Modulation of nucleosome structure and position allows the manipulation of this roadblock and thus ultimately the regulation of DNA transcription or repair. A vast range of proteins cooperate to achieve this regulation. Even if the resulting chromatin biology is highly complex, just like mosaic art, no matter how splendid it is, the art piece is made of simple building blocks. And just as the color and shapes of these building blocks determine the range of mosaics that can be made, the fundamental molecular properties of the nucleosome and chromatin factors determine the dynamic network of interactions that underlie chromatin biology. In this regard, the determination of the crystal structure of nucleosome in 1997 has truly been a scientific breakthrough discovery (Figure 1.1a)⁹. Since then it has been shown that the architecture and structure of the nucleosome

is nearly independent of DNA sequence and constant among different species. Yet the efficiency of this roadblock can be altered through a number of mechanisms, most notable through incorporation of post translational modifications (PTMs) and the replacement of the canonical histones by histone variant proteins.

PTMs alter nucleosome structure and function

PTMs are covalent modifications that can be dynamically added on and removed from residue side chains as a means to regulate protein function in response to the cell environment. While histone PTMs are mostly known for serving as binding anchors for various chromatin factors, a number of histone PTMs, in particular those that occur in the histone core rather than the tail, act by influencing nucleosome structure directly. Core modifications often alter histone-DNA and/or histone-histone interactions thus changing nucleosome dynamics and stability. For example, acetylation of H3 lysine 56, which is located close to the nucleosomal DNA, interrupts the local histone-DNA interaction, increases DNA accessibility, and ultimately affects gene expression and DNA repair¹⁰⁻¹¹. Similarly, acetylation of H4 lysine 91 removes an essential salt bridge to H2B E68 in the tetramer-dimer interface, resulting in higher sensitivity toward DNA digestion by micrococcal nuclease¹². This modification occurs before the H3-H4 tetramer is assembled on DNA and is thus suggested to regulate nucleosome assembly¹³. These examples illustrate how small structural changes, i. e. addition of a simple acetyl group, can have pronounced effects on overall nucleosome stability and function.

Histone variants alter nucleosome structure and function

The largest structural changes are brought about by substitution of the canonical histones by variant histones with different amino acid sequence. For example, macroH2A, an H2A variant, contains an additional ~250 amino acids at its C-terminus that includes a macro domain and thus offers new functionalities to the nucleosome¹⁴⁻¹⁵.

Unlike canonical histones, whose synthesis and assembly into nucleosomes is highly coupled with DNA replication, histone variants are expressed and incorporated into chromatin in a replication-independent manner. In addition, histone variants are usually encoded by single copy genes whereas the canonical forms are encoded by multiple copy genes. Variants are less abundant but play a vital role in regulating chromatin function by replacing their canonical forms from nucleosome at defined genomic loci. For example, centromeres contain specialized nucleosomes that include histone H3 variant CENP-A. Due to the presence of this variant these nucleosomes have an altered structure and can form interactions with other centromeric proteins. The crystal structure of the centromeric nucleosomes (pdb: 3AN2) showed that it only organizes 121 bp DNA due to the shorter α N helix of the variant ¹⁶ and the resulting higher flexibility of entry/exit nucleosomal DNA may facilitate the binding of centromeric DNA-binding proteins.

The first histone variant, sperm H2B, was identified in the 70s of 20th century ¹⁷. There are now in total 18 groups of histones variants recorded in Histone Database for H2A, H2B, H3 and H4 ¹⁸. Overall, H2A and H3 have the most variants, while H2B and H4 have only a few variants. Sequence differences to the canonical form can be large, such as in the case of macroH2A, but are typically limited to few amino acid substitutions. For example, histone variant H2A.X contains an additional C-terminal motif that contains a serine that is phosphorylated in the response to DNA damage, thereby signaling and recruiting DNA repair proteins to the damage site ¹⁹.

The incorporation of histone variants is mediated by nucleosome remodelers and dedicated histone chaperone proteins. For H2A.Z, a variant involved in transcription regulation and mostly positioned at promoters, it has been shown that Swr1, part of the SWR1 remodeling complex, can load the H2A.Z-H2B dimer in two steps on a tetrasome, the complex of DNA and the H3-H4 tetramer ²⁰. SWR1 is thought to first unpeel the DNA from the nucleosome, using energy from ATP, releasing the canonical H2A-H2B dimers in the process, after which the variant dimer is incorporated.

The structures of several variant nucleosomes have been determined and some have shown surprisingly little structural changes to the canonical form. While a wide array of biochemical and biophysical

data have indicated that H2A.Z forms mobile and dynamic nucleosomes, the crystal structure of the H2A.Z nucleosome (pdb: 1F66) is virtually identical to that of the canonical nucleosome ²¹, highlighting the need for complementary studies that are sensitive to dynamics.

This thesis is focused on H2A.B, which is one of the most divergent histone variants ²². Incorporation of H2A.B into nucleosomes has been shown to induce a more open conformation of the nucleosome in which the DNA is more accessible ²³. Functionally, it has been linked to various processes, including RNA splicing and active transcription ²⁴. To provide a solid fundamental basis to understand its biological function, a study on the structure and dynamics of the H2A.B-H2B dimer and the H2A.B nucleosome is described in this thesis. The remainder of this introduction describes the current knowledge on this enigmatic histone variant.

Discovery of histone variant H2A.B

As human females have two X chromosomes, while males have one X and one Y chromosome, one of the two X chromosomes from female somatic cells is randomly subjected to permanent inactivation to ensure comparable gene expression from the X chromosome ²⁵. This sex-determining system is consistent in most of mammals. Inactivation of one of the X-chromosomes is achieved by heterochromatin formation and further compaction and results in the formation of so-called Barr bodies in the nucleus. While the exact mechanism of inactivation is not fully resolved yet, a crucial role is played by the long non-coding RNA Xist ²⁶. In addition, the inactive X (Xi) chromatin contains specialized nucleosomes with H2A variant macroH2A²⁷. This variant is thought to promote the repression of transcription by amino acid substitutions within the histone fold that stabilize the nucleosome, as well as by promoting recruitment of histone deacetylases (HDACs) via the macrodomain ²⁸. In the hope to find a counterpart to macroH2A that is repelled from Xi, Chadwick and Willard searched the genome database and found a group of distant H2A homologs ^{22,29}. Expression of the identified gene was confirmed by Northern blot and reverse transcription PCR in a variety of cell lines and tissues. Transfection of

a GFP fusion protein into primary human fibroblasts showed that it was excluded from the Xi chromosome in female cells. Further analysis proved that the protein is a nucleosomal core histone ²². The identified protein was originally named H2A.Bbd, for Barr-body deficient H2A variant. The simplified name H2A.B has been slowly adopted during the past two decades. H2A.B contains 114 amino acids and shares only 48% identity to canonical H2A, making it one of the most divergent histone variants up to date.

The sequence of H2A.B differs strongly from canonical H2A

Up to date, fifteen H2A.B sequences have been reported over twelve mammalian species, and their sequence alignment against human canonical H2A is shown in Figure 1.1c ¹⁸. Three key differences between the H2A.B family and canonical H2A emerge from sequence comparisons. First, most sequence differences occur in the N-terminal tail of H2A.B, which is also a highly variable region within the H2A.B family. A common overall feature of the H2A.B N-terminal tail is the lack of lysine residues as opposed to canonical H2A. Histone lysines are common targets for PTMs, such as acetylation, methylation, and ubiquitination, which are essential in defining chromatin function. In the case of H2A, N-terminal K13/15 ubiquitination plays a crucial role in the DNA damage response ³⁰⁻³¹, while N-terminal K5 acetylation is essential in transcription regulation and chromatin remodeling ³². Notably, also the histone fold domain of H2A.B is scarce in lysines, with only one lysine present instead of four in the canonical H2A. Together, this absence of modifiable residues suggests that the variant is regulated in a manner different from other members of H2A family and that it exerts its specific impact on chromatin function independent of PTMs.

The second key difference in H2A.B is its missing C-terminal tail and truncated C-terminal docking domain. Absence of the C-terminal tail again removes a number of well-known modification sites in H2A, in particular K118/K119 sites that are involved in transcriptional repression ³³. In the canonical nucleosome, the H2A docking domain mediates the interaction between H2A-H2B dimer and H3-H4 tetramer

⁹. In particular, the segment absent in H2A.B binds to the H3 α N helix in the canonical nucleosome. Furthermore, the truncated docking domain in H2A.B contains several sequence differences compared to H2A, raising the question whether this will also translate into structural differences in the H2A.B nucleosome.

A third common feature of H2A.B compared to other H2A histones is the reduced negative charge of the acidic patch, a negatively charged area on nucleosome surface formed by six H2A residues (E56, E61, E64, D90, E91, and E92) and two H2B residues (E102 and E110) ⁹. The acidic patch functions as a binding surface for the H4 tail from a neighboring nucleosome, thereby mediating higher order chromatin folding ³⁴. Moreover, the acidic patch is the key binding site for a wide variety of nucleosome binding proteins, such as LANA, RCC1, SIR3, HMGN2, etc. ³⁵⁻³⁸. Just as the H4 tail, these proteins all use a strategically located arginine residue to form hydrogen bonds with acidic patch residues³⁹. In human H2A.B three of the six residues that form acidic patch are either neutralized or charge-swapped: E61 to K, E91 to R, and E92 to L. Within the H2A.B family, several species have also a D90 to N substitution, while mouse H2A.B retains E92 in its sequence shifted by one residue position. The reduced negative charge on H2A.B-nucleosome surface will disfavor or even inhibit binding of the many acidic-patch binding chromatin factors including the H4 tail, thus affecting chromatin function and higher-order structure (see also below).

Last but not least, it has been noticed that H2A.B is less basic compared to H2A, due to the loss of lysines and a series of substitution of charged residues ^{23, 40} (see Figure 1.1c). The reduced electropositive potential may affect DNA binding directly. Moreover, the extensively altered pattern of charged residues in H2A.B may influence the stability of its dimer with H2B and/or its octamer with H2B, H3, and H4 ⁴¹. A more detailed examination of this hypothesis is discussed in **Chapter 3**. In addition, we note that one of minor groove anchoring arginines in H2A is shifted one residue position in most H2A.B sequences, which may alter the interaction with nucleosomal DNA. This is further investigated in **Chapter 4**.

H2A.B induces opening structure at DNA entry/exit ends of nucleosome

To investigate how the sequence differences in H2A.B impact nucleosome structure, Bao et al. reconstituted H2A.B nucleosome *in vitro* and studied the organization of the nucleosomal DNA using fluorescence resonance energy transfer (FRET) experiments and micrococcal nuclease (MNase) digestion assays²³. Both methods indicated that H2A.B nucleosome have a relatively open structure in which DNA entry/exit ends are far away from each other, even at low ionic strength. In particular, the MNase digestion assays showed that only ~118 bp were left intact after digestion of the free DNA ends while 146 bp are protected in the conventional case. Similar results were obtained later using the high-affinity, artificial '601' DNA sequence rather than the 5S DNA sequence⁴². Subsequent studies using atomic force microscopy (AFM) confirmed that less DNA is tightly bound to the histone octamer core in H2A.B nucleosomes, but arrived at ~130 bp DNA being protected⁴²⁻⁴³, possibly due to different experimental conditions. Electron cryo-microscopy, atomic force microscopy, SAXS, and small-angle neutron scattering experiments have further confirmed the less organized DNA entry/exit ends in the variant nucleosome (see Figure 1.1b)^{42, 44-46}. In an attempt to dissect the structural basis for the opened structure, various mutation studies have been conducted. A chimeric protein made by replacing H2A N-terminal tail with the H2A.B N-terminal tail showed no changes of the DNase digestion profile, proving that the elongated N-terminal tail of H2A.B has no impact on the opening structure of the variant nucleosome⁴². Similarly, a chimeric protein containing H2A histone fold and the H2A.B docking domain as well as a series of H2A mutants in which the docking domain was truncated at different positions (marked as triangles in Figure 1.1c) were constructed^{23, 42, 47}. These studies indicated that the truncation itself is not responsible for the open structure of H2A.B nucleosome. They also suggested that this region does not significantly contribute to nucleosome stability, but rather that the sequence difference within the truncated domain are responsible for this. In addition, replacement of the H2A.B docking domain with the H2A docking domain could rescue the orientation of the nucleosomal DNA entry/exit ends, further supporting the crucial

role of H2A docking domain in stabilizing the nucleosome structure⁴². It should be noted that it has not been tested whether extension of the H2A.B docking domain with the missing H2A C-terminal region can similarly rescue the formation of a closed nucleosome.

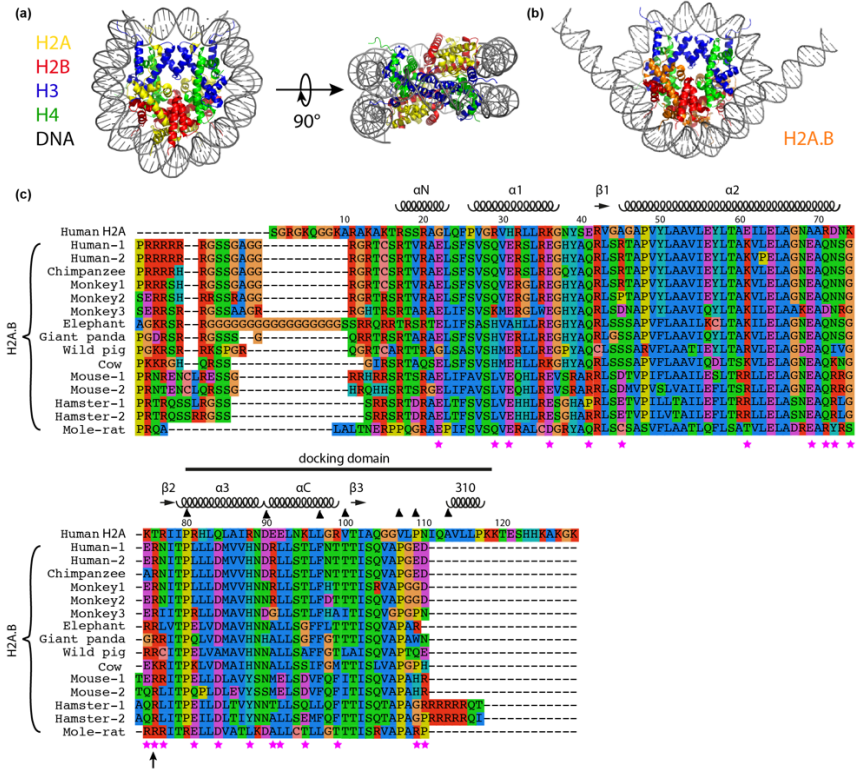


Figure 1.1. Nucleosome structure and sequence alignment of histone variant H2A.B with canonical H2A from human. (a) Top (left) and side (right) views on the crystal structure of canonical nucleosome (pdb: 2PYO). Color coding of the histones indicated in the figure. (b) Model of H2A.B nucleosome, made by opening the 15 bp of entry/exit DNA ends of the canonical crystal structure. (c) Sequence alignment of fifteen H2A.B sequences with human canonical H2A. Secondary structure elements based on the nucleosome crystal structure are indicated above the sequence. The docking domain of H2A is marked as black bar. Substitutions that alter the charge within the histone fold domain, based on the difference of human H2A.B and H2A, are marked as pink star below the sequence. The shifted minor groove arginine is marked as black arrow below the sequence. C-terminal docking domain truncation sites from literature are marked as black triangles above the sequence.

H2A.B prevents chromatin fiber formation

To examine the impact of H2A.B on higher-order chromatin structure, a few studies investigated the folding of H2A.B nucleosomal arrays. As discussed above, H2A.B nucleosomes will have a reduced acidic patch on their surface. Since the acidic patch is a crucial binding platform for the N-terminal H4 tail from an adjacent nucleosome to mediate the higher-order compaction of chromatin fibers⁴⁸, the loss of the three glutamic acids in the acidic patch of H2A.B nucleosomes can be expected to interrupt the nucleosome-nucleosome interactions. Indeed, H2A.B nucleosomes arrays were shown to fold as “beads on a string”, rather than a compacted fibre⁴⁹, in sharp contrast to arrays of histone variant H2A.Z nucleosomes, which fold more readily into fibres than canonical ones due to the more extended acidic patch of H2A.Z⁵⁰. Another study demonstrated by ultracentrifugation sedimentation experiments that an H2A.B mutant in which the three glutamic acids are restored, H2A.B-EE/E, fold as a fibre and that this folding of a H2A.B-EE/E nucleosome array is due to the restored acidic patch and not due to stably wrapping of 146 bp DNA within the mutant nucleosome³⁴. *In vivo* data for chromatin conformation containing H2A.B nucleosomes is lacking. However, H2A.B was found to be enriched at the actively transcribed genes in cells, which suggests H2A.B incorporation is related to a less compacted chromatin structure *in vivo*⁵¹⁻⁵².

Assembly, disassembly and remodelling of H2A.B nucleosomes *in vivo*

In the studies described above H2A.B nucleosomes were reconstituted *in vitro*. While canonical nucleosomes can be reconstituted from refolded histone octamers, the H2A.B containing octamer is not stable, so that reconstitution is carried out from refolded H2A.B-H2B dimers and H3-H4 tetramers that are deposited on DNA using salt gradient dialysis. How H2A.B nucleosomes are formed *in vivo* is yet unknown. While histone variants are typically incorporated by dedicated chaperones and/or remodelers, no such protein has been identified for H2A.B. However, histone chaperone NAP-I can mediate reversible

dimer exchange of H2A.B-H2B from the nucleosome at physiological ionic strength *in vitro* ⁵³.

Interestingly, H2A.B-H2B dimers may be much more dynamically incorporated than for canonical dimers. Photobleaching experiments on cells expressing GFP-H2A and GFP-H2A.B showed faster fluorescence recovery for H2A.B at the photobleached area, indicating higher mobility of the variant compared to the canonical form ^{45, 54}. Such high dynamic transfer of H2A.B is perhaps directly related to the instability of H2A.B nucleosome as measured by sedimentation and force-extension experiments ^{42, 54}.

Viewing the H2A.B nucleosome as a more labile roadblock, one could hypothesize that there is also a reduced need for remodeling of the nucleosome, at least for remodelers that typically open chromatin such as SWI/SNF. Indeed, while SWI/SNF binds to the H2A.B nucleosome with the same efficiency as to the canonical nucleosome, it is not capable of remodeling the variant nucleosome ⁴⁴. Experiments using domain swapped mutants showed that neither the elongated N-terminal tail of H2A.B nor docking domain are fully responsible for preventing SWI/SNF remodeling ⁴². For the related RSC remodeler, the truncated docking domain of H2A.B turned out to play a larger role in the resistance to remodeling ⁴⁷. Still, these results suggest that the histone fold part of H2A.B plays a role in preventing the remodeling ability of at least SWI/SNF. Interestingly, since the acidic patch has been implicated in remodeling by the ISWI remodeler (unrelated to SWI/SNF), the reduced acidic patch in H2A.B may be a molecular reason for the failure in remodeling.

H2A.B associates with active transcription and RNA splicing

The open DNA entry/exit end conformation of H2A.B nucleosome, together with the less compacted chromatin structure of H2A.B nucleosome arrays suggests the DNA in the variant chromatin is more accessible and thus more amenable to transcription, DNA replication, and repair. Indeed, increased transcription rate has been demonstrated *in vitro* using the H2A.B nucleosome arrays compared to the canonical arrays ³⁴. An extensive study by Tolstorukov *et al.* using *in vivo* ChIP-

seq data from HeLa cells showed that H2A.B was enriched in the bodies of actively expressed genes and its enrichment correlated with the gene expression level ⁵¹. Since depletion of H2A.B caused both up- and down-regulated gene expression and even a net decrease in transcription, the mild effects of H2A.B depletion suggested H2A.B is not directly regulating transcription. The authors used mass spectrometry to identify proteins that are specific to H2A.B enriched chromatin, revealing a number of RNA processing factors and spliceosome components. By depleting H2A.B, the frequency of exons included in mature transcript was elevated and the intronic read density was increased, which are consistent with a decrease in splicing efficiency ⁵¹. Similarly, mouse H2A.B was found to be enriched at actively transcribed genes and associated with mRNA splicing ^{24, 55-56}. In particular, one study demonstrated by using RNA-pull down assays that unlike canonical H2A or H2A.Z, the H2A.B N-terminal tail possesses RNA binding affinity and confirmed a direct interaction *in vivo* through cross-linking and immune-precipitation assays ⁵⁶. Further support for involvement in RNA splicing comes from H2A.B knock-out mouse study, showing loss of proper RNA Pol II localization and changes in pre-mRNA splicing ⁵⁷. Altogether there is substantial evidence for a role of H2A.B in RNA splicing, although the precise molecular details are still to be elucidated. In addition, H2A.B is also suggested to function in cell memory⁵⁸ and DNA repair⁵⁹.

H2A.B-like H2A variants: H2A.L, H2A.P and H2A.Q

In recent years, several other H2A variants have been identified within the mammalian lineage that like H2A.B have a shortened C-terminus with a truncated docking domain. These variants H2A.L, H2A.Q, H2A.P are closely related to H2A.B and supposed to share a common ancestor, H2A.R ⁶⁰. These short histone variants reside on the X-chromosome in most species ⁶⁰⁻⁶¹. As H2A.B, these short H2A variants lack the acidic patch that is characteristic for canonical H2A. Other features of H2A.B are mirrored also in the other members of the short H2A histone variant family. Indeed, H2A.L incorporated nucleosomes have been reported to associate with 130 bp DNA and its nucleosome array shows a beads on the string conformation ⁶². Even more dramatic

than in H2A.B, extensive charge-altering substitutions in H2A.P reduce the theoretical isoelectric point below 5, where it is 11 for canonical H2A⁶⁰. Two conserved arginine residues in H2A that contact the DNA minor-groove are lost in H2A.P and H2A.Q proteins, similar to the altered position of one of these in H2A.B. Finally, these short H2A variants do not appear in all mammals, suggesting they are functionally non-essential or redundant with each other and they may have different functions between different mammals⁶⁰.

Perspective and outline of this thesis

To summarize, H2A.B is one of the most divergent histone variants, which induces an open structure in its nucleosome and is involved in active gene transcription and mRNA splicing. While the functional role of H2A.B is slowly emerging, high-resolution structural characterization of the H2A.B nucleosome is still lacking. Yet, detailed knowledge on the structure and dynamics of the variant dimer and nucleosomes is essential for understanding its functional properties. While the crystal structure of the H2A.B-H2B heterodimer was solved very recently⁶³, a structure for the H2A.B nucleosome is still unavailable and unlikely to be solved at high resolution by crystallography or cryo-electron microscopy, due to the reduced DNA organization, higher dynamics and lower stability. In this work, we aimed to investigate the structure and dynamics of the H2A.B-H2B dimer and nucleosome at atomic resolution using NMR spectroscopy. Over the course of this decade, and in particular thanks to the introduction of NMR and isotope labeling techniques suitable for high molecular weight systems, nucleosomes have become amenable to NMR studies (for a recent review see reference⁶⁴). **Chapter 2** reviews the state-of-the-art isotope labeling methods that have extended the capability of solution state NMR to study macromolecular complexes. **Chapter 3** describes the solution structure of the H2A.B-H2B as well as a mutational study correlating surface charge of the heterodimer to its thermostability. The structural and dynamics properties of the H2A.B nucleosome are reported in **Chapter 4**, highlighting the impact of a position change at a minor groove anchoring arginine in the variant nucleosome. Next, the acidity of the canonical acidic patch is

investigated in **Chapter 5** to provide experimental mapping of the electrostatics. The thesis is concluded with a general discussion of the findings in **Chapter 6** followed by a summary.

References

1. Kossel, A., Über eine peptoartigen Bestandteil des zellkerns. *Z. Physiol. Chem.* **1884**, *8*, 511-515.
2. Phillips, D. M.; Johns, E. W., A Fractionation of the Histones of Group F2a from Calf Thymus. *Biochem J* **1965**, *94*, 127-30.
3. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171* (4356), 737-8.
4. Bram, S.; Ris, H., On the structure of nucleohistone. *J Mol Biol* **1971**, *55* (3), 325-36.
5. Sahasrabuddhe, C. G.; Van Holde, K. E., The effect of trypsin on nuclease-resistant chromatin fragments. *J Biol Chem* **1974**, *249* (1), 152-6.
6. Williamson, R., Properties of rapidly labelled deoxyribonucleic acid fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells. *J Mol Biol* **1970**, *51* (1), 157-68.
7. Hewish, D. R.; Burgoyne, L. A., Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Biophys Res Commun* **1973**, *52* (2), 504-10.
8. Kornberg, R. D., Chromatin structure: a repeating unit of histones and DNA. *Science* **1974**, *184* (4139), 868-71.
9. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, *389* (6648), 251-60.
10. Xu, F.; Zhang, K.; Grunstein, M., Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell* **2005**, *121* (3), 375-85.
11. Zhang, L.; Serra-Cardona, A.; Zhou, H.; Wang, M.; Yang, N.; Zhang, Z.; Xu, R. M., Multisite Substrate Recognition in Asf1-Dependent Acetylation of Histone H3 K56 by Rtt109. *Cell* **2018**, *174* (4), 818-830 e11.

12. Ye, J.; Ai, X.; Eugeni, E. E.; Zhang, L.; Carpenter, L. R.; Jelinek, M. A.; Freitas, M. A.; Parthun, M. R., Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly. *Mol Cell* **2005**, *18* (1), 123-30.
13. Yang, X.; Yu, W.; Shi, L.; Sun, L.; Liang, J.; Yi, X.; Li, Q.; Zhang, Y.; Yang, F.; Han, X.; Zhang, D.; Yang, J.; Yao, Z.; Shang, Y., HAT4, a Golgi apparatus-anchored B-type histone acetyltransferase, acetylates free histone H4 and facilitates chromatin assembly. *Mol Cell* **2011**, *44* (1), 39-50.
14. Pehrson, J. R.; Fried, V. A., MacroH2A, a core histone containing a large nonhistone region. *Science* **1992**, *257* (5075), 1398-400.
15. Kozlowski, M.; Corujo, D.; Hothorn, M.; Guberovic, I.; Mandemaker, I. K.; Blessing, C.; Sporn, J.; Gutierrez-Triana, A.; Smith, R.; Portmann, T.; Treier, M.; Scheffzek, K.; Huet, S.; Timinszky, G.; Buschbeck, M.; Ladurner, A. G., MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. *EMBO Rep* **2018**, *19* (10).
16. Tachiwana, H.; Kagawa, W.; Shiga, T.; Osakabe, A.; Miya, Y.; Saito, K.; Hayashi-Takanaka, Y.; Oda, T.; Sato, M.; Park, S. Y.; Kimura, H.; Kurumizaka, H., Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **2011**, *476* (7359), 232-5.
17. Strickland, M.; Strickland, W. N.; Brandt, W. F.; Von Holt, C.; Wittmann-Liebold, B., The complete amino-acid sequence of histone H2B(3) from sperm of the sea urchin *Parechinus angulosus*. *Eur J Biochem* **1978**, *89* (2), 443-52.
18. Draizen, E. J.; Shaytan, A. K.; Marino-Ramirez, L.; Talbert, P. B.; Landsman, D.; Panchenko, A. R., HistoneDB 2.0: a histone database with variants--an integrated resource to explore histones and their variants. *Database (Oxford)* **2016**, *2016*.
19. Ismail, I. H.; Hendzel, M. J., The gamma-H2A.X: is it just a surrogate marker of double-strand breaks or much more? *Environ Mol Mutagen* **2008**, *49* (1), 73-82.
20. Giaimo, B. D.; Ferrante, F.; Herchenrother, A.; Hake, S. B.; Borggrefe, T., The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* **2019**, *12* (1), 37.
21. Suto, R. K.; Clarkson, M. J.; Tremethick, D. J.; Luger, K., Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat Struct Biol* **2000**, *7* (12), 1121-4.
22. Chadwick, B. P.; Willard, H. F., A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* **2001**, *152* (2), 375-84.

23. Bao, Y.; Konesky, K.; Park, Y. J.; Rosu, S.; Dyer, P. N.; Rangasamy, D.; Tremethick, D. J.; Laybourn, P. J.; Luger, K., Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J* **2004**, *23* (16), 3314-24.
24. Soboleva, T. A.; Nekrasov, M.; Pahwa, A.; Williams, R.; Huttley, G. A.; Tremethick, D. J., A unique H2A histone variant occupies the transcriptional start site of active genes. *Nat Struct Mol Biol* **2011**, *19* (1), 25-30.
25. Sierra, I.; Anguera, M. C., Enjoy the silence: X-chromosome inactivation diversity in somatic cells. *Curr Opin Genet Dev* **2019**, *55*, 26-31.
26. Loda, A.; Heard, E., Xist RNA in action: Past, present, and future. *PLoS Genet* **2019**, *15* (9), e1008333.
27. Costanzi, C.; Pehrson, J. R., Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **1998**, *393* (6685), 599-601.
28. Chakravarthy, S.; Gundimella, S. K.; Caron, C.; Perche, P. Y.; Pehrson, J. R.; Khochbin, S.; Luger, K., Structural characterization of the histone variant macroH2A. *Mol Cell Biol* **2005**, *25* (17), 7616-24.
29. Naylor, J. A.; Buck, D.; Green, P.; Williamson, H.; Bentley, D.; Giannelli, F., Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Hum Mol Genet* **1995**, *4* (7), 1217-24.
30. Horn, V.; Uckelmann, M.; Zhang, H.; Eerland, J.; Aarsman, I.; le Paige, U. B.; Davidovich, C.; Sixma, T. K.; van Ingen, H., Structural basis of specific H2A K13/K15 ubiquitination by RNF168. *Nat Commun* **2019**, *10* (1), 1751.
31. Mattioli, F.; Vissers, J. H.; van Dijk, W. J.; Ikpa, P.; Citterio, E.; Vermeulen, W.; Marteijn, J. A.; Sixma, T. K., RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. *Cell* **2012**, *150* (6), 1182-95.
32. Poplawski, A.; Hu, K.; Lee, W.; Natesan, S.; Peng, D.; Carlson, S.; Shi, X.; Balaz, S.; Markley, J. L.; Glass, K. C., Molecular insights into the recognition of N-terminal histone modifications by the BRPF1 bromodomain. *J Mol Biol* **2014**, *426* (8), 1661-76.
33. Zhou, W.; Zhu, P.; Wang, J.; Pascual, G.; Ohgi, K. A.; Lozach, J.; Glass, C. K.; Rosenfeld, M. G., Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell* **2008**, *29* (1), 69-80.

34. Zhou, J.; Fan, J. Y.; Rangasamy, D.; Tremethick, D. J., The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat Struct Mol Biol* **2007**, *14* (11), 1070-6.
35. Mucke, K.; Paulus, C.; Bernhardt, K.; Gerrer, K.; Schon, K.; Fink, A.; Sauer, E. M.; Asbach-Nitzsche, A.; Harwardt, T.; Kieninger, B.; Kremer, W.; Kalbitzer, H. R.; Nevels, M., Human cytomegalovirus major immediate early 1 protein targets host chromosomes by docking to the acidic pocket on the nucleosome surface. *J Virol* **2014**, *88* (2), 1228-48.
36. Makde, R. D.; England, J. R.; Yennawar, H. P.; Tan, S., Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **2010**, *467* (7315), 562-6.
37. 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 Å resolution. *Science* **2011**, *334* (6058), 977-82.
38. Kato, H.; van Ingen, H.; Zhou, B. R.; Feng, H.; Bustin, M.; Kay, L. E.; Bai, Y., Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A* **2011**, *108* (30), 12283-8.
39. Kalashnikova, A. A.; Porter-Goff, M. E.; Muthurajan, U. M.; Luger, K.; Hansen, J. C., The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface* **2013**, *10* (82), 20121022.
40. Eirin-Lopez, J. M.; Ishibashi, T.; Ausio, J., H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *FASEB J* **2008**, *22* (1), 316-26.
41. Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I., Protein stability and surface electrostatics: a charged relationship. *Biochemistry* **2006**, *45* (9), 2761-6.
42. Doyen, C. M.; Montel, F.; Gautier, T.; Menoni, H.; Claudet, C.; Delacour-Larose, M.; Angelov, D.; Hamiche, A.; Bednar, J.; Faivre-Moskalenko, C.; Bouvet, P.; Dimitrov, S., Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *EMBO J* **2006**, *25* (18), 4234-44.
43. Montel, F.; Fontaine, E.; St-Jean, P.; Castelnovo, M.; Faivre-Moskalenko, C., Atomic force microscopy imaging of SWI/SNF action: mapping the nucleosome remodeling and sliding. *Biophys J* **2007**, *93* (2), 566-78.

44. Angelov, D.; Verdel, A.; An, W.; Bondarenko, V.; Hans, F.; Doyen, C. M.; Studitsky, V. M.; Hamiche, A.; Roeder, R. G.; Bouvet, P.; Dimitrov, S., SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays. *EMBO J* **2004**, *23* (19), 3815-24.
45. Arimura, Y.; Kimura, H.; Oda, T.; Sato, K.; Osakabe, A.; Tachiwana, H.; Sato, Y.; Kinugasa, Y.; Ikura, T.; Sugiyama, M.; Sato, M.; Kurumizaka, H., Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci Rep* **2013**, *3*, 3510.
46. Sugiyama, M.; Arimura, Y.; Shirayama, K.; Fujita, R.; Oba, Y.; Sato, N.; Inoue, R.; Oda, T.; Sato, M.; Heenan, R. K.; Kurumizaka, H., Distinct features of the histone core structure in nucleosomes containing the histone H2A.B variant. *Biophys J* **2014**, *106* (10), 2206-13.
47. Shukla, M. S.; Syed, S. H.; Goutte-Gattat, D.; Richard, J. L.; Montel, F.; Hamiche, A.; Travers, A.; Faivre-Moskalenko, C.; Bednar, J.; Hayes, J. J.; Angelov, D.; Dimitrov, S., The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling. *Nucleic Acids Res* **2011**, *39* (7), 2559-70.
48. Dorigo, B.; Schalch, T.; Kulangara, A.; Duda, S.; Schroeder, R. R.; Richmond, T. J., Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* **2004**, *306* (5701), 1571-3.
49. Montel, F.; Menoni, H.; Castelnovo, M.; Bednar, J.; Dimitrov, S.; Angelov, D.; Faivre-Moskalenko, C., The dynamics of individual nucleosomes controls the chromatin condensation pathway: direct atomic force microscopy visualization of variant chromatin. *Biophys J* **2009**, *97* (2), 544-53.
50. Fan, J. Y.; Rangasamy, D.; Luger, K.; Tremethick, D. J., H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. *Mol Cell* **2004**, *16* (4), 655-61.
51. Tolstorukov, M. Y.; Goldman, J. A.; Gilbert, C.; Ogryzko, V.; Kingston, R. E.; Park, P. J., Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol Cell* **2012**, *47* (4), 596-607.
52. Ioudinkova, E. S.; Barat, A.; Pichugin, A.; Markova, E.; Sklyar, I.; Pirozhkova, I.; Robin, C.; Lipinski, M.; Ogryzko, V.; Vassetzky, Y. S.; Razin, S. V., Distinct distribution of ectopically expressed histone variants H2A.Bbd and MacroH2A in open and closed chromatin domains. *PLoS One* **2012**, *7* (10), e47157.
53. Okuwaki, M.; Kato, K.; Shimahara, H.; Tate, S.; Nagata, K., Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. *Mol Cell Biol* **2005**, *25* (23), 10639-51.

54. Gautier, T.; Abbott, D. W.; Molla, A.; Verdel, A.; Ausio, J.; Dimitrov, S., Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep* **2004**, *5* (7), 715-20.
55. Chen, Y.; Chen, Q.; McEachin, R. C.; Cavalcoli, J. D.; Yu, X., H2A.B facilitates transcription elongation at methylated CpG loci. *Genome Res* **2014**, *24* (4), 570-9.
56. Soboleva, T. A.; Parker, B. J.; Nekrasov, M.; Hart-Smith, G.; Tay, Y. J.; Tng, W. Q.; Wilkins, M.; Ryan, D.; Tremethick, D. J., A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS Genet* **2017**, *13* (2), e1006633.
57. Anuar, N. D.; Kurscheid, S.; Field, M.; Zhang, L.; Rebar, E.; Gregory, P.; Buchou, T.; Bowles, J.; Koopman, P.; Tremethick, D. J.; Soboleva, T. A., Gene editing of the multi-copy H2A.B gene and its importance for fertility. *Genome Biol* **2019**, *20* (1), 23.
58. Wu, B. J.; Dong, F. L.; Ma, X. S.; Wang, X. G.; Lin, F.; Liu, H. L., Localization and expression of histone H2A variants during mouse oogenesis and preimplantation embryo development. *Genet Mol Res* **2014**, *13* (3), 5929-39.
59. Sansoni, V.; Casas-Delucchi, C. S.; Rajan, M.; Schmidt, A.; Bonisch, C.; Thomae, A. W.; Staeger, M. S.; Hake, S. B.; Cardoso, M. C.; Imhof, A., The histone variant H2A.Bbd is enriched at sites of DNA synthesis. *Nucleic Acids Res* **2014**, *42* (10), 6405-20.
60. Molaro, A.; Young, J. M.; Malik, H. S., Evolutionary origins and diversification of testis-specific short histone H2A variants in mammals. *Genome Res* **2018**, *28* (4), 460-473.
61. Ferguson, L.; Ellis, P. J.; Affara, N. A., Two novel mouse genes mapped to chromosome Yp are expressed specifically in spermatids. *Mamm Genome* **2009**, *20* (4), 193-206.
62. Syed, S. H.; Boulard, M.; Shukla, M. S.; Gautier, T.; Travers, A.; Bednar, J.; Faivre-Moskalenko, C.; Dimitrov, S.; Angelov, D., The incorporation of the novel histone variant H2AL2 confers unusual structural and functional properties of the nucleosome. *Nucleic Acids Res* **2009**, *37* (14), 4684-95.
63. Dai, L.; Xie, X.; Zhou, Z., Crystal structure of the histone heterodimer containing histone variant H2A.Bbd. *Biochem Biophys Res Commun* **2018**, *503* (3), 1786-1791.

64. Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J., Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr Opin Struct Biol* **2015**, 32, 113-22.

Chapter 2. Isotope-labeling strategies for solution NMR studies of macromolecular assemblies

This chapter is based on:

Zhang, H. & van Ingen, H. Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. *Current Opinion in Structural Biology*. 2016 June; 38:75-82. Doi: 10.1016/j.sbi.2016.05.008.

Abstract

Proteins come together in macromolecular assemblies, recognizing and binding to each other through their structures, and operating on their substrates through their motions. Detailed characterization of these processes is particularly suited to NMR, a high-resolution technique sensitive to structure, dynamics, and interactions. Advances in isotope-labeling have enabled such studies to an ever-increasing range of systems. Here we highlight recent applications and bring to the fore the range of options to produce labeled proteins and to control the specific placement of isotopes. The increased labeling control and affordability, together with the possibility to combine strategies will further deepen and extend the range of protein assembly investigations.

Introduction

Proper cellular functioning depends critically on networks of biomolecular interactions. Proteins at the nodes of these networks interact with and operate on other proteins, nucleic acids, and small-molecule ligands. Thus, understanding protein function at the molecular level is a key goal in life sciences research. Structural biologists and biochemists pursue this goal by investigating the structures, dynamics, and interactions of proteins. The key technologies used include crystallography, nuclear magnetic resonance spectroscopy (NMR), electron paramagnetic resonance, cryo-electron microscopy, and small-angle scattering. NMR has the unique advantages that it allows to study proteins and protein interactions at atomic resolution, in solution, and that it is exquisitely sensitive to a wide range of protein motions. Such studies require the incorporation of NMR-active isotopes of nitrogen (^{15}N) and carbon (^{13}C), sometimes in combination with deuterium (^2H), to allow residue and atom-specific interpretation of the NMR spectrum.

Here, we review recent developments in isotopic labeling strategies in solution-state NMR, focusing on the study of macromolecular assemblies (Figure 2.1). The size of such complexes and/or the complexity of the protein of interest generally require different approaches from the conventional uniform [^2H , ^{13}C , ^{15}N]-labeling, briefly reviewed below. These strategies require restricted placement of isotopes in order to reduce the number of signals, usually in combination with deuteration of unwanted signals to enhance sensitivity. Here, we review recent progress in and highlight examples of selective labeling of methyl-groups, defined protein segments, or specific subunits. These strategies are applied separately or in combination to achieve high-quality spectra for demanding systems. Finally, we review ^{19}F fluorine labeling and isotopic labeling in cell-free systems, yeasts and insect cells that enable NMR studies of challenging eukaryotic proteins.

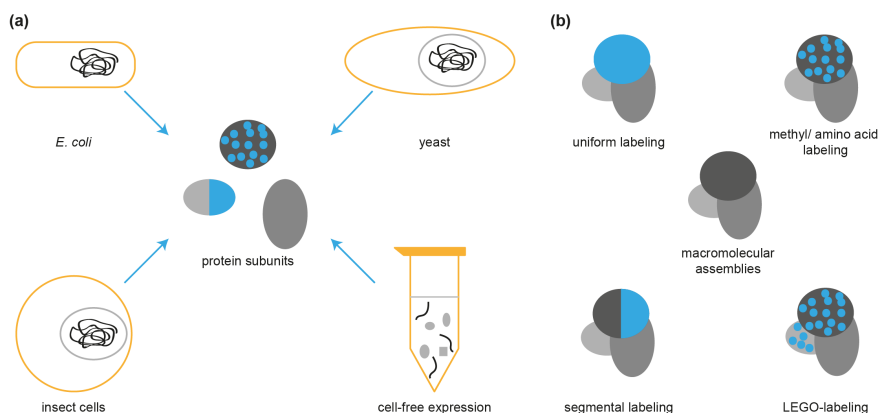


Figure 2.1. Overview of labeling strategies available for the study of macromolecular protein assemblies. Schematic overview of (a) different expression hosts available to produce isotope-label proteins with ^2H , ^{13}C and ^{15}N , and (b) different labeling schemes that can be applied. Blue proteins are NMR-active, isotope-labeled, gray proteins are unlabeled (or deuterated) and NMR-inactive. Expression in *E. coli* is compatible with all labeling methods, cell-free expression with uniform, methyl-selective, amino-acid selective and segmental labeling, yeast-based expression with uniform and methyl-selective labeling, insect-cell-based expression with uniform or amino-acid selective and methyl-selective labeling. Notably, reconstitution of the complex takes place *in vivo* after expression of subunits in LEGO-NMR labeling, whereas the other cases depicted in (b) require reconstitution *in vitro*.

Conventional uniform labelling

In the typical uniform labeling strategy, proteins are overexpressed by manual induction of a suitable T7-based *E. coli* strain¹, grown in M9 minimal medium supplemented with ^{13}C -labeled glucose and $^{15}\text{NH}_4\text{Cl}$ as the sole carbon and nitrogen sources, respectively. Proteins larger than 20-25 kDa are typically deuterated by using $^2\text{H}_2\text{O}$ (D_2O) in the cell growth medium instead of $^1\text{H}_2\text{O}$, optionally combined with the use of fully deuterated and ^{13}C -labeled glucose as the carbon source. Combining [$^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$]-labeling and transverse relaxation optimized spectroscopy (TROSY²), allows structural and dynamical characterization of proteins in complexes in the 100 kDa size range, for recent examples see³⁻⁴, up to 1 MDa⁵⁻⁶. Additionally, uniform labeling is valuable to study individual subunits in the ‘divide-and-

conquer' strategy. For single-chain proteins beyond 50 kDa, however, the sheer amount of signals complicates the spectra, and assignment becomes increasingly difficult.

In case of large assemblies, simple uniform labeling can be exploited to selectively observe highly flexible regions, as is nicely illustrated in a recent study on the nucleosome ⁷. Its histone H3 subunit has a highly flexible N-terminal tail that is effectively decoupled from the slow overall molecular tumbling of the nucleosome (~220 kDa). Due to the large overall size, signals of the rigid part of uniformly [¹⁵N,¹³C]-labeled H3 are effectively broadened beyond detection, leaving a simplified spectrum of N-terminal tail. Using this approach, Stützer *et al.* were able to show that the H3 tail interacts with linker DNA and that this reduces the modifiability of the histone tail.

As an alternative to manual induction, auto-induction media have been developed offering overexpression in an unattended manner, better reproducibility, and higher levels of soluble protein expression ⁸. Auto-induction media are composed of glucose, lactose and glycerol as carbon sources, triggering T7-based expression strains to be automatically induced by lactose after consuming all glucose present. For uniform ¹³C or ²H-labeling, such media are prohibitively expensive due to the need for labeled lactose. Recently, Guthertz and his colleagues showed that only the glucose moiety of lactose needs to be isotope-labeled, taking advantage of the inability of *E. coli* BL21 to metabolize the galactose moiety ⁹. Specifically labeled lactose was synthesized from unlabeled galactose and ¹³C or ²H-labeled glucose, and used to produce uniformly ¹³C or ²H-labeled proteins.

Interestingly, O'Brien *et al.* proposed a novel method to produce deuterated proteins in H₂O medium ¹⁰. The uniform ²H, ¹⁵N, ¹³C-labeling is achieved by adding ²H, ¹⁵N, ¹³C labeled nutrients prior to IPTG induction in the H₂O M9 medium where the unlabeled nutrients are exhausted. This approach was optimized to achieve 80% deuteration for ²H, ¹⁵N uniform labeling, however is less sufficient for triple ²H, ¹⁵N, ¹³C labeling. Nevertheless, this approach provides a more cost-effective and feasible uniform as well as methyl-specific isotope labeling approach for NMR studies.

Methyl-TROSY labelling

The method of choice for the quantitative study of high molecular weight systems (> 100 kDa) is the specific labeling of methyl groups in a highly deuterated background ¹¹⁻¹². Methyl groups are ideal candidates to be specifically isotopic labeled because they are abundant, found both in the core and on the surface of protein structures ¹³; they carry three protons, and their symmetry and rapid rotation can be exploited to yield intense and well-resolved NMR signals ¹⁴⁻¹⁵. Originally developed in the Kay lab for Ile- $\delta 1$, Leu, Val methyl groups, this labeling strategy requires perdeuterated proteins, into which specific [^1H , ^{13}C]-labeled methyl groups are introduced using deuterated amino acids precursors that only [^1H , ^{13}C]-labeled on the methyl group of interest ¹⁶⁻¹⁷. Methyl-labeling has since been extended to Ile- $\gamma 2$, Ala, Met, Thr methyl groups ¹⁸⁻²² and is thoroughly reviewed in ²³.

Developments during the last 5-6 years have focused on reducing overlap and increasing sensitivity of methyl-TROSY spectra by independently labeling Leu and Val methyl groups, and extending this capability to the stereo-specific labeling of these prochiral methyl groups ²⁴. In the original protocol, these methyl groups cannot be separated as they originate from a common precursor. Lichteneker *et al.* developed protocols to selectively label Val or Leu methyl groups using custom synthesized Leu precursors ²⁵⁻²⁶. Selective and stereo-specific labeling of Val methyl groups was achieved by Mas *et al.* using specifically labeled 2-acetolactate as Leu/Val precursor together with addition of perdeuterated Leu in the culture medium to prevent conversion of the precursor to labeled Leu ²⁷. In a third approach, the culture medium is supplemented with custom synthesized stereo-specifically labeled Leu and Val amino acids rather than their precursors ²⁸. With this approach fully independent labeling of either pro-*R* or pro-*S* methyl group of either Leu or Val, e.g. pro-*R* Leu- $\delta 1$ with pro-*S* Val- $\gamma 2$, is possible by proper choice of amino acid supplement. The Boissbouvier lab recently developed a protocol where Ala- β , Ile- $\delta 1$, Leu-pro*S* and Val-pro*S* are simultaneously methyl labeled, relying on a custom synthesized Ile-precursor to avoid co-incorporation incompatibility and isotopic scrambling ²⁹. They demonstrate the suitability of this scheme to measure methyl-methyl

distances for structural studies of high molecular weight systems. A different approach was developed by Miyanoiri et al. using auxotrophic *E.coli* strain of which biosynthesis pathways of Ile, Leu and Val are blocked to achieve stereo-specific labeling of $^{13}\text{CH}_3$ -Ile, -Val, and -Leu without any amino acid scrambling³⁰.

Survey of recent literature shows many great examples of how this labeling strategy can generate exciting insights in the structure-dynamics-function relationship of protein-protein, protein-DNA, and protein-small molecule complexes involved in protein folding³¹⁻³³, regulation of protein expression³⁴⁻³⁸, protein signal-transduction³⁹⁻⁴⁰ and protein secretion⁴¹⁻⁴³. We highlight here the work from the Kalodimos lab on the interaction of the 50 kDa trigger factor (TF) chaperone with a 48 kDa unfolded substrate, alkaline phosphatase (PhoA)⁴⁴. Taking advantage of the modular nature of the PhoA-TF complex, Saio *et al.* were able to show that three TF molecules are required to interact with the entire length of PhoA, resulting in a ~200 kDa complex in solution. Using methyl-group labeled samples as the cornerstone in their NMR data collection and analysis, high-resolution NOE-based structures were determined for each TF bound to a PhoA segment. The resulting structures show how the same substrate-binding region in the chaperone engages different hydrophobic stretches of the unfolded PhoA.

Segmental labelling

Isotope-labeling of selected segments of a protein can greatly reduce the complexity of NMR spectra. Labeled and unlabeled protein segments are produced separately, and then fused via a thioester-intermediate to ultimately form a native peptide bond (Figure 2.2a,b). Rooted in native chemical ligation where both parts are produced synthetically⁴⁵, recombinant protein segments are fused using either inteins⁴⁶⁻⁴⁹ or sortase⁵⁰. Both methods require a judicious choice of the ligation point, typically in a domain-connecting loop.

Inspired by protein splicing, the intein-based approaches rely on the use of internal protein domains (inteins) that can excise themselves from a protein in a traceless manner. In expressed protein ligation (EPL⁵¹⁻⁵²), the required thioester intermediate is formed after expression of the N-terminal protein fused to an intein, allowing subsequent ligation

kDa hexameric ClpB chaperone illustrates the dramatic improvement in spectral quality in a segmental methyl-selective labeled complex (right) over the uniformly methyl-selective labeled complex (left). Color coding of the assembly cartoon as in Figure 2.1. Figure adapted from ⁶⁴ with permission from the authors.

with the C-terminal part (Figure 2.2a). In protein trans-splicing (PTS), both parts of the protein are fused to a split intein, and expressed either separately, or sequentially from different promoters, to allow differential labeling ^{48, 53}. The split intein-fusions are reassembled *in vitro* or *in vivo* to an active intein that excises itself, resulting in a native, fused target protein ⁵³ (Figure 2.2a). Notably, intein activity in PTS may depend critically on the protein context and unwanted “cross-labeling” may occur when splicing is carried out *in vivo* ⁵⁴.

Development in intein-based segmental protein production has focused mainly on the identification of better split inteins for PTS ⁵⁵⁻⁵⁷. Recently, a highly active and extremely stable split intein was designed promising higher yields and increased robustness in PTS ⁵⁸. In addition, generic gene insert was designed containing a split intein, termed PTS cassette, to screen split intein insertion sites for any target proteins under the control of T7 promoter ⁵⁹.

An attractive alternative to intein-based segmental labeling is the *in vitro* ligation approach based on the transpeptidase Sortase A (SrtA) ⁵⁰, in which protein segments are produced with or without isotopic labeling, purified separately and ligated *in vitro*, without risking cross-labeling contamination. The sortase enzyme recognizes an LPXTG motif on the N-terminal segment and catalyzes the formation of a new peptide bond with the C-terminal part (Figure 2.2b). To highlight, Bobby lab used this powerful method to study ligand-bromodomain interaction at high resolution by strategically labelling on the C-terminal bromodomain whereas the N-terminal bromodomain remained unlabeled ⁶⁰. Recently, the Sattler lab developed a modified ligation protocol, addressing the reversibility of sortase reaction ⁶¹. Using a centrifugal concentrator to continuously remove the cleaved glycine and a clever combination of cleavable and non-cleavable purification-tags, ligation efficiency for tested proteins (a 32 kDa dual RRM-domain protein and the 57 kDa Hsp90 chaperone) was improved up to two-fold.

While EPL, PTS, and SrtA methods have been successfully applied for segmental isotope labeling of multidomain proteins, it is more

challenging to apply to single domain globular proteins. This is because the split fragments of globular proteins are usually insoluble, which requires extra refolding steps. To solve this, a new approach using asparaginyl endopeptidases (AEP) was proposed⁶²⁻⁶³ (Figure 2.2c). Compared to sortase, AEP recognizes a shorter motif of NGL on the N-terminal segment of the protein and leaves a shorter ligation tag in the catalyzed protein ligation, which is less likely to disturb the solubility of the split fragments. In the demonstrated case of MAP, the two fragments were folded and purified before ligated by AEP *in vitro*. This new strategy, in combination with PTS, provides new possibilities for production of more complex protein conjugates with various biophysical probes.

Recent work from Rosenzweig *et al.* on the substrate recognition of the 580-kDa hexameric ClpB chaperone demonstrates the dramatic spectral improvement segmental labeling can offer⁶⁴ (Figure 2.2d). The N-terminal domain (NTD, 16 kDa) of the ClpB monomer (97 kDa) was expressed as an intein-fusion, with methyl-group specific isotope-labeling, whereas the remainder of ClpB was fully deuterated. The ligated, segmentally labeled ClpB monomer was subsequently reassembled into its functional hexameric form. The resulting high-quality methyl-TROSY spectra were used to determine microscopic binding affinities of a client protein to two separate sites on ClpB. Together with biochemical assays, these results established the NTD as a protein aggregate sensor that binds client protein before they are shuttled through the ClpB active channel for unfolding.

LEGO-NMR subunit labelling

Protein complexes are typically reconstituted *in vitro*, permitting the selective labeling of one or more subunits. This approach may fail for complexes for which the individual subunits have poor solubility. The LEGO-NMR strategy was recently introduced to overcome this problem⁶⁵. In a method akin to *in vivo* PTS, all subunits are co-expressed in a single *E. coli* cell from two plasmids, one inducible by arabinose with glycerol as carbon source, and the other by IPTG with glucose as carbon source. This setup permits the selective labeling of a subset of subunits and the *in vivo* assembly of labeled and unlabeled subunits into a functional complex. Mund *et al.* demonstrated this

technique to label, express and generate oligomers (LEGO) on a ~75 kDa complex, comprised of 7 subunits, which was selectively [^2H , ^{15}N]-labeled on three or single subunits, allowing precise mapping of an RNA binding site. Furthermore, compatibility with selective methyl-labeling was neatly demonstrated by preparation of a complex with selective methyl-labeling of Met in 3 subunits and of Ile- $\delta 1$ in the remaining 4 subunits.

Fluorine-labelling

As an alternative to $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ isotope-labeling, incorporation of ^{19}F isotopes can offer a highly sensitive probe of conformational changes, dynamics and interactions because of its high abundance, gyromagnetic ratio and chemical shift range (for a recent review see ⁶⁶). Uniform labeling with fluorinated amino acids analogs is achieved using bacterial strains auxotrophic for the substituted amino acid, or using the amber-codon approach to achieve site-specific labeling. Alternatively, fluorinated tags, such as 3-bromo-1,1,1-trifluoroacetone, are attached to cysteine-thiol groups or other labile groups. Recently, chemical shift sensitivity of CF_3 tags has been compared to optimize resolution ⁶⁷. CF_3 tags with distinct chemical shifts were also used for differential ^{19}F labeling of proteins to study individual behavior of each protein in their mixtures ⁶⁸. Combination of paramagnetic and ^{19}F labeling was recently demonstrated to obtain precise long-range distance measurements ⁶⁹. Furthermore, enzymatic ^{19}F labeling of glutamine side chain carboxamide group by transglutaminase was developed to study the drug-protein and protein-protein interactions, as demonstrated on the complexes of about 100 kDa ⁷⁰. The advantages of ^{19}F labeling are nicely illustrated in recent studies where the chemical shift sensitivity of ^{19}F was exploited to identify different conformational states of GPCRs ⁷¹⁻⁷² and substrate-arrestin complexes ⁷³.

Isotope-labelling in yeast and insect cells

Expression in *E. coli* is widely used due to its high-level of protein production and cheap growth media. It may fail, however, to produce functional recombinant proteins, especially in case they require

eukaryotic folding machineries, glycosylation or other post-translational modifications. Cells from higher organisms, most commonly yeasts and baculovirus infected insect cells ⁷⁴ are necessarily used as expression systems to isotope-label these proteins, permitting NMR studies of otherwise intractable protein assemblies.

Expression in yeast is attractive because of the low-cost minimal growth medium and relatively high protein expression yields. Recently, selective [¹H,¹³C]-labeling of Ile- δ 1 methyl groups in perdeuterated proteins has been described in glucose-controlled *Kluyveromyces lactis* ⁷⁵ and methanol-controlled *Pichia pastoris* ⁷⁵⁻⁷⁶. The 42 kDa maltose-binding protein was perdeuterated to high levels ($\geq 90\%$) with Ile- δ 1 labeling efficiency of 45% and 67% for *P. pastoris* and *K. lactis*, respectively. For both systems, methyl-selective Leu/Val labeling was $<5\%$, although significant improvement is possible through co-expression of metabolic enzymes or labeled Leu/Val supplementation ⁷⁵.

Isotope-labeling in insect cells requires the use of labeled amino acids as medium-supplement. The associated high costs are raised even further for large proteins requiring deuterated amino acids. Recently, protocols for cheaper media have been proposed based on custom-made isotope-labeled yeast extracts, demonstrating the feasibility of uniform ¹⁵N-labeling ⁷⁷, and uniform [²H,¹³C,¹⁵N]-labeling ⁷⁸. Opitz *et al.* achieved $>80\%$ ¹³C/¹⁵N incorporation and $\sim 60\%$ deuteration, producing samples suitable for triple resonance experiments and detailed structural analysis ⁷⁸. Sitarska and colleagues optimized a protocol based on commercially available isotope-labeled algae extracts, resulting in triple-labeled proteins with similar efficiency and costs compared to the yeast-based method ⁷⁹.

Here, we highlight recent studies of solubilized membrane proteins that are expressed and isotope-labeled on specific amino acids in insect cells ⁸⁰⁻⁸⁵. Nygaard *et al.* used specific ¹³C-labeling of Met methyl groups to study the conformational heterogeneity of a detergent-GPCR complex in diverse ligand-bound states ⁸⁰. In a subsequent study, the GPCR was embedded in lipid bilayer nanodiscs and deuterated up to 90% using a combination of ²H-labeled algae extracts and ²H-amino acids ⁸¹. Recently, the Grzsiak lab studied the β 1-adrenergic receptor GPCR as a 100 kDa detergent-GPCR complex using specific ¹⁵N-labeling of Val residues, resulting in highly quality TROSY spectra

where 21 out of 28 possible Val probes could be resolved and assigned⁸². Ligand binding caused chemical shift changes at the opposite end of the GPCR, which correlated linearly to the G-protein activation efficiency of each ligand, demonstrating an allosteric coupling between the extracellular ligand binding site and the intracellular G-protein binding site.

Cell-free isotope-labelling

The exemplification of cell-free based isotope-labeling is stereo-array isotope-labeling (SAIL) where a cocktail of specifically [²H, ¹³C, ¹⁵N]-labeled amino acids is used to produce proteins with optimal NMR properties⁸⁶. The SAIL method takes full advantage of the lack of isotope scrambling in cell-free protein synthesis and the smaller amounts of amino acid supplementation required, compared to *in vivo* expression. Other advantages of cell-free expression are that it offers possibility to express toxic proteins, to improve protein production by adjusting the cell-extract with various factors⁸⁷, and to produce solubilized membrane-proteins without co-purification of endogenous lipids⁸⁸.

Recently, three new strategies have been put forward to optimize labeling of large proteins in cell-free expression. First, combination of cell-free expression with segmental labeling was proposed to generate multi-domain proteins with a specific pattern of amino acid labeling restricted to each domain⁸⁹. This was demonstrated on a two-domain protein, where a ¹⁵N-Lys labeled intein-fusion was ligated using EPL to a [¹³C, ¹⁵N]-Lys labeled domain. Second, the high cost of selective methyl-group labeling has been reduced greatly, making use of hydrolyzed methyl-labeled inclusion bodies derived from *E. coli* to replace commercial labeled amino acids⁹⁰. This approach was illustrated on an Ile- δ 1, Val/Leu-proS methyl-labeled eukaryotic membrane protein, toxic to *E. coli*. While this second method still relies partially on the cellular expression, the most recent strategy uses additional branched chain aminotransferase IlveE to directly convert precursors into L-Val and L-Leu (and potentially L-Ile as well) for the synthesis of the target protein in cell-free system⁹¹.

Conclusion

Here, we highlighted the increasing range of options regarding expression system and labeling strategy that is available for solution NMR studies of protein complexes. The availability of affordable deuteration and methyl-labeling protocols for non-*E. coli* based expression, as well as the LEGO-NMR approach, widen the application window to otherwise intractable systems. Control over the restricted placement of isotopes offers an extremely valuable degree of flexibility, in particular when both backbone and methyl-TROSY spectra are of good quality. The ‘best’ labeling strategy remains case-dependent: the size and behavior of complex and its subunits, the question at hand, and the spectral quality required versus costs and time affordable will dictate the strategy chosen. We anticipate that especially the combination of labeling strategies, such as segmental methyl-labeling, will prove extraordinarily powerful in the dissection of the inner workings of Nature’s molecular machines.

Acknowledgements

We thank all members of the MacBio group for stimulating discussions. This work was supported by a VIDI grant from the Dutch Science Foundation NWO to HvI (NWO-CW VIDI 723.013.010).

References

1. Studier, F. W.; Moffatt, B. A., Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **1986**, *189* (1), 113-30.
2. 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* **1997**, *94* (23), 12366-71.
3. Brewer, K. D.; Bacaj, T.; Cavalli, A.; Camilloni, C.; Swarbrick, J. D.; Liu, J.; Zhou, A.; Zhou, P.; Barlow, N.; Xu, J.; Seven, A. B.; Prinslow, E. A.; Voleti, R.; Haussinger, D.; Bonvin, A. M.; Tomchick, D. R.; Vendruscolo, M.; Graham, B.; Sudhof, T. C.; Rizo, J., Dynamic binding mode of a Synaptotagmin-1-SNARE complex in solution. *Nat Struct Mol Biol* **2015**, *22* (7), 555-64.

4. Lakomek, N. A.; Kaufman, J. D.; Stahl, S. J.; Louis, J. M.; Grishaev, A.; Wingfield, P. T.; Bax, A., Internal dynamics of the homotrimeric HIV-1 viral coat protein gp41 on multiple time scales. *Angew Chem Int Ed Engl* **2013**, 52 (14), 3911-5.
5. Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wuthrich, K., NMR analysis of a 900K GroEL GroES complex. *Nature* **2002**, 418 (6894), 207-11.
6. Libich, D. S.; Fawzi, N. L.; Ying, J.; Clore, G. M., Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. *Proc Natl Acad Sci U S A* **2013**, 110 (28), 11361-6.
7. Stutzer, A.; Liokatis, S.; Kiesel, A.; Schwarzer, D.; Sprangers, R.; Soding, J.; Selenko, P.; Fischle, W., Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Mol Cell* **2016**, 61 (2), 247-59.
8. Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **2005**, 41 (1), 207-34.
9. Guthertz, N.; Klopp, J.; Winterhalter, A.; Fernandez, C.; Gossert, A. D., Auto-inducing media for uniform isotope labeling of proteins with (15)N, (13)C and (2)H. *J Biomol NMR* **2015**, 62 (2), 169-77.
10. O'Brien, E. S.; Lin, D. W.; Fuglestad, B.; Stetz, M. A.; Gosse, T.; Tommos, C.; Wand, A. J., Improving yields of deuterated, methyl labeled protein by growing in H₂O. *J Biomol NMR* **2018**, 71 (4), 263-273.
11. Sprangers, R.; Kay, L. E., Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **2007**, 445 (7128), 618-22.
12. Wiesner, S.; Sprangers, R., Methyl groups as NMR probes for biomolecular interactions. *Curr Opin Struct Biol* **2015**, 35, 60-7.
13. Bordo, D.; Argos, P., Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J Mol Biol* **1991**, 217 (4), 721-9.
14. Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E., Cross-correlated relaxation enhanced ¹H[¹³C] NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **2003**, 125 (34), 10420-8.
15. Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E., Methyl TROSY: explanation and experimental verification. *Magn Reson Chem* **2003**, 41 (10), 843-852.

16. Gardner, K. H.; Kay, L. E., Production and incorporation of N-15, C-13, H-2 (H-1-delta 1 methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society* **1997**, *119* (32), 7599-7600.
17. Tugarinov, V.; Kay, L. E., An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* **2004**, *28* (2), 165-72.
18. Ruschak, A. M.; Velyvis, A.; Kay, L. E., A simple strategy for (1)(3)C, (1)H labeling at the Ile-gamma2 methyl position in highly deuterated proteins. *J Biomol NMR* **2010**, *48* (3), 129-35.
19. Gelis, I.; Bonvin, A. M.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G., Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **2007**, *131* (4), 756-69.
20. Isaacson, R. L.; Simpson, P. J.; Liu, M.; Cota, E.; Zhang, X.; Freemont, P.; Matthews, S., A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc* **2007**, *129* (50), 15428-9.
21. Ayala, I.; Sounier, R.; Use, N.; Gans, P.; Boisbouvier, J., An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR* **2009**, *43* (2), 111-9.
22. Velyvis, A.; Ruschak, A. M.; Kay, L. E., An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One* **2012**, *7* (9), e43725.
23. Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J., Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr Opin Struct Biol* **2015**, *32*, 113-22.
24. Gans, P.; Hamelin, O.; Sounier, R.; Ayala, I.; Dura, M. A.; Amaro, C. D.; Noirclerc-Savoye, M.; Franzetti, B.; Plevin, M. J.; Boisbouvier, J., Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl* **2010**, *49* (11), 1958-62.
25. Lichteneker, R. J.; Weinhaupl, K.; Reuther, L.; Schorghuber, J.; Schmid, W.; Konrat, R., Independent valine and leucine isotope labeling in Escherichia coli protein overexpression systems. *J Biomol NMR* **2013**, *57* (3), 205-9.
26. Lichteneker, R. J.; Coudeville, N.; Konrat, R.; Schmid, W., Selective isotope labelling of leucine residues by using alpha-ketoacid precursor compounds. *Chembiochem* **2013**, *14* (7), 818-21.

27. Mas, G.; Crublet, E.; Hamelin, O.; Gans, P.; Boisbouvier, J., Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J Biomol NMR* **2013**, *57* (3), 251-62.
28. Miyanoiri, Y.; Takeda, M.; Okuma, K.; Ono, A. M.; Terauchi, T.; Kainosho, M., Differential isotope-labeling for Leu and Val residues in a protein by *E. coli* cellular expression using stereo-specifically methyl labeled amino acids. *J Biomol NMR* **2013**, *57* (3), 237-49.
29. Kerfah, R.; Plevin, M. J.; Pessey, O.; Hamelin, O.; Gans, P.; Boisbouvier, J., Scrambling free combinatorial labeling of alanine-beta, isoleucine-delta1, leucine-proS and valine-proS methyl groups for the detection of long range NOEs. *J Biomol NMR* **2015**, *61* (1), 73-82.
30. Miyanoiri, Y.; Ishida, Y.; Takeda, M.; Terauchi, T.; Inouye, M.; Kainosho, M., Highly efficient residue-selective labeling with isotope-labeled Ile, Leu, and Val using a new auxotrophic *E. coli* strain. *J Biomol NMR* **2016**, *65* (2), 109-19.
31. Libich, D. S.; Tugarinov, V.; Clore, G. M., Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proc Natl Acad Sci U S A* **2015**, *112* (29), 8817-23.
32. Sekhar, A.; Rosenzweig, R.; Bouvignies, G.; Kay, L. E., Mapping the conformation of a client protein through the Hsp70 functional cycle. *Proc Natl Acad Sci U S A* **2015**, *112* (33), 10395-400.
33. Karagoz, G. E.; Duarte, A. M.; Akoury, E.; Ippel, H.; Biernat, J.; Moran Luengo, T.; Radli, M.; Didenko, T.; Nordhues, B. A.; Veprintsev, D. B.; Dickey, C. A.; Mandelkow, E.; Zweckstetter, M.; Boelens, R.; Madl, T.; Rudiger, S. G., Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. *Cell* **2014**, *156* (5), 963-74.
34. Neu, A.; Neu, U.; Fuchs, A. L.; Schlager, B.; Sprangers, R., An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nat Chem Biol* **2015**, *11* (9), 697-704.
35. Nishikawa, J. L.; Boeszoermenyi, A.; Vale-Silva, L. A.; Torelli, R.; Posteraro, B.; Sohn, Y. J.; Ji, F.; Gelev, V.; Sanglard, D.; Sanguinetti, M.; Sadreyev, R. I.; Mukherjee, G.; Bhyravabhotla, J.; Buhrlage, S. J.; Gray, N. S.; Wagner, G.; Naar, A. M.; Arthanari, H., Inhibiting fungal multidrug resistance by disrupting an activator-Mediator interaction. *Nature* **2016**, *530* (7591), 485-9.
36. Zhou, B. R.; Jiang, J.; Feng, H.; Ghirlando, R.; Xiao, T. S.; Bai, Y., Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell* **2015**, *59* (4), 628-38.

37. Drogemuller, J.; Strauss, M.; Schweimer, K.; Jurk, M.; Rosch, P.; Knauer, S. H., Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci Rep* **2015**, *5*, 16428.
38. Lou, Y. C.; Weng, T. H.; Li, Y. C.; Kao, Y. F.; Lin, W. F.; Peng, H. L.; Chou, S. H.; Hsiao, C. D.; Chen, C., Structure and dynamics of polymyxin-resistance-associated response regulator PmrA in complex with promoter DNA. *Nat Commun* **2015**, *6*, 8838.
39. Mazhab-Jafari, M. T.; Marshall, C. B.; Smith, M. J.; Gasmi-Seabrook, G. M.; Stathopoulos, P. B.; Inagaki, F.; Kay, L. E.; Neel, B. G.; Ikura, M., Oncogenic and RASopathy-associated K-RAS mutations relieve membrane-dependent occlusion of the effector-binding site. *Proc Natl Acad Sci U S A* **2015**, *112* (21), 6625-30.
40. Tokunaga, Y.; Takeuchi, K.; Takahashi, H.; Shimada, I., Allosteric enhancement of MAP kinase p38 alpha's activity and substrate selectivity by docking interactions. *Nature Structural & Molecular Biology* **2014**, *21* (8), 704-711.
41. McShan, A. C.; Kaur, K.; Chatterjee, S.; Knight, K. M.; De Guzman, R. N., NMR Identification of the Binding Surfaces Involved in the Salmonella and Shigella Type III Secretion Tip-Translocon Protein-Protein Interactions. *Proteins* **2016**.
42. Chaudhury, S.; de Azevedo Souza, C.; Plano, G. V.; De Guzman, R. N., The LcrG Tip Chaperone Protein of the Yersinia pestis Type III Secretion System Is Partially Folded. *J Mol Biol* **2015**, *427* (19), 3096-109.
43. Chaudhury, S.; Nordhues, B. A.; Kaur, K.; Zhang, N.; De Guzman, R. N., Nuclear Magnetic Resonance Characterization of the Type III Secretion System Tip Chaperone Protein PcrG of Pseudomonas aeruginosa. *Biochemistry* **2015**, *54* (43), 6576-85.
44. Saio, T.; Guan, X.; Rossi, P.; Economou, A.; Kalodimos, C. G., Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **2014**, *344* (6184), 1250494.
45. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776-9.
46. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **1998**, *95* (12), 6705-10.
47. Severinov, K.; Muir, T. W., Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J Biol Chem* **1998**, *273* (26), 16205-9.

48. Yamazaki, T.; Otomo, T.; Oda, N.; Kyogoku, Y.; Uegaki, K.; Ito, N.; Ishino, Y.; Nakamura, H., Segmental isotope labeling for protein NMR using peptide splicing. *Journal of the American Chemical Society* **1998**, *120* (22), 5591-5592.
49. Evans, T. C., Jr.; Benner, J.; Xu, M. Q., Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **1998**, *7* (11), 2256-64.
50. Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A., Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **2004**, *126* (9), 2670-1.
51. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **2003**, *72*, 249-89.
52. Berrade, L.; Camarero, J. A., Expressed protein ligation: a resourceful tool to study protein structure and function. *Cell Mol Life Sci* **2009**, *66* (24), 3909-22.
53. Muona, M.; Aranko, A. S.; Raulinaitis, V.; Iwai, H., Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro. *Nat Protoc* **2010**, *5* (3), 574-87.
54. Volkmann, G.; Iwai, H., Protein trans-splicing and its use in structural biology: opportunities and limitations. *Mol Biosyst* **2010**, *6* (11), 2110-21.
55. Aranko, A. S.; Oeemig, J. S.; Zhou, D.; Kajander, T.; Wlodawer, A.; Iwai, H., Structure-based engineering and comparison of novel split inteins for protein ligation. *Mol Biosyst* **2014**, *10* (5), 1023-34.
56. Aranko, A. S.; Wlodawer, A.; Iwai, H., Nature's recipe for splitting inteins. *Protein Eng Des Sel* **2014**, *27* (8), 263-71.
57. Thiel, I. V.; Volkmann, G.; Pietrokovski, S.; Mootz, H. D., An atypical naturally split intein engineered for highly efficient protein labeling. *Angew Chem Int Ed Engl* **2014**, *53* (5), 1306-10.
58. Stevens, A. J.; Brown, Z. Z.; Shah, N. H.; Sekar, G.; Cowburn, D.; Muir, T. W., Design of a Split Intein with Exceptional Protein Splicing Activity. *J Am Chem Soc* **2016**, *138* (7), 2162-5.
59. Zettler, J.; Eppmann, S.; Busche, A.; Dikovskaya, D.; Dotsch, V.; Mootz, H. D.; Sonntag, T., SPLICEFINDER - a fast and easy screening method for active protein trans-splicing positions. *PLoS One* **2013**, *8* (9), e72925.
60. Williams, F. P.; Milbradt, A. G.; Embrey, K. J.; Bobby, R., Segmental Isotope Labelling of an Individual Bromodomain of a Tandem Domain BRD4 Using Sortase A. *PLoS One* **2016**, *11* (4), e0154607.

61. Freiburger, L.; Sonntag, M.; Hennig, J.; Li, J.; Zou, P.; Sattler, M., Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *J Biomol NMR* **2015**, *63* (1), 1-8.
62. Mikula, K. M.; Tascon, I.; Tommila, J. J.; Iwai, H., Segmental isotopic labeling of a single-domain globular protein without any refolding step by an asparaginyl endopeptidase. *FEBS Lett* **2017**, *591* (9), 1285-1294.
63. Mikula, K. M.; Krumwiede, L.; Pluckthun, A.; Iwai, H., Segmental isotopic labeling by asparaginyl endopeptidase-mediated protein ligation. *J Biomol NMR* **2018**, *71* (4), 225-235.
64. Rosenzweig, R.; Farber, P.; Velyvis, A.; Rennella, E.; Latham, M. P.; Kay, L. E., ClpB N-terminal domain plays a regulatory role in protein disaggregation. *Proc Natl Acad Sci U S A* **2015**, *112* (50), E6872-81.
65. Mund, M.; Overbeck, J. H.; Ullmann, J.; Sprangers, R., LEGO-NMR spectroscopy: a method to visualize individual subunits in large heteromeric complexes. *Angew Chem Int Ed Engl* **2013**, *52* (43), 11401-5.
66. Kitevski-LeBlanc, J. L.; Prosser, R. S., Current applications of ¹⁹F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* **2012**, *62*, 1-33.
67. Ye, L.; Larda, S. T.; Frank Li, Y. F.; Manglik, A.; Prosser, R. S., A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in (1)(9)F NMR studies of proteins. *J Biomol NMR* **2015**, *62* (1), 97-103.
68. Edwards, J. M.; Derrick, J. P.; van der Walle, C. F.; Golovanov, A. P., (19)F NMR as a Tool for Monitoring Individual Differentially Labeled Proteins in Complex Mixtures. *Mol Pharm* **2018**, *15* (7), 2785-2796.
69. Matei, E.; Gronenborn, A. M., (19)F Paramagnetic Relaxation Enhancement: A Valuable Tool for Distance Measurements in Proteins. *Angew Chem Int Ed Engl* **2016**, *55* (1), 150-4.
70. Hattori, Y.; Heidenreich, D.; Ono, Y.; Sugiki, T.; Yokoyama, K. I.; Suzuki, E. I.; Fujiwara, T.; Kojima, C., Protein (19)F-labeling using transglutaminase for the NMR study of intermolecular interactions. *J Biomol NMR* **2017**, *68* (4), 271-279.
71. Ye, L.; Van Eps, N.; Zimmer, M.; Ernst, O. P.; Prosser, R. S., Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* **2016**, *533* (7602), 265-8.

72. Manglik, A.; Kim, T. H.; Masureel, M.; Altenbach, C.; Yang, Z.; Hilger, D.; Lerch, M. T.; Kobilka, T. S.; Thian, F. S.; Hubbell, W. L.; Prosser, R. S.; Kobilka, B. K., Structural Insights into the Dynamic Process of beta2-Adrenergic Receptor Signaling. *Cell* **2015**, *161* (5), 1101-11.
73. Yang, F.; Yu, X.; Liu, C.; Qu, C. X.; Gong, Z.; Liu, H. D.; Li, F. H.; Wang, H. M.; He, D. F.; Yi, F.; Song, C.; Tian, C. L.; Xiao, K. H.; Wang, J. Y.; Sun, J. P., Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and (19)F-NMR. *Nat Commun* **2015**, *6*, 8202.
74. Saxena, K.; Dutta, A.; Klein-Seetharaman, J.; Schwalbe, H., Isotope labeling in insect cells. *Methods Mol Biol* **2012**, *831*, 37-54.
75. Miyazawa-Onami, M.; Takeuchi, K.; Takano, T.; Sugiki, T.; Shimada, I.; Takahashi, H., Perdeuteration and methyl-selective (1)H, (13)C-labeling by using a Kluyveromyces lactis expression system. *J Biomol NMR* **2013**, *57* (3), 297-304.
76. Clark, L.; Zahm, J. A.; Ali, R.; Kukula, M.; Bian, L.; Patrie, S. M.; Gardner, K. H.; Rosen, M. K.; Rosenbaum, D. M., Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris. *J Biomol NMR* **2015**, *62* (3), 239-45.
77. Meola, A.; Deville, C.; Jeffers, S. A.; Guardado-Calvo, P.; Vasiliauskaite, I.; Sizun, C.; Girard-Blanc, C.; Malosse, C.; van Heijenoort, C.; Chamot-Rooke, J.; Krey, T.; Guittet, E.; Petres, S.; Rey, F. A.; Bontems, F., Robust and low cost uniform (15)N-labeling of proteins expressed in Drosophila S2 cells and Spodoptera frugiperda Sf9 cells for NMR applications. *J Struct Biol* **2014**, *188* (1), 71-8.
78. Opitz, C.; Isogai, S.; Grzesiek, S., An economic approach to efficient isotope labeling in insect cells using homemade 15N-, 13C- and 2H-labeled yeast extracts. *J Biomol NMR* **2015**, *62* (3), 373-85.
79. Sitarska, A.; Skora, L.; Klopp, J.; Roest, S.; Fernandez, C.; Shrestha, B.; Gossert, A. D., Affordable uniform isotope labeling with (2)H, (13)C and (15)N in insect cells. *J Biomol NMR* **2015**, *62* (2), 191-7.
80. Nygaard, R.; Zou, Y.; Dror, R. O.; Mildorf, T. J.; Arlow, D. H.; Manglik, A.; Pan, A. C.; Liu, C. W.; Fung, J. J.; Bokoch, M. P.; Thian, F. S.; Kobilka, T. S.; Shaw, D. E.; Mueller, L.; Prosser, R. S.; Kobilka, B. K., The dynamic process of beta(2)-adrenergic receptor activation. *Cell* **2013**, *152* (3), 532-42.
81. Kofuku, Y.; Ueda, T.; Okude, J.; Shiraishi, Y.; Kondo, K.; Mizumura, T.; Suzuki, S.; Shimada, I., Functional dynamics of deuterated beta2 -adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. *Angew Chem Int Ed Engl* **2014**, *53* (49), 13376-9.

82. Isogai, S.; Deupi, X.; Opitz, C.; Heydenreich, F. M.; Tsai, C. J.; Brueckner, F.; Schertler, G. F.; Veprintsev, D. B.; Grzesiek, S., Backbone NMR reveals allosteric signal transduction networks in the beta1-adrenergic receptor. *Nature* **2016**, *530* (7589), 237-41.
83. Okude, J.; Ueda, T.; Kofuku, Y.; Sato, M.; Nobuyama, N.; Kondo, K.; Shiraishi, Y.; Mizumura, T.; Onishi, K.; Natsume, M.; Maeda, M.; Tsujishita, H.; Kuranaga, T.; Inoue, M.; Shimada, I., Identification of a Conformational Equilibrium That Determines the Efficacy and Functional Selectivity of the mu-Opioid Receptor. *Angew Chem Int Ed Engl* **2015**, *54* (52), 15771-6.
84. Yoshiura, C.; Ueda, T.; Kofuku, Y.; Matsumoto, M.; Okude, J.; Kondo, K.; Shiraishi, Y.; Shimada, I., Elucidation of the CCR1- and CCR5-binding modes of MIP-1alpha by application of an NMR spectra reconstruction method to the transferred cross-saturation experiments. *J Biomol NMR* **2015**, *63* (4), 333-40.
85. Minato, Y.; Suzuki, S.; Hara, T.; Kofuku, Y.; Kasuya, G.; Fujiwara, Y.; Igarashi, S.; Suzuki, E.; Nureki, O.; Hattori, M.; Ueda, T.; Shimada, I., Conductance of P2X4 purinergic receptor is determined by conformational equilibrium in the transmembrane region. *Proc Natl Acad Sci U S A* **2016**, *113* (17), 4741-6.
86. Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Guntert, P., Optimal isotope labelling for NMR protein structure determinations. *Nature* **2006**, *440* (7080), 52-7.
87. Rosenblum, G.; Cooperman, B. S., Engine out of the chassis: cell-free protein synthesis and its uses. *FEBS Lett* **2014**, *588* (2), 261-8.
88. Etzkorn, M.; Raschle, T.; Hagn, F.; Gelev, V.; Rice, A. J.; Walz, T.; Wagner, G., Cell-free expressed bacteriorhodopsin in different soluble membrane mimetics: biophysical properties and NMR accessibility. *Structure* **2013**, *21* (3), 394-401.
89. Michel, E.; Skrisovska, L.; Wuthrich, K.; Allain, F. H., Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology. *Chembiochem* **2013**, *14* (4), 457-66.
90. Linser, R.; Gelev, V.; Hagn, F.; Arthanari, H.; Hyberts, S. G.; Wagner, G., Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. *J Am Chem Soc* **2014**, *136* (32), 11308-10.
91. Lazarova, M.; Lohr, F.; Rues, R. B.; Kleebach, R.; Dotsch, V.; Bernhard, F., Precursor-Based Selective Methyl Labeling of Cell-Free Synthesized Proteins. *ACS Chem Biol* **2018**, *13* (8), 2170-2178.

Chapter 3. Variant H2A.B-H2B histone dimer is more stable than the canonical dimer

This chapter is based on:

Heyi Zhang, Sofie Dubbeldam, Hugo van Ingen. Variant H2A.B-H2B histone dimer is more stable than the canonical dimer. *Submitted*.

Contributions of authors:

Buffer screening assays for thermostability comparison between histone dimers were carried out by Sofie Dubbelman.

Abstract

Histone variants incorporated nucleosomes with specific structural and functional properties offer the opportunity to regulate chromatin biology. Histone variant H2A.B is one of the most divergent histone variants known to date and has been shown to be associated with regulation of gene transcription, DNA replication, and RNA splicing. Compared to those with the canonical H2A, nucleosomes containing H2A.B display a ‘loosened’ conformation in which the entry and exit DNA are partly unwrapped from the histone octamer core. In this study, we studied the structure, dynamics and stability of the H2A.B-H2B heterodimer with solution NMR. Using backbone chemical shift and sparse NOE data in combination with CS-Rosetta modelling, we determined the solution structure of H2A.B-H2B. The dimer shows the canonical histone fold with highly flexible and disordered tails. The solution structure of the histone core is well-defined and matches the recently determined H2A.B-H2B crystal structure. The H2A.B-H2B core has an overall reduced electrostatic surface potential, compared to the canonical H2A-H2B dimer. Based on comparative thermal stability assays on a series of mutants, we show that H2A.B-H2B dimer has increased thermostability compared to the canonical dimer, and that stability correlates well with overall net charge of the dimer core. Structural analysis highlights that the distinct electrostatic properties of H2A.B-H2B dimer not only destabilize DNA binding but may also weaken the interaction with the H3-H4 tetramer. The increased dimer stability may facilitate the dynamic and chaperone-free exchange of the H2A.B-H2B dimer from the nucleosome.

Introduction

Eukaryotic cells have evolved a number of mechanisms to regulate the basic processes that need access to the DNA, such as transcription, replication and repair. Next to chromatin remodeling and the installation of post-translation modifications, the incorporation of variant histone proteins is a key mechanism to confer specific functional properties to chromatin at defined loci. These histone variants replace the abundant, canonical form of H2A, H2B, H3 or H4 histones that make up the histone octamer core in the nucleosome, the basic unit of chromatin. Histone variants, which unlike canonical histones can be expressed throughout the whole cell cycle, have been identified for all four core histones.^{1,2,3} Due to their different amino acid sequence, histone variants can alter nucleosome structure and dynamics, promote or inhibit nucleosome-protein interactions, or provide different post-translational modification sites to regulate chromatin function.^{4,5,6} For example, H3 variant CENP-A is essential for centromere identity,⁷ while histone H2A variant H2A.X signals DNA damage by phosphorylation on its C-terminal serine which as a result recruits and accumulates DNA repair proteins to damage sites.⁸ H2A.B is one of the most divergent H2A histone variants,⁹⁻¹⁰ sharing only 48% sequence identity to the canonical H2A.¹¹ While most sequence differences with the canonical H2A occur in the N- and C-terminal tails, there are also a considerable number of substitutions in the core of the protein: 36 for the human variant and 45 for the mouse variant (Figure 3.1a). The majority of these substitutions involve reversal, loss or addition of charge. Notably, sequence conservation for H2A.B orthologues is significantly lower than for canonical H2A or H2B (Figure 3.1a, 1b).

As discussed in Chapter 1, a wide array of biochemical and low-resolution structural methods have been used to show that the DNA ends in H2A.B-incorporated nucleosomes are transiently unwrapped from the histone octamer, resulting in the stable association of 110-130 bp instead of 147bp DNA as in the conventional nucleosomes¹²⁻¹⁴. In addition, H2A.B nucleosome arrays are less compact and may thus improve transcription efficiency and facilitate DNA replication and repair^{14-20,21}. Despite the importance of H2A.B in these essential cell functions, the molecular basis of how this variant confers its special

properties to the nucleosome is not fully resolved. The partly neutralized “acidic patch”, a key interaction site for proteins that bind the canonical nucleosome²² and site of nucleosome-nucleosome interactions, as well as the near-absence of lysine residues, which are main sites for post-translational modifications (see also Figure 3.1a), impact the functional properties of the H2A.B nucleosome. Nevertheless, these changes are unlikely to affect nucleosome structure directly. In this regard, previous studies have highlighted the impact of the altered and truncated C-terminal docking domain (see Figure 3.1a) of H2A.B on its less compact structure of nucleosomes^{12-13, 23-24}. The histone fold core of H2A.B-H2B may thus additionally contribute to peculiar properties of the H2A.B nucleosome.

Recently, the crystal structure of the H2A.B-H2B dimer core was solved, which showed that dimer has the canonical histone fold and significantly reduced positive charge on its DNA binding surface compared to the canonical dimer.²⁵ For canonical H2A-H2B, the solution structure showed overall the same fold as in the nucleosome but with considerable structural heterogeneity, suggesting that intrinsic structural flexibility of the H2A-H2B dimer facilitates its association with DNA and other histone partners into nucleosomes.²⁶ The extent of intrinsic flexibility in H2A.B-H2B is unknown but may give clues to understand the destabilizing influence on the nucleosome and the increased exchange dynamics of this dimer.

Here, we studied the structure, dynamics and stability of the H2A.B-H2B dimer in solution. Using chemical shift and sparse NOE data based Rosetta protein folding, the solution structure of the H2A.B-H2B dimer was determined. We find that H2A.B-H2B folds into the canonical histone fold with a well-defined core region and disordered, highly flexible N- and C-terminal extensions, including part of the docking domain. The H2A.B $\alpha 2$ helix is bent in solution as was observed in the crystal. The structure further shows that the interaction surface with H3-H4 has an altered electrostatic potential that may contribute to their looser association in addition to the truncation. Surprisingly, we find that the variant dimer has significantly increased thermal stability compared to the canonical dimer. Through a mutational analysis we show that reduced electrostatic repulsion between H2A.B and H2B monomers causes enhanced stability, which may contribute to the increased exchange dynamics for H2A.B.

Materials and methods

Histone protein production and refolding. *Drosophila melanogaster* (*Dm.*) canonical histones H2A (Uniprot-id: P84051) and H2B (Uniprot-id: P02283) and *Homo sapiens* (*Hs.*) H2A.B (Uniprot: P0C5Z0) were expressed in *E. coli* BL21 Rosetta 2 (DE3) cells (Novagen) and purified under denaturing conditions from inclusion bodies by extraction in 6 M guanidine chloride, followed by size-exclusion chromatography in buffer A (7 M urea, 50 mM NaPi, 1 mM EDTA, 150 mM NaCl, pH 7.5) using a Superdex 75 column (GE) and ion exchange with a salt gradient from buffer A to buffer A supplemented with 1 M NaCl.²⁷ Histones used for NMR studies were produced in D₂O or H₂O-based M9 minimal medium containing desired isotopes. Histones used for thermal shift assays were produced in LB medium. Histone dimers were refolded from equimolar mixes of denatured purified histones by dialysis to 2 M NaCl at room temperature and subsequent purification using size-exclusion chromatography over a Superdex 200 column (GE) in 2M NaCl buffer.²⁷ Purified dimers were stored at 4 °C before buffer exchange to lower salt concentration for NMR or thermal stability studies.

Site-directed mutagenesis. Site-directed mutagenesis was performed on the pET-21b plasmid containing the *Drosophila* H2A gene or on the pET-3a plasmid containing human H2A.B gene. For each mutation, a pair of complementary mutagenic primers was designed and used to amplify the entire plasmid in a thermocycling reaction using in-house produced Pfu DNA polymerase. The template plasmid was digested with DpnI after reaction, and the generated nicked circular DNA was used to transform *E.coli* DH5 α cells. Successful transformations were used to extract plasmids after which correctness of mutations was verified by sequencing.

NMR spectroscopy. All NMR experiments were performed at 35 °C on a Bruker Avance III spectrometer operating at 20.0 (21.1) T corresponding to 850 (900) MHz ¹H Larmor frequency, equipped with a cryo-probe. Samples for H2A.B and H2B backbone assignment contained ~0.15 mM H2A.B-H2B refolded from either fractionally deuterated, uniformly ¹⁵N/¹³C-labeled H2A.B and unlabeled H2B or

vice versa. H2A.B data was collected in NMR buffer 1 (20 mM MES, 50 mM NaCl, pH 6, 0.1 mM PMSF, 5 mM DTT), while H2B data was collected at higher salt levels to promote stability of the complex (NMR buffer 2: 20 mM NaPO₄, 200 mM NaCl, pH 6.5, 5 mM β -mercaptoethanol). ¹H-¹⁵N HSQC fingerprint spectra showed negligible chemical shift differences between these conditions. Samples for measurement of ¹⁵N T₁ and T₂ relaxation times contained 0.15-0.3 mM dimer with either ¹⁵N-labeled H2A.B or H2B.

Backbone resonances for H2A.B and H2B in the H2A.B-H2B heterodimer were assigned using standard sequential backbone assignment approach based on the set of 3D TROSY versions of HNCO, HN(CA)CO, HNCACB, HNCA, HNCB, HN(CO)CA experiments. In total, 89% and 91% of the H_N, N, C _{α} , C _{β} , C' backbone resonances were assigned for the entire sequence of H2A.B and H2B respectively. Backbone resonances for the core regions of H2A.B (V31-F101) and H2B (Y34-K122) were assigned to 99% and 93% completion, respectively. No chemical shift assignment could be obtained for the first 8 residues (PRRRRRRG) and C19 in the N-terminal tail together with S28, F29 of H2A.B, and P1, S5, G6 in the N-terminal tail together with S53 in the core and T119, S120 in the C-terminus of H2B, either due to extensive overlap or missing connectivity information.

A 3D ¹⁵N-edited NOESY with 200 ms mixing time was recorded on a sample containing 0.38 mM dimer in NMR buffer 2. The dimer was refolded from perdeuterated ¹⁵N-labeled H2B and unlabeled H2A.B to selectively record intermolecular NOEs.

All NMR data were processed using Bruker Topspin, or NMRPipe²⁸ and analyzed using NMRFAM-Sparky²⁹. Secondary structure of H2A.B-H2B dimer was predicted by TALOS-N using H_N, N, C _{α} , C _{β} , C' chemical shifts.³⁰

Structure calculations. Calculation of the structure of the H2A.B-H2B dimer was carried out in a two-step approach. First, an initial model based only on backbone chemical shifts was obtained using CS-Rosetta.^{31,32} Then, this model was used to assign intermolecular NOEs between H2A.B and H2B. These NOEs were subsequently used as additional constraints in a final CS-Rosetta structure calculation, after which modeling of the N- and C-terminal tails yielded the final solution

structure of H2A.B-H2B. For the first step, the H_N , N, C_α , C_β , C' backbone chemical shifts of the core regions of H2A.B (V31-F101) and H2B (Y34-K122) were used to calculate 3000 structures of the H2A.B-H2B histone fold core using the CS-Rosetta webserver (<https://csrosetta.bmr.b.wisc.edu/csrosetta/submit>). To allow Rosetta to fold the dimeric core, the two proteins were connected by a random coil (Gly)₈ poly-glycine linker into a single chain. From the 20 lowest energy models, the 10 structures with the lowest C_α RMSD to the lowest energy model were selected as the best models.

The initial CS-Rosetta ensemble was used to semi-automatically assign intermolecular NOEs in the ^{15}N -edited NOESY recorded on a H2A.B- $[\text{U-}^2\text{H}, ^{15}\text{N}]$ -H2B sample using an in-house written python script. All H2B H_N protons were assigned based on backbone (H_N , N) chemical shifts. One intermolecular NOE could be assigned unambiguously based on the backbone assignments as an H_N - H_N contact. For all other observed NOEs, we relied on chemical shifts statistics, correspondence to the initial structural model, and networking support to assign the H2A.B proton, in a procedure akin to that implemented in CYANA.³³ Briefly, candidate assignments were compiled based on correspondence to the atom and residue specific average chemical shift as deposited in the BMRB using a chemical shift tolerance of 0.4 ppm. These candidate assignments were then filtered according to their match to the structural model, requiring that the candidate NOE was satisfied in at least 60% of models in the initial CS-Rosetta ensemble using a distance cutoff of 7 Å for intense peaks ($S/N \geq 10$) and 8 Å for weaker peaks ($S/N < 10$). This resulted in 35 unambiguously assigned NOEs in which there was only one assignment possibility. These assignments were used as anchors to extend the assignments in a network approach in which cross peaks with ^1H chemical shifts within 0.01 ppm were assigned to same H2A.B proton wherever structurally supported, resulting in 44 assigned intermolecular NOEs in total. For 20 intense cross peaks ($S/N \geq 10$) multiple assignment possibilities were obtained and could thus not be assigned.

The NOEs were converted to distance constraints based on a calibration of H2B backbone H_N - H_N NOEs and average helical i , $i+1$, $i+2$, $i+3$ distances. Each NOE was converted into one of five categories of upper distance restraints: ≤ 2.4 , 2.4–3, 3–4, 4–5 and 5–7 Å.

Intensities of NOEs involving methyl protons were divided by three per methyl group before calibration.

In final step of the structure calculation, the distance restraints derived from the intermolecular NOEs were used together with the chemical-shift generated fragments to calculate the structure of the H2A.B-H2B core region (H2A.B V31-F101 and H2B Y34-K122 connected by a poly-glycine linker) using the Rosetta *AbInitio* protocol as described previously³⁴

(www.rosettacommons.org/demos/latest/public/abinitio_w_chemicals_hift_noe/README). In total, 8000 structures were generated and scored on the basis of their the full-atom and distance restraint energy (E) by Rosetta. For each model, backbone chemical shifts of $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{13}C , $^1\text{H}_\text{N}$ and ^{15}N were predicted by SPARTA+. ³⁵ The correspondence between predicted and experimental chemical shifts was used to rescore the initial Rosetta energy to a rescored energy (E') according to:

$$E' = E + c \times \chi_{CS}^2$$

where c is a weighting factor set to 0.25, and

$$\chi_{CS}^2 = \sum_i \sum_j (\delta_{i,j}^{exp} - \delta_{i,j}^{pred})^2 / \sigma_{i,j}^2$$

where $\delta_{i,j}^{exp}$ is the backbone chemical shift observed from NMR experiments for each atom type i ($^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{13}C , $^1\text{H}_\text{N}$, ^{15}N) for a given residue j ; and $\delta_{i,j}^{pred}$ is the backbone chemical shift predicted by SPARTA+; and $\sigma_{i,j}^2$ is the uncertainty of $\delta_{i,j}^{pred}$. The model with the lowest E' was selected as the reference model, and C_α -RMSD of each model to the reference model was calculated. The 10 models with the lowest C_α -RMSD from the 20 lowest energy models were selected as the final structure ensemble for the histone fold core. Finally, the poly-glycine linker was removed and disordered N-and C-terminal tails of H2A.B and H2B were added using MODELLER³⁶ to result in the final solution structure of H2A.B-H2B.

Electrostatics were calculated using the adaptive Poisson-Boltzmann solver³⁷ using the lowest-energy structure. All structure images were created using open source PyMOL (The PyMOL Molecular Graphics System, Version 1.7, Schrödinger, LLC).

Thermal stability assays. A buffer series with pH range of 6 to 8 and salt concentrations from 50 mM to 250 mM NaCl was made to test thermo-stability of H2A.B-H2B dimer. For each well on 96-well plate, final dimer concentration was 5 μ M (corresponding to a 30-fold dilution from the stock) with 1000-fold diluted Sypro Orange (Sigma-Aldrich) in a 25 μ L final volume. Thermo-stability of mutant dimers (~10-20-fold diluted from stock) were measured with same combinations in 20 mM phosphate buffer at pH 6.5 with 200 mM NaCl. Measurements were performed in duplicate with the Biorad CFX96 Real-Time System. The temperature gradient was increased from 20°C to 80°C at 1°C/ minute and the fluorescence signal was recorded using FRET channel. The melting temperature was derived from the maximum of the first derivative of the absorbance signal using the Bio-Rad CFX manager program.

Results

H2A.B-H2B heterodimer has the canonical histone-fold core in solution. To study structure and dynamics of H2A.B-H2B in solution, we refolded the heterodimer using either isotope-labeled H2A.B with unlabeled H2B or vice versa and studied these using NMR spectroscopy. Of note, in this study we used *Drosophila* H2B that carries four conservative substitutions in the core region compared to the human sequence (Figure 3.1b). Both H2A.B and H2B give rise to high-quality TROSY fingerprint spectra, confirming that the dimers are well-folded (Figure S3.1). Near-complete backbone assignment could be obtained for both histones. Chemical shift indices (CSI) obtained from the experimental C α and C β chemical shifts confirm the presence of the characteristic histone fold, a central long α -helix flanked by two shorter helices, for both histones (Figure 3.1c). In addition, the C-terminal α C helix of each histone is present together with the two β -strands that form two intermolecular β -sheets. The CSI further indicates that the N-terminal tail of both H2A.B and H2B and C-terminal end of H2A.B docking domain (F101-D114) are disordered in solution. Overall, the secondary structure matches well to the crystal structure of the variant H2A.B-H2B heterodimer²⁵. Interestingly, in solution the dimer lacks the N-terminal α N-helix that is observed in

the H2A.B-H2B crystal. Similarly, the α N-helix is also unfolded in the canonical H2A-H2B heterodimer in solution, while it is a defined structural element in crystal structures of the nucleosome^{23,26,38}. Secondary structure prediction by TALOS-N³⁰ based on the experimental backbone chemical shift values confirmed the observations from CSI, albeit that in both H2A.B and H2B lower propensities for the presence of the β 2 strand were found (Figure 3.1c). Overall, these data indicate that the H2A.B-H2B dimer forms a well-folded histone fold domain with disordered tails. It further suggests the H2A.B α N helix is a labile secondary structure element that can be stabilized by contacts to the DNA, as seen for canonical H2A, or by crystal contacts as seen in the H2A.B-H2B crystal structure.

Backbone chemical shifts do not suffice to define the H2A.B-H2B core structure. Initially, we followed the approach used by Moriwaki *et al.*²⁶ to calculate the structure of the dimer based on backbone chemical shifts using CS-Rosetta. This approach relies on the Rosetta protein folding engine, supplemented by chemical shift based selection of protein fragments that are assembled into a 3D fold and followed by chemical shift based scoring of the final solutions. Using the experimental H_N , N, C_α , C_β , and C' backbone chemical shifts, 3000 models for the H2A.B-H2B core were generated using the BMRB CS-Rosetta web server. While the calculation converged, superposition of the 10 models with the lowest C_α -RMSD from the lowest energy 20 models showed that the H2A.B α 1 and H2B α C helices are ill-defined within the otherwise reasonably defined histone fold-core (Figure S3.2). While such structural heterogeneity could be intrinsic feature of the H2A.B-H2B dimer, it could also be the result of limitations in the modeling approach and input data.

We thus decided to refine this initial model with sparse distance restraints obtained from an intermolecular NOESY experiment. Exploiting the *in vitro* reconstitution of the H2A.B-H2B complex, we refolded dimers from unlabeled H2A.B and perdeuterated and ^{15}N -labeled H2B. This labeling scheme ensures that only intermolecular NOEs will be observed in a standard ^{15}N -NOESY experiment, with the obvious exception of intramolecular H_N - H_N NOEs. To assign these intermolecular NOEs to their corresponding ^1H atoms in H2A.B, we took advantage of the initial CS-Rosetta model as a structural reference

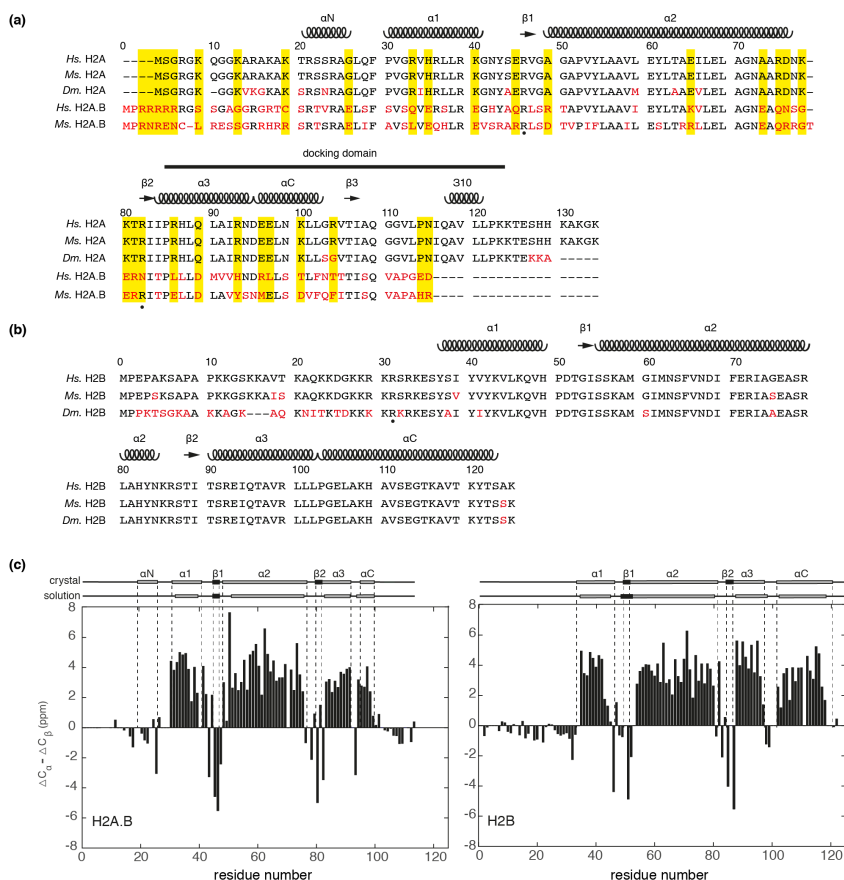


Figure 3.1. Sequence analysis and secondary structure of the H2A.B-H2B heterodimer. (a, b) Sequence alignment of human, mouse, and drosophila H2A, H2A.B (a), and H2B (b) sequences; *Hs.* – *Homo sapiens*; *Mm.* – *Mus musculus*; *Dm.* – *Drosophila melanogaster*. Sequence differences from *Hs.* are highlighted in red. Secondary structure in the human canonical nucleosome (PDB-id: 3afa) is indicated above the sequence. Minor groove arginines are marked with a black dot. In (a) substitutions involving charged residues are highlighted in yellow. (c). Chemical shift indices based on C α and C β chemical shifts for H2A.B and H2B in the H2A.B-H2B heterodimer, confirming the presence of the histone-fold in solution. Secondary structure elements in the crystal structure of the H2A.B-H2B dimer (PDB-id 6a7u) and as predicted by TALOS-N are indicated above the bar graph and labeled. Dashed lines are drawn to indicate the beginnings and ends of each secondary structure element observed in the crystal structure.

in a semi-automatic assignment procedure akin to that implemented in CYANA³³ (see Materials and Methods).

This allowed the unambiguous assignment of 44 intermolecular NOEs, among which 8 involve the H2B α C helix, and 36 involve the converged core region of the dimer. For example, clear intermolecular H_N - H_α and H_N - H_β contacts are observed for H2B residues G50 and S52, indicating formation of β -sheet between the H2B β 1-strand and H2A.B (Figure 3.2a,b). Of interest, several clear intermolecular NOEs were observed for residues in the H2B α C helix, including H106 and A107, indicating the structure of the H2B α C helix is well-defined within the H2A.B-H2B heterodimer (Figure 3.2c,d). Unfortunately, no intermolecular NOEs to residues in the H2A.B α 1 helix could be established, likely because it is positioned relatively far from H2B H_N atoms. The extent of intermolecular NOEs observed for the H2B α C helix demonstrate that it has a well-defined position in the core of the dimer. We thus conclude that backbone chemical shifts do not suffice to properly define the H2A.B-H2B structure.

Solution structure of the H2A.B-H2B heterodimer highlights altered electrostatic potential. The intermolecular NOEs and backbone chemical shifts were combined in a Rosetta calculation according to established protocols (see Materials and Methods). The calculation converged well (Figure S3.3), resulting in a final ensemble of 10 structures of the H2A.B-H2B core (Figure 3.3a). Structural statistics are reported in Supplementary Table S3.1. The histone-fold core as well as the orientation of the H2B α C-helix are well-defined (heavy atom backbone RMSD 1.1 ± 0.2 Å). Inclusion of the NOE data resulted in both local and overall better definition of the structure compared to using backbone chemical shifts only (Figure S3.4). The position of H2A.B α 1-helix within the core is somewhat less well-defined. Since backbone dynamics of this helix, as judged from ^{15}N T_1/T_2 ratios, are not significantly different from the rest of the core (Figure S3.5), this is most likely due to the lack of intermolecular NOE data. Notably, these data further support that the H2A.B docking domain is flexible and unstructured after the α C helix.

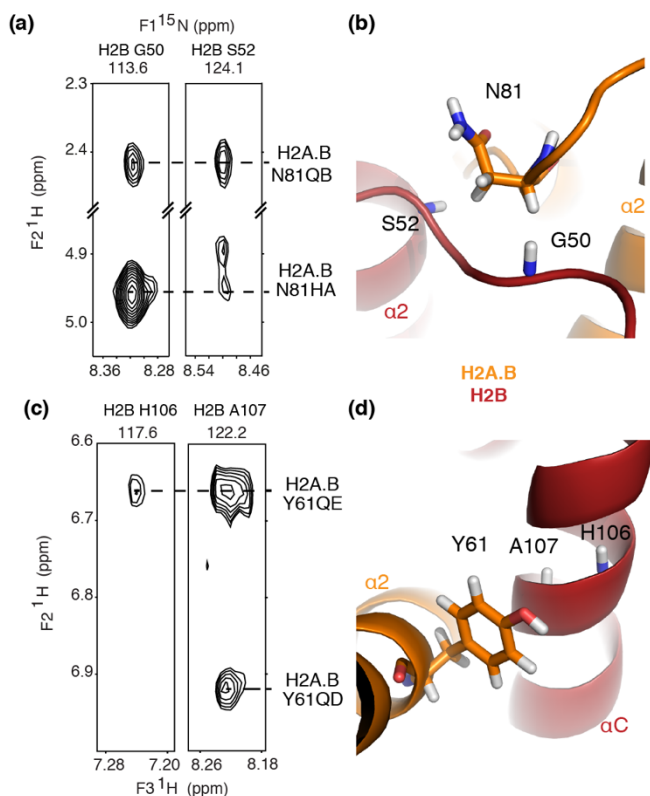


Figure 3.2. Intermolecular NOEs between H2A.B and H2B within the histone core support a well-defined conformation. (a, c) Strips from the 3D ^{15}N -edited NOESY recorded on heterodimers with perdeuterated ^{15}N -labeled H2B and unlabeled H2A.B showing intermolecular NOEs between H2B H_N and H2A.B protons. Assignment of H2B H_N and H2A.B resonances are indicated. (b, d) Zoom on the lowest-energy CS-Rosetta model corresponding to the intermolecular contact shown in the NOESY spectrum. NOEs shown in (a) restrain the H2A.B $\beta 2$ -H2B $\beta 1$ β -strand, NOEs in (c) restrain the H2B αC -helix position.

The solution structure agrees well to the recent crystal structure of the H2A.B-H2B core with backbone RMSD of 1.7 Å (Figure 3.3b). The H2A.B $\alpha 2$ helix is slightly bent (average bend angle $25 \pm 8^\circ$ over the whole ensemble), as was observed also in the crystal structure (20° bend angle). While Dai *et al.* hypothesized the bend may be due to the crystal packing, its observation here demonstrates it is an intrinsic property of the dimer. Dai *et al.* have pointed out that this bend is less prominent in the canonical H2A structure in the nucleosomal

conformation, causing a slight change in the position of the DNA-binding L2 loop. Careful analysis shows however that, while the $\alpha 2$ -helix in canonical H2A is indeed less bent (17° bend angle), slight changes in the relative position of the $\alpha 2$ helix within the dimer core also contribute to this difference in position. This is also apparent from comparison of the solution H2A.B-H2B structure with the canonical H2A-H2B structure in the nucleosome (Figure 3.3c). Interestingly, the solution structure of canonical H2A-H2B dimer also shows an increased bend angle for the H2A $\alpha 2$ helix ($32 \pm 9^\circ$), suggesting that the quaternary state of the H2A-H2B dimer, whether it is an isolated dimer or incorporated into nucleosomes, is a greater influence on the structure than the H2A variant type.

The most striking difference between the solution and crystal structure of H2A.B-H2B is the absence of the H2A.B αN helix in solution (see Figure 3.3b). Compared to canonical H2A, most amino acid substitutions occur on the outward facing surface of the core. Notably, one of the major structural differences between the canonical and variant dimer was shown by Dai *et al.* to be the substitution of E92 in H2A to L96 in H2A.B, which disrupts the hydrogen bonding to H2B residues E105 and L106.²⁵ A changed hydrogen bonding in solution is supported by the large amide backbone chemical shifts perturbations that are observed for the corresponding H2B residues between canonical and variant dimer (Figure S3.6). In particular for the glutamate a large upfield shift of the $^1\text{H}_\text{N}$ resonance is observed, from 10.12 ppm in the canonical dimer to 9.15 ppm in the variant dimer, supporting reduced hydrogen bonding.

For sake of completeness, we modelled the disordered tails to the structure of the folded core, highlighting the length of these extension compared to the globular core (Figure 3.4). Backbone dynamics derived from ^{15}N relaxation experiments support the highly flexible and unstructured nature of the tails compared to the relatively rigid folded-core (Figure S3.5).

Dai *et al.* highlighted the reduced electropositive charge at the DNA binding surface of the H2A.B-H2B core in their crystal structure²⁵, which is also apparent in solution (Figure 3.5a). In addition, a reduced negative potential is obvious for the region between the $\alpha 2$, $\alpha 3$ helix of H2A.B and αC of H2B, which forms the acidic patch on the canonical dimer (Figure 3.5b). Several charge neutralizing or charge

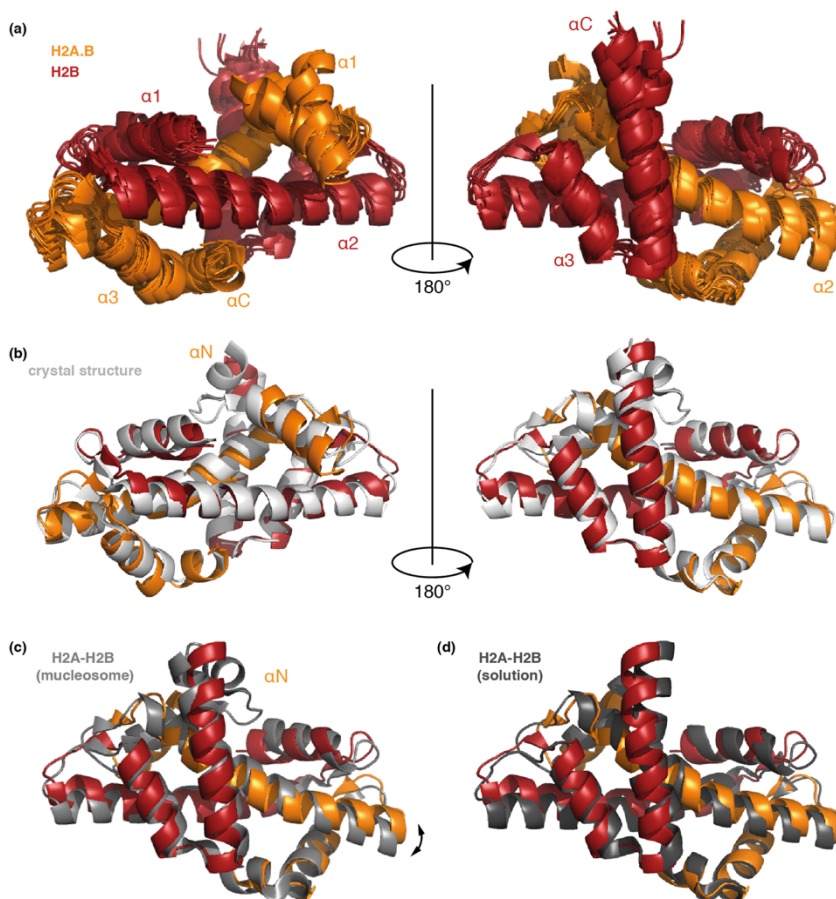


Figure 3.3. Solution structure of the H2A.B-H2B core. (a) Superposition of the final ensemble of 10 structural models calculated by NOE-supplemented CS-Rosetta. Color coding indicated in the figure; most prominent secondary structure elements are labeled. (b) Comparison of the H2A.B-H2B solution and crystal structure (PDB-id 6a7u). The bent H2A.B $\alpha2$ helix is indicated. (c) Comparison of the variant dimer solution structure and the canonical H2A-H2B dimer from nucleosome crystal structure (PDB-id 3afa) highlighting the difference in H2A.B $\alpha2$ orientation. (d) Comparison of the H2A.B-H2B and H2A-H2B solution structures (PDB-id 2rvq). Comparisons in (b), (c), and (d) were made using the closest-to-mean structure as the best representative structure of the final H2A.B-H2B ensemble. Structures were superimposed on heavy backbone atoms in the core region of the dimer H2A.B V31-L100 and H2B Y34-S121. Heavy atom backbone RMSDs for the superpositions shown in (b), (c) and (d) are 1.7, 1.8 and 1.9 Å, respectively.

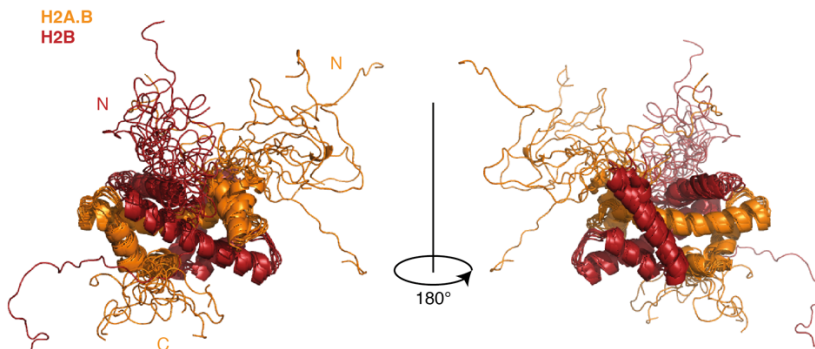


Figure 3.4. Solution structure of the H2A.B-H2B heterodimer including N- and C-terminal tails. Color coding as in Figure 3.3. The position of the N- and C-terminus is indicated.

reversal substitution in H2A.B (H2A E61, E91, E92 to H2A.B K65, R95, L96) significantly reduce the net charge of this patch. Closer inspection of the surface potential reveals that the H2A.B-H2B dimer exhibits a unique acidic surface on its H3-H4 interacting surface, which is positively charged in the canonical dimer. The change to a net negative charge is the result of substitution of Q84 and R88 in the H2A $\alpha 3$ helix with D88 and H92 in H2A.B. Within the canonical octamer, H2A R88 is involved in a long-range electrostatic interaction with H3 E94 and E105 on the adjacent H3-H4 dimer (Figure 3.5c). Within an H2A.B-nucleosome, this electrostatic interaction would either be neutralized or transformed into a repulsive interaction, depending on the charge-state of H92 (Figure 3.5d). These changes may thus contribute to the destabilization of the H2A.B containing octamers and nucleosomes. Previous studies using H2A and H2A.B domain-swapped proteins showed that a H2A chimera including the H2A.B $\alpha 3$ helix and the rest of the docking domain does not allow refolding of the histone octamer using salt dialysis¹².

Decreased net positive charge of the H2A.B-H2B core results in higher thermostability. We established that the H2A.B-H2B dimer retains a well-folded core in solution and has distinctly altered electrostatics compared to the canonical dimer. Such changes in surface electrostatics have been shown to impact protein stability even

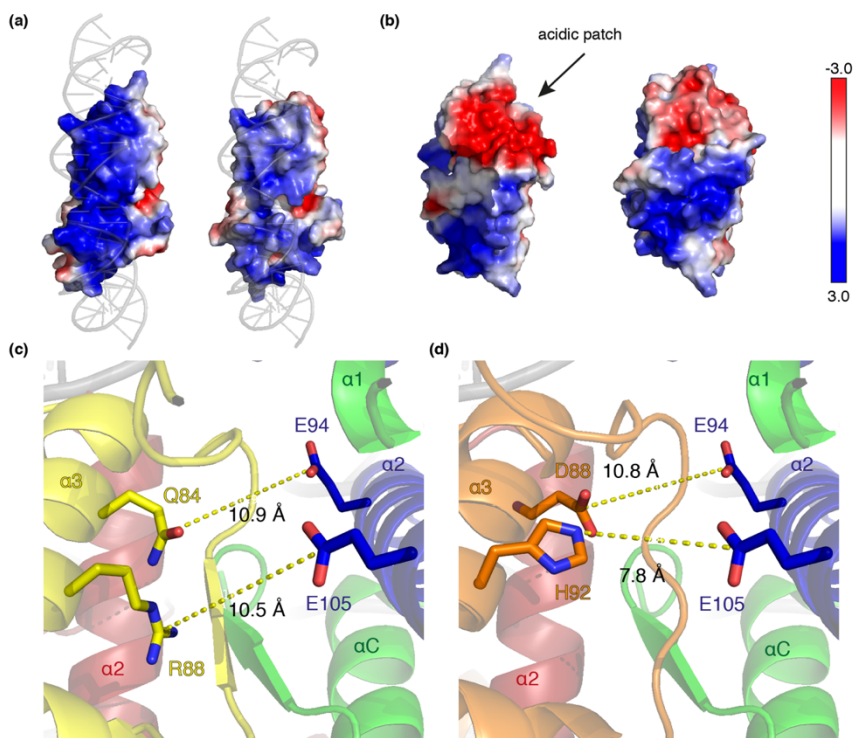


Figure 3.5. Electrostatic surface of H2A.B-H2B dimer compared to canonical H2A-H2B dimer. (a, b) View on the DNA binding surface (a) and acidic patch region (b) of H2A-H2B (left) and the H2A.B-H2B solution structure (right) with colour coding according to the electrostatic potential highlighting the reduced positive charge on the variant. The nucleosomal DNA is traced in light grey. Colour bar indicates the electro potential scale. (c) Zoom on the canonical nucleosome structure (PDB-id 3afa) focussing on an attractive electrostatic interaction in the interface between H2A-H2B and H3-H4. Selected side chains are shown as sticks. Colour coding: H2A – yellow; H2B – red; H3 – blue; H4 – green. (d) Same view as in (c) on a model of a H2A.B-containing nucleosome. H2A.B shown in orange. The docking domain C-terminal extension of H2A.B was modelled in the nucleosomal conformation of canonical H2A.

when these do not affect packing of the hydrophobic core³⁹. Given that the structure and packing of the hydrophobic core of the variant dimer is highly similar to the canonical dimer, we hypothesized that the altered electrostatics may impact the stability of the dimer. Variant and canonical heterodimer stability was assessed by measuring their unfolding temperatures in a thermostability assay (TSA). Over a wide range of buffer conditions with varying pH and ionic strengths, we

observed that the variant dimer unfolds at significantly elevated temperatures compared to the canonical dimer (Figure 3.6a,b). H2A.B-H2B dimer unfolds at around 63 °C and H2A-H2B dimer unfolds at around 50 °C. While both types of dimers prefer higher salt concentrations (250 mM NaCl) over lower salt concentration (50 mM NaCl), H2A-H2B shows a stronger salt-induced stabilization effect than H2A.B-H2B. The canonical H2A-H2B dimer is the most stable in pH 7 buffers while the variant dimer prefers pH 6. For canonical H2A-H2B the melting temperatures and their buffer dependence compare well to previous data obtained from dimers isolated from chicken erythrocytes⁴⁰.

As is apparent from the electrostatic potentials displayed in Figure 3.5a, the core region of H2A.B carries overall fewer positively charged residues than H2A (7 vs. 12, respectively). The net charge of the core region is -1 for H2A.B and +4 for H2A, while the net charge for H2B is +6. This suggests that electrostatic repulsion between canonical H2A and H2B histone cores may result in a lower dimer thermostability compared to the variant in which there is slight overall electrostatic attraction.

To verify this hypothesis, we performed TSA assays on several H2A-H2B and H2A.B-H2B mutant dimers in which charged residues in H2A were substituted for the corresponding residues in H2A.B or vice versa. Two sites of charge reversal were selected to reduce the net charge of H2A maximally by a single mutation. Dimers refolded from H2A mutant K35E and K74E with net charge of +2 showed a significant increase in thermostability, which rescued half of the difference in melting temperatures between the canonical and variant dimers (Figure 3.6c). Dimers refolded from the reciprocal H2A.B mutants E40K and E79K with net charge +1 showed reduced thermostability compared to H2A.B-H2B dimer, thus qualitatively confirming our expectations. An H2A H30E mutant protein showed a melting temperature similar to that of wild-type protein, likely because at pH of the measurements (pH 6.5) the histidine is not fully protonated. Increase of the net positive charge of the canonical dimer by removal of glutamic acid E90 or E91 in the acidic patch H2A, either by substitution to arginine (E90R, net charge +6) or to leucine (E91L, net charge +5) from H2A.B, results in decreased melting temperatures. Overall these data reveal a clear trend of increasing unfolding

temperature with decreasing net charge of the dimer core. We thus conclude that the reduced net charge of H2A.B core and the resulting reduced electrostatic repulsion with H2B is responsible for the increased thermostability of the variant dimer.

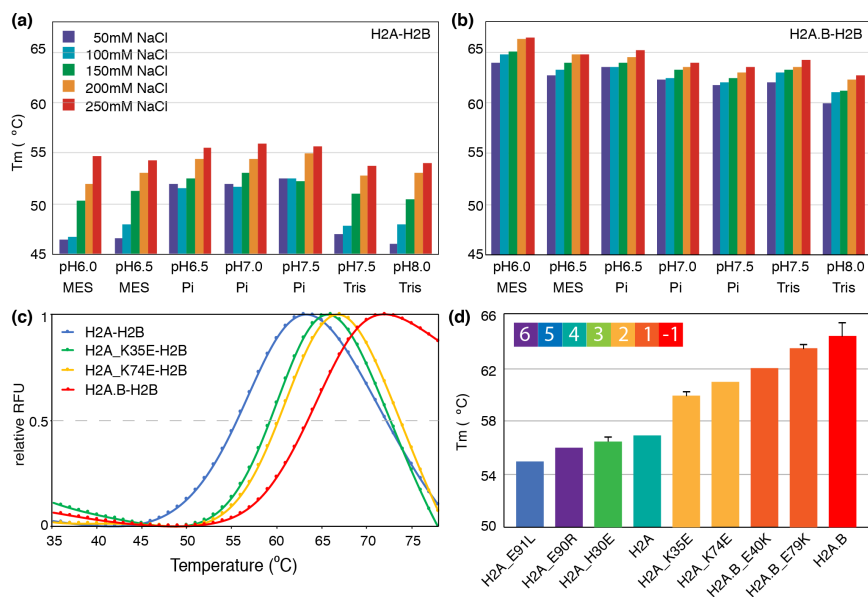


Figure 3.6. Thermostability of variant H2A.B-H2B and canonical H2A-H2B heterodimer. (a, b) Melting temperatures of H2A-H2B dimer (a) and H2A.B-H2B (b) across various pH and ionic strength conditions. Colour coding indicated. (c) Melting curves of H2A.B-H2B, H2A-H2B and selected mutant dimers as measured in thermal shift assay. Colour coding indicated. (d) Melting temperatures of H2A-H2B and H2A.B-H2B mutant dimers. The type of H2A protein incorporated in the dimer is shown on the x-axis. The net charge of the core region of H2A histones is indicated. Error bars based on duplicate experiments. Melting curves for (c, d) were measured in 20 mM NaPi buffer at pH 6.5 with 200 mM NaCl.

Discussion

Histone variant H2A.B is one of the most divergent histone variants that have evolved. Despite the specialized function and distinct properties of H2A.B-containing chromatin and nucleosomes, our understanding of the underlying molecular basis is yet limited. High resolution structural information on H2A.B-containing nucleosomes is

lacking, likely due to the destabilizing effect of H2A.B on the interactions to DNA and H3-H4. Here, we used solution-based method to study the structure, dynamics and stability of the H2A.B-H2B dimer. We find that H2A.B-H2B is structurally highly comparable to the free canonical dimer, yet with a markedly higher thermostability. Mutational analysis showed that the stability is correlated to the net charge of the dimer core, strongly suggesting that reduced electrostatic repulsion between the histone-fold domains of H2A.B and H2B is responsible for its increased thermostability. Since the change in charge is conserved across most of the 15 reported H2A.B sequences, including the key H30E, K35E, E91L and K74E substitutions, it is likely that the stability differences between canonical and variant dimers observed in this study are conserved across the H2A.B family. To obtain the solution structure of H2A.B-H2B heterodimer we made use of a CS-Rosetta⁴¹ driven approach, similar to the structure determination of the canonical H2A-H2B dimer²⁶. Since addition of sparse data, i.e. either from RDC and/or NOE data^{34,42,43} or evolutionary sequence information⁴⁴ can critically improve performance, we here included intermolecular NOEs as additional constraints in the calculations. This resulted in an overall better defined structural ensemble compared to using backbone chemical shift data only, in particular with respect to the position of the H2B α C helix. For the canonical H2A-H2B structure that was based on chemical shift data only, variable positions of the H2B α C helix and H2A α 1 helix were taken as sign of intrinsic flexibility²⁶. To what extent inclusion of NOE data would improve the definition of helices position for the canonical dimer remains to be established.

The solution structure of the isolated H2A.B-H2B heterodimer features a well-defined histone-fold core with disordered N- and C-terminal histones tails. The H2A.B α N-helix and the C-terminal extension of the docking domain are unfolded and likely only fold upon incorporation into a nucleosome. The chemical shift data do now show indications of transient folding for these regions, pointing to disorder-to-order transition upon binding to either DNA or H3-H4 in the nucleosome. Notably, these regions, in particular the α N helix, likely retain increased flexibility within the nucleosome, as was recently shown for canonical H2A in a solid-state NMR investigation of sedimented nucleosomes⁴⁵. Furthermore, the L2 loop arginine (R80) in

H2A.B is shifted by one residue compared to H2A (see the alignment in Figure 3.1a). Since the H2A R76 at L2 loop inserts DNA minor groove at SHL 5.5²³, a change in its position could directly influence DNA opening. From our structure it would be expected that this arginine clashes with the nucleosomal DNA, thus potentially forcing the DNA in a more open conformation. Whether the L2 loop and thus the R80 position, is maintained within the H2A.B-nucleosome, or will adapt to the nucleosomal DNA remains to be explored. Finally, we have highlighted here the increased negative charge in the H2A.B $\alpha 3$ helix that faces an acidic surface in the H3-H4 tetramer in the context of the histone octamer. The resulting electrostatic repulsion could further contribute to destabilization of H2A.B containing nucleosome and to the failure in refolding the H2A.B containing histone octamer by salt-dialysis²⁷.

We have found that the isolated H2A.B-H2B dimer has significantly higher thermostability than the canonical dimer (13 °C degree higher T_m). For most proteins a higher thermostability can be interpreted as a sign of overall increased thermodynamic stability⁴⁶. Indeed, stability measurements of the canonical dimer at constant temperature (25 °C) have been in reasonable agreement with thermal melting data^{40,47}. This thus further supports an increased thermodynamic stability of the H2A.B-H2B dimer. Strikingly, the H2A.B-containing nucleosome was recently shown to also have a slightly higher thermostability (5 °C higher T_m) compared to the canonical nucleosome⁴⁸. Here, it seems unlikely to directly translate to increase stability at ambient temperature given the more complex nature of this assembly and the extensive experimental evidence for reduced intermolecular interactions within the H2A.B-nucleosome.^{12,13,14,19,25} We thus hypothesize that the relative free energy of H2A.B-nucleosome is decreased on one hand by the reduced interaction to DNA and H3-H4, and on the other hand by the increased stability of the free H2A.B-H2B dimer. Our results suggest that these effects are at least in part interrelated since both originate from the lower positive charge of H2A.B. The net effect of these changes is to promote a shift in equilibrium towards dissociated, isolated H2A.B-H2B dimers and thus to promote opening up of the nucleosome and chromatin. Indeed, the free histone pool contains a relatively large amount of highly diffusive isolated H2A.B-H2B dimers^{21,19}. In addition, these free energy changes

likely also result in decreased energy barriers between states, facilitating increased exchange dynamics from nucleosomes to the free histone pool^{17,49}.

In summary, we presented the solution structure of the H2A variant H2A.B-H2B dimer, highlighting extensive changes in its surface electrostatics within an otherwise regular histone-fold, and showed that H2A.B-H2B has higher thermostability than the canonical dimer. This structure provides an important building block for a further detailed structural characterization of the assembly and exchange of the H2A.B nucleosome and provides new insights in its impact on chromatin structure and function.

Accession codes

The ensemble of solution structures is deposited in the Protein Data Bank under accession codes 6SMM. The NMR chemical shift assignment and NOE distance restraints have been deposited in the Biological Magnetic Resonance Data Bank under accession code 34426. Proteins used in this study are *Dm.* H2A (Uniprot-id: P84051) and *Dm.* H2B (Uniprot-id: P02283) and *Hs.* H2A.B (Uniprot: P0C5Z0).

Acknowledgement

This research was financially supported by The Dutch Research Council NWO (VIDI grant 723.013.010 to HvI). We thank Dr. Karolin Luger (University of Colorado) for providing the H2A.B plasmid, Alma Svatos for her help in the chemical shift assignment, Dr. Jorgen Schaarschmidt and prof. Alexandre Bonvin for help in setting up the Rosetta calculations, Bart van den Berg, Renske Snoeks, Kerven Dammers and Kim Fernandez for their help in making some of the mutants used in this study, Anneloes Blok and Mark Daniels for technical assistance, and the NMR group members for their stimulating support.

References

1. Draizen, E. J.; Shaytan, A. K.; Marino-Ramirez, L.; Talbert, P. B.; Landsman, D.; Panchenko, A. R., HistoneDB 2.0: a histone database with variants--an integrated resource to explore histones and their variants. *Database (Oxford)* **2016**, 2016.
2. Singh, R.; Bassett, E.; Chakravarti, A.; Parthun, M. R., Replication-dependent histone isoforms: a new source of complexity in chromatin structure and function. *Nucleic Acids Res* **2018**, 46 (17), 8665-8678.
3. Long, M.; Sun, X.; Shi, W.; Yanru, A.; Leung, S. T. C.; Ding, D.; Cheema, M. S.; MacPherson, N.; Nelson, C. J.; Ausio, J.; Yan, Y.; Ishibashi, T., A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res* **2019**.
4. Talbert, P. B.; Henikoff, S., Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* **2017**, 18 (2), 115-126.
5. Cheema, M. S.; Ausio, J., The Structural Determinants behind the Epigenetic Role of Histone Variants. *Genes (Basel)* **2015**, 6 (3), 685-713.
6. Zink, L. M.; Hake, S. B., Histone variants: nuclear function and disease. *Curr Opin Genet Dev* **2016**, 37, 82-89.
7. Palmer, D. K.; O'Day, K.; Wener, M. H.; Andrews, B. S.; Margolis, R. L., A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J Cell Biol* **1987**, 104 (4), 805-15.
8. Rogakou, E. P.; Pilch, D. R.; Orr, A. H.; Ivanova, V. S.; Bonner, W. M., DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **1998**, 273 (10), 5858-68.
9. Eirin-Lopez, J. M.; Ishibashi, T.; Ausio, J., H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *FASEB J* **2008**, 22 (1), 316-26.
10. Gonzalez-Romero, R.; Mendez, J.; Ausio, J.; Eirin-Lopez, J. M., Quickly evolving histones, nucleosome stability and chromatin folding: all about histone H2A.Bbd. *Gene* **2008**, 413 (1-2), 1-7.
11. Chadwick, B. P.; Willard, H. F., A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* **2001**, 152 (2), 375-84.
12. Bao, Y.; Konesky, K.; Park, Y. J.; Rosu, S.; Dyer, P. N.; Rangasamy, D.; Tremethick, D. J.; Laybourn, P. J.; Luger, K., Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J* **2004**, 23 (16), 3314-24.

13. Doyen, C. M.; Montel, F.; Gautier, T.; Menoni, H.; Claudet, C.; Delacour-Larose, M.; Angelov, D.; Hamiche, A.; Bednar, J.; Faivre-Moskalenko, C.; Bouvet, P.; Dimitrov, S., Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *EMBO J* **2006**, *25* (18), 4234-44.
14. Arimura, Y.; Kimura, H.; Oda, T.; Sato, K.; Osakabe, A.; Tachiwana, H.; Sato, Y.; Kinugasa, Y.; Ikura, T.; Sugiyama, M.; Sato, M.; Kurumizaka, H., Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci Rep* **2013**, *3*, 3510.
15. Montel, F.; Menoni, H.; Castelnovo, M.; Bednar, J.; Dimitrov, S.; Angelov, D.; Faivre-Moskalenko, C., The dynamics of individual nucleosomes controls the chromatin condensation pathway: direct atomic force microscopy visualization of variant chromatin. *Biophys J* **2009**, *97* (2), 544-53.
16. Zhou, J.; Fan, J. Y.; Rangasamy, D.; Tremethick, D. J., The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat Struct Mol Biol* **2007**, *14* (11), 1070-6.
17. Angelov, D.; Verdel, A.; An, W.; Bondarenko, V.; Hans, F.; Doyen, C. M.; Studitsky, V. M.; Hamiche, A.; Roeder, R. G.; Bouvet, P.; Dimitrov, S., SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays. *EMBO J* **2004**, *23* (19), 3815-24.
18. Soboleva, T. A.; Parker, B. J.; Nekrasov, M.; Hart-Smith, G.; Tay, Y. J.; Tng, W. Q.; Wilkins, M.; Ryan, D.; Tremethick, D. J., A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS Genet* **2017**, *13* (2), e1006633.
19. Tolstorukov, M. Y.; Goldman, J. A.; Gilbert, C.; Ogryzko, V.; Kingston, R. E.; Park, P. J., Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol Cell* **2012**, *47* (4), 596-607.
20. Sansoni, V.; Casas-Delucchi, C. S.; Rajan, M.; Schmidt, A.; Bonisch, C.; Thomae, A. W.; Staeger, M. S.; Hake, S. B.; Cardoso, M. C.; Imhof, A., The histone variant H2A.Bbd is enriched at sites of DNA synthesis. *Nucleic Acids Res* **2014**, *42* (10), 6405-20.
21. Gautier, T.; Abbott, D. W.; Molla, A.; Verdel, A.; Ausio, J.; Dimitrov, S., Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep* **2004**, *5* (7), 715-20.
22. McGinty, R. K.; Tan, S., Nucleosome structure and function. *Chem Rev* **2015**, *115* (6), 2255-73.

23. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, *389* (6648), 251-60.
24. Shukla, M. S.; Syed, S. H.; Goutte-Gattat, D.; Richard, J. L.; Montel, F.; Hamiche, A.; Travers, A.; Faivre-Moskalenko, C.; Bednar, J.; Hayes, J. J.; Angelov, D.; Dimitrov, S., The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling. *Nucleic Acids Res* **2011**, *39* (7), 2559-70.
25. Dai, L.; Xie, X.; Zhou, Z., Crystal structure of the histone heterodimer containing histone variant H2A.Bbd. *Biochem Biophys Res Commun* **2018**, *503* (3), 1786-1791.
26. Moriwaki, Y.; Yamane, T.; Ohtomo, H.; Ikeguchi, M.; Kurita, J.; Sato, M.; Nagadoi, A.; Shimojo, H.; Nishimura, Y., Solution structure of the isolated histone H2A-H2B heterodimer. *Sci Rep* **2016**, *6*, 24999.
27. Luger, K.; Rechsteiner, T. J.; Richmond, T. J., Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* **1999**, *304*, 3-19.
28. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995**, *6* (3), 277-93.
29. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325-7.
30. Shen, Y.; Bax, A., Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **2013**, *56* (3), 227-41.
31. Shen, Y.; Vernon, R.; Baker, D.; Bax, A., De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* **2009**, *43* (2), 63-78.
32. Shen, Y.; Bryan, P. N.; He, Y.; Orban, J.; Baker, D.; Bax, A., De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci* **2010**, *19* (2), 349-56.
33. Guntert, P.; Buchner, L., Combined automated NOE assignment and structure calculation with CYANA. *J Biomol NMR* **2015**, *62* (4), 453-71.
34. Lange, O. F.; Rossi, P.; Sgourakis, N. G.; Song, Y.; Lee, H. W.; Aramini, J. M.; Ertekin, A.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D., Determination of

solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc Natl Acad Sci U S A* **2012**, *109* (27), 10873-8.

35. Shen, Y.; Bax, A., SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* **2010**, *48* (1), 13-22.

36. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* **2016**, *86*, 2 9 1-2 9 37.

37. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **2004**, *32* (Web Server issue), W665-7.

38. Tachiwana, H.; Kagawa, W.; Osakabe, A.; Kawaguchi, K.; Shiga, T.; Hayashi-Takanaka, Y.; Kimura, H.; Kurumizaka, H., Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc Natl Acad Sci U S A* **2010**, *107* (23), 10454-9.

39. Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I., Protein stability and surface electrostatics: a charged relationship. *Biochemistry* **2006**, *45* (9), 2761-6.

40. Karantza, V.; Baxevanis, A. D.; Freire, E.; Moudrianakis, E. N., Thermodynamic studies of the core histones: ionic strength and pH dependence of H2A-H2B dimer stability. *Biochemistry* **1995**, *34* (17), 5988-96.

41. Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A., Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A* **2008**, *105* (12), 4685-90.

42. Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D., NMR structure determination for larger proteins using backbone-only data. *Science* **2010**, *327* (5968), 1014-8.

43. ElGamacy, M.; Riss, M.; Zhu, H.; Truffault, V.; Coles, M., Mapping Local Conformational Landscapes of Proteins in Solution. *Structure* **2019**, *27* (5), 853-865 e5.

44. Tang, Y.; Huang, Y. J.; Hopf, T. A.; Sander, C.; Marks, D. S.; Montelione, G. T., Protein structure determination by combining sparse NMR data with evolutionary couplings. *Nat Methods* **2015**, *12* (8), 751-4.

45. Xiang, S.; le Paige, U. B.; Horn, V.; Houben, K.; Baldus, M.; van Ingen, H., Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **2018**, *57* (17), 4571-4575.
46. Rees, D. C.; Robertson, A. D., Some thermodynamic implications for the thermostability of proteins. *Protein Sci* **2001**, *10* (6), 1187-94.
47. Gloss, L. M.; Placek, B. J., The effect of salts on the stability of the H2A-H2B histone dimer. *Biochemistry* **2002**, *41* (50), 14951-9.
48. Taguchi, H.; Horikoshi, N.; Arimura, Y.; Kurumizaka, H., A method for evaluating nucleosome stability with a protein-binding fluorescent dye. *Methods* **2014**, *70* (2-3), 119-26.
49. Okuwaki, M.; Kato, K.; Shimahara, H.; Tate, S.; Nagata, K., Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. *Mol Cell Biol* **2005**, *25* (23), 10639-51.

Supplements

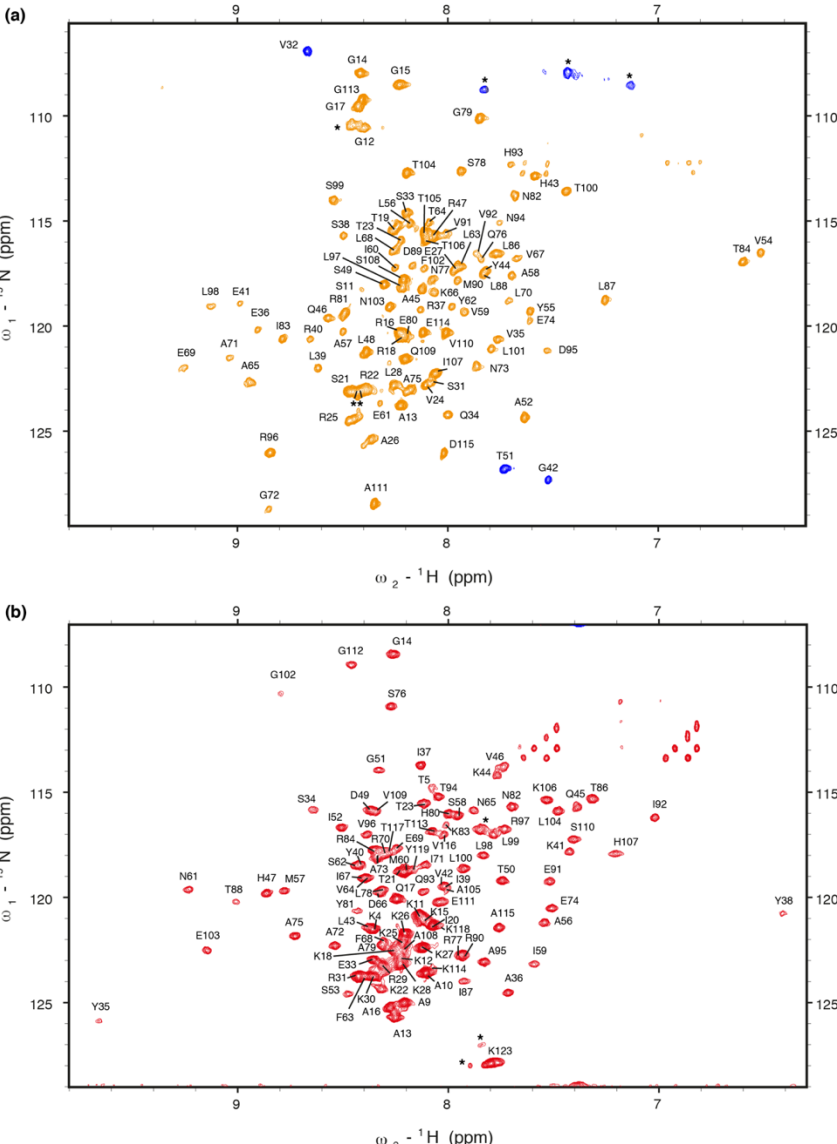


Figure S3.1. 1H-15N TROSY spectra of (a) H2A.B and (b) H2B within the H2A.B-H2B heterodimer. Assignments are indicated, unassigned backbone resonances are indicated with a *. Negative folded peaks are shown in blue.

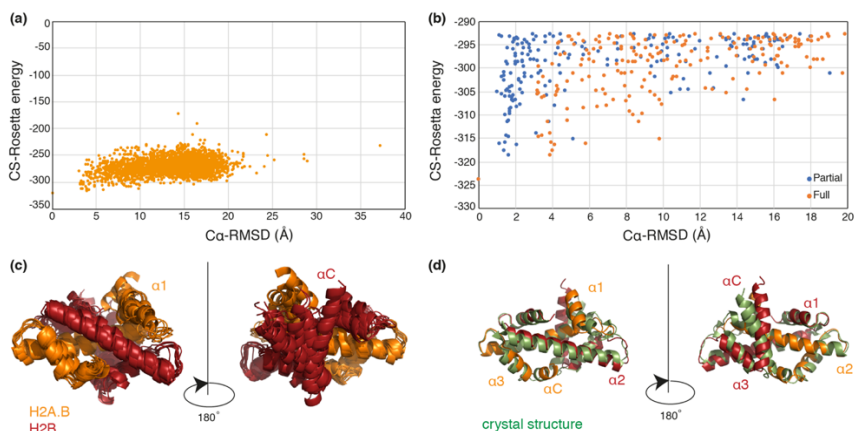


Figure S3.2. CS-Rosetta solution structure of H2A.B-H2B. (a,b) CS-Rosetta energy plotted versus C α -RMSD to the lowest energy structure. In (a) C α -RMSD was calculated over the complete histone core shown for all 3000 calculated solutions. In (b) a zoom of the plot in (a) is shown with C α -RMSD calculated over the complete histone core in orange, and RMSD excluding the H2B α C and H2A.B α 1 helices in blue. (c) Superposition of the 10 CS-Rosetta models with the lowest C α -RMSD from the lowest energy 20 models, showing ill-defined positions of H2B α C and H2A.B α 1-helices within an otherwise converged structure. (d) Structural comparison between lowest-energy CS-Rosetta model and the H2A.B-H2B crystal structure (PDB-id 6a7u).

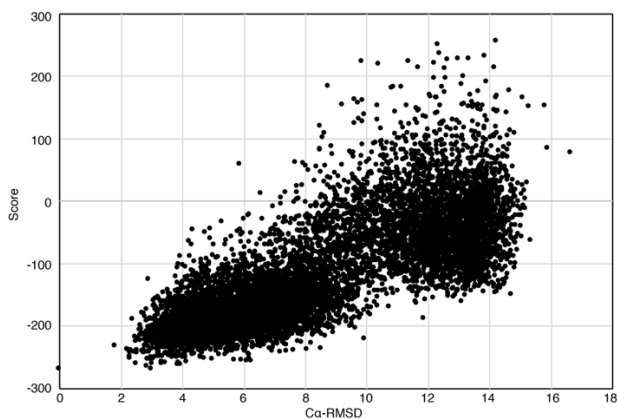


Figure S3.3. Final re-scored Rosetta energy including both chemical shift and NOE correspondence plotted versus C α -RMSD to the lowest-energy structure of the H2A.B-H2B core. In total 8000 solutions were calculated. The RMSD is calculated over the complete histone-core.

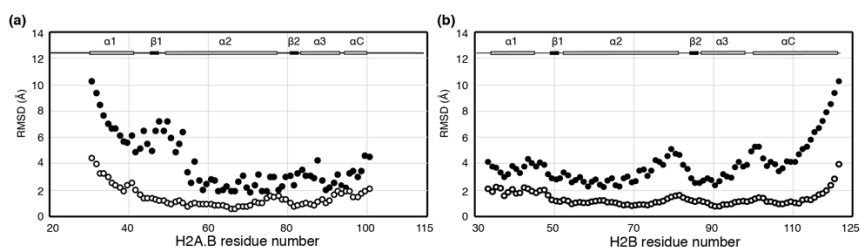


Figure S3.4. Impact of inclusion of intermolecular NOE data on CS-Rosetta calculated structure. (a, b) Average per-residue heavy atom backbone RMSD, averaged over the 10 models in either the final ensemble calculated without/with (filled/blank circles) NOE restraints. Average per-residues RMSDs are plotted for H2A.B (a) and H2B (b) core regions. Secondary structure elements in the final solution structure are plotted on top.

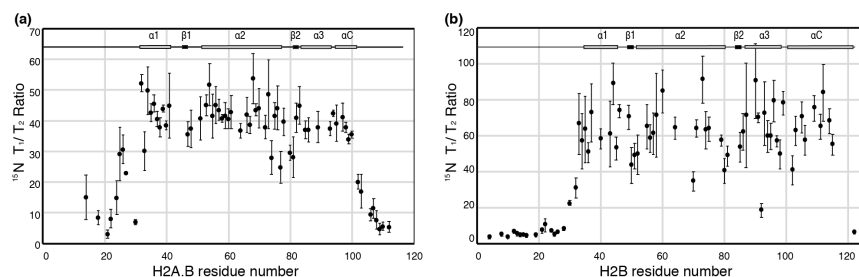


Figure S3.5. Backbone dynamics of H2A.B-H2B dimer. (a, b) Ratio of backbone amide ^{15}N - T_1 and ^{15}N - T_2 values for H2A.B (a) and H2B (b), illustrating the high flexibility of the H2A.B N- and C-terminal tails, as well as the H2B N-terminal tail. The average T_1/T_2 -ratio for residues the folded core of H2B are elevated compared to those in the folded parts of H2A.B, most likely due the higher salt concentration in the sample used for measuring H2B dynamics. Secondary structure elements in the final solution structure are plotted on top.

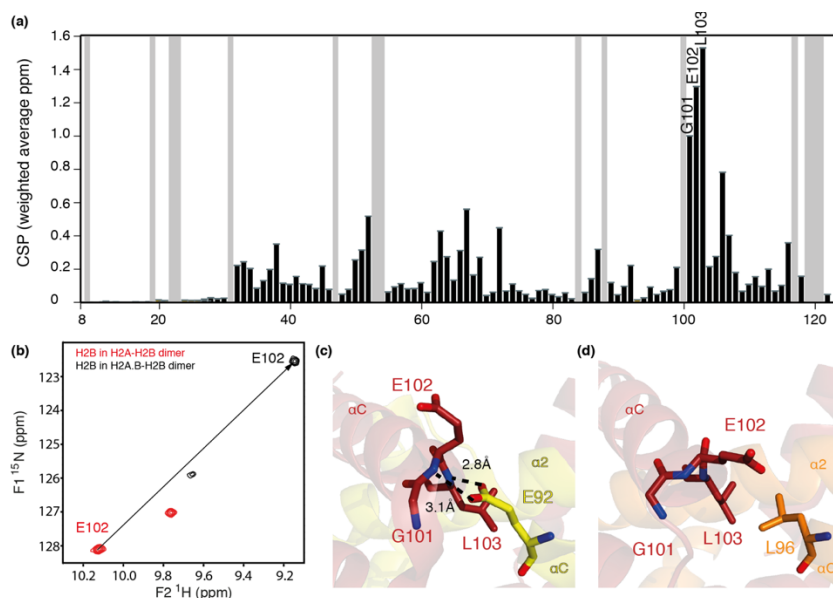


Figure S3.6. Chemical shift perturbations of H2B residues in canonical and variant dimers caused by sequence variance between H2A and H2A.B. (a) Chemical shift perturbations (CSPs) for H2B residues between two canonical and variant dimers. Unassigned residues shown in grey. Residues with CSP larger than 1ppm are labeled. (b) Overlay of H2B TROSY spectra within canonical H2A-H2B dimer (red) and variant H2A.B-H2B dimer (black), zooming in on the resonance of E102 in H2B. (c,d) Comparison of solution structure of H2A.B-H2B (d) and crystal structure of H2A-H2B (c) zooming in on the loss of a hydrogen-bond due to substitution of H2A E92 to H2A.B L96.

Table S3.1. Structural statistics for the core region of isolated H2A.B-H2B heterodimer

A. Restraint information	
number of intermolecular NOEs	44
B. Average RMS deviation from experimental restraints	
All experimental distance restraints (Å)	0.08 ± 0.24
C. Coordinate RMS deviation^a (Å)	
Average RMSD to mean	
Ordered backbone atoms	1.08 ± 0.24
Ordered heavy atoms	1.54 ± 0.25
Global backbone atoms	1.48 ± 0.25
Global all heavy atoms	1.92 ± 0.33
Average Pairwise RMSD	
Ordered backbone atoms	1.61 ± 0.35
Ordered all heavy atoms	2.30 ± 0.36
Global backbone atoms	2.21 ± 0.38
Global all heavy atoms	2.91 ± 0.41

^a Ordered regions correspond to residues V31-F101 of H2A.B and Y34-K122 of H2B.

Chapter 4. Structure and dynamics of the H2A.B variant nucleosome

This chapter is based on:

Heyi Zhang, Yiran Lin, Jan Huertas, Vlad Cojocaru, Vincenzo Lobbia, Ulric B. le Paige, Hugo van Ingen. Structure and dynamics of the H2A.B variant nucleosome. *In preparation*.

Contributions of authors:

MNase digestion assays were performed with Yiran Lin; NMR study of H2A.B N-terminal tail in the nucleosome was performed with Vincenzo Lobbia; Energy minimization of H2A.B nucleosome models was performed by Jan Huertas and Vlad Cojocaru; 601 DNA used in nucleosome reconstitutions was provided by Ulric B. le Paige.

Abstract

Incorporation of histone variants into nucleosomes confers distinct functional properties to nucleosomes by their unique amino acid sequences, contributing to the regulation of chromatin functions such as transcription and DNA repair. Histone variant H2A.B is a highly divergent H2A variant and is strongly linked to active transcription and RNA splicing. H2A.B nucleosomes are known to have an ‘open’ conformation in which the nucleosomal DNA is partially released from the histone octamer core at the entry/ exit ends. Yet, the exact molecular reason for such opened structure is unknown, nor is there high-resolution structural and dynamical data of H2A.B nucleosomes available. Here, we report an NMR-driven investigation of the structure and dynamics of H2A.B nucleosomes, showing that the H2A.B truncated docking domain is folded within the variant nucleosome as in the canonical conformation and forms a stably folded interface with the H3-H4 subunit. Our NMR data further indicate that the H3 N-terminal tail in H2A.B nucleosomes has increased flexibility and is overall less bound to DNA compared to canonical nucleosomes. Finally, we identify a register shift of a DNA minor groove anchoring arginine in H2A.B and show using micrococcal nuclease digestion assays that this shift contributes to destabilization and DNA opening of the nucleosome. Together, these results provide new insights into the molecular properties of the H2A.B nucleosome.

Introduction

Histone proteins are among the most highly conserved proteins in nature, reflecting their cardinal importance in packaging and protecting the genetic material of eukaryotes in the form of chromatin. The bulk of chromatin is made of four canonical histones: H2A, H2B, H3 and H4, and together they form a histone octamer around which ~150 base pairs of DNA is wrapped, creating the nucleosome. To regulate chromatin biology, nucleosomes are the binding target of many nuclear proteins and are actively altered in various ways, one of which is by a special class of histone proteins, histone variants, to confer specific structural and functional properties to chromatin at particular locations in the genome.

Histone variants have been discovered for all four core histone types ¹⁻² and can substitute their canonical forms in the nucleosome, typically through the action of specific histone chaperones and remodeller proteins ³. Through their distinct amino acid sequences, incorporation of histone variants bestows special functional and conformational properties to nucleosome. For example, histone H3 variant CENP-A is crucial in assembling the centrosome as it facilitates binding of centromeric DNA-binding proteins ⁴ and histone H2A variant H2A.X plays a major role in signalling DNA damage through phosphorylation of a unique serine in its C-terminus ⁵.

One of the most recently discovered and the most divergent histone variants is H2A variant H2A.B or H2A.Bbd (Barr body deficient) (Figure 4.1a). Originally identified as a variant that is excluded from the inactive X chromosome ⁶, it is now tightly associated to active transcription and mRNA splicing ⁷⁻¹². These functional roles may ultimately result from the pronounced promotion of open chromatin structure by H2A.B. When incorporated into nucleosomal arrays, H2A.B was shown to prevent formation of a compact chromatin fibre ¹³. At the level of isolated nucleosomes, a series of biochemical and low-resolution structural studies have shown that, compared to canonical nucleosomes, nucleosomes containing H2A.B have a significantly shorter stretch of DNA tightly bound to the histone octamer (with the last turn of DNA dissociated from the octamer) and that the DNA ends are at a much wider angle relative to each other (Figure 4.1b,c) ¹⁴⁻¹⁶.

The sequence differences between canonical H2A and the variant H2A.B must underlie the molecular basis of its open nucleosomal conformation, yet up to date an exact description of this relationship is lacking. Since the last turn of nucleosomal DNA is bound to H3 in the canonical nucleosome, there is not an *a priori* clear-cut answer to this question. At only 48% sequence identity to canonical H2A, H2A.B's distinct features are the elongated and arginine rich N-terminal tail, a truncated C-terminal docking domain, the absence of acidic patch residues, and an overall less basic nature. Previous studies have pointed to the truncated C-terminal docking domain as main cause of the open nucleosome conformation^{14-15, 17}. This region supports multiple interactions with H3 and H4 that stabilizes the histone octamer core and is truncated by ~ 19 residues in H2A.B resulting in a loss of several of these stabilizing interactions¹⁸. A series of experiments including MNase digestion and DNase footprinting assays, as well as FRET measurement of reconstituted nucleosomes using either H2A docking domain mutants or H2A/H2A.B chimera proteins resulted in the conclusion that H2A.B specific sequence differences from the $\alpha 3$ helix to the C-terminus are causing the open structure of the nucleosome either directly or indirectly^{14-15, 17}. In addition, recent structures of the H2A.B-H2B heterodimers have pointed to the reduced positive electrostatic potential on the DNA binding surface which may promote DNA unwrapping¹⁹ (see Chapter 3).

Here, we aimed to characterize the structural and dynamical properties of H2A.B nucleosomes at atomic resolution, combining solution-state NMR, biochemical assays and molecular dynamics (MD) simulations. We show that the elongated H2A.B N-terminal tail is highly dynamic whereas on the C-terminal side only the ultimate residue is highly mobile. Our NMR data indicate that the H2A.B truncated docking domain is folded and interfaced to H3-H4 as in canonical nucleosomes. NMR relaxation data support an opened nucleosome structure in which the DNA is at least transiently released from the octamer surface. As a result, the H3 N-terminal tail is more dynamic and overall less DNA-bound than in canonical nucleosomes, which is supported by a preliminary analysis of ongoing MD simulations. Finally, we identified a register shift of a key DNA binding arginine residue in H2A.B that contributes directly to unwrapping of the nucleosomal DNA. Together,

this study offers new insights into the unusual properties of the H2A.B nucleosome.

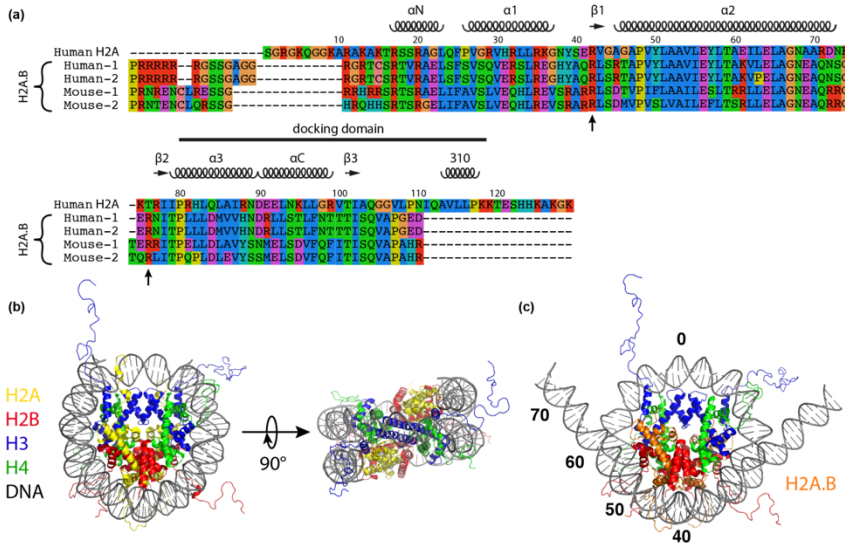


Figure 4.1. Histone variant H2A.B. (a) Sequence alignment of the most studied human and mouse H2A.B sequences with the human canonical H2A. Secondary structural elements of canonical H2A in the context of nucleosome are indicated above the sequences. The definition of docking domain is marked as a black bar. Positions of anchoring arginine's in canonical nucleosome are shown with arrows below the sequences. (b) Overall structure of the canonical nucleosome (PDB id: 1KX5). Colour coding is indicated. (c) A simplistic model for the H2A.B nucleosome with an arbitrary conformation for the terminal 15 bp of nucleosomal DNA. The dyad DNA base pair is marked as position 0 and the terminal three turns of the DNA are marked at positions 40-70.

Results

H2A.B N-terminal tail is highly flexible and transiently DNA-bound

To investigate the structural and dynamical properties of the elongated and arginine rich H2A.B N-terminal tail, we reconstituted H2A.B nucleosome using uniformly ^{15}N labelled H2A.B and unlabelled H2B, H3, H4, and DNA. This isotope-labelling pattern in combination with the large mass of nucleosome (> 200 kDa) permits the selective observation of the highly flexible parts of H2A.B. Thirteen peaks were observed in the 2D NH TROSY fingerprint spectrum (Figure 4.2a).

Comparison to the H2A.B assignments in the H2A.B-H2B heterodimer (see Chapter 3) allowed transfer of seven assignments for G11-G14, R17, T18 and D114, together with the blob assignment of the six consecutive arginine's (R2-R7). The two remaining peaks did not overlap with any peak in the dimer spectrum and could only be assigned tentatively to S9 and/or S10, and C19. Apart from the C-terminal residue D114, the observed resonances together constitute the majority of the twenty N-terminal residues, corresponding to the expected size of the N-terminal H2A.B tail and demonstrating the high flexibility of this region. No resonances for residues 21-25 could be observed suggesting the α N-helix is formed as in the canonical nucleosome and the H2A.B-H2B crystal structure ¹⁸⁻¹⁹, while it is highly flexible and disordered within the variant dimer free in solution (see Chapter 3).

Notably, resonances for R15 and G16 are absent and resonances for G14 and R17 have very low signal intensity, indicating the limited mobility for the GRGR box in the nucleosome which is possibly caused by its attachment to the nucleosomal DNA. Further evidence for transient DNA binding comes from the small chemical shift changes of the tail resonances in the nucleosome when compared to the dimer. Interestingly, these changes are in the downfield direction whereas upfield changes were observed for canonical H2A in a mononucleosome sediment ²⁰, signifying a change in DNA interaction mode. In addition, in both sediment and solution, peak doubling was observed for several resonances in the canonical H2A tail, reflecting the interaction to the different sites of the non-palindromic 601-DNA ²⁰⁻²¹. For H2A.B such peak doubling is not apparent suggesting that at least for the observed resonances DNA binding is less intimate.

Assignment of H2A.B docking domain methyl TROSY signals

We next examined the conformation and dynamics of the H2A.B C-terminal docking domain within the nucleosome. Starting from the end of H2A.B α C helix, this region comprises residues F101 to D114. This region is unfolded and highly dynamic within the context of the dimer (see Chapter 3), yet the corresponding part in canonical H2A forms a defined interface to H3-H4 tetramer in nucleosome, including an intermolecular β -sheet (see Figure 4.1). We thus wanted to probe the conformation of this region in the variant nucleosome. The tail-focused

NMR experiment described above showed only one observable resonance for this region, the C-terminus residue D114 (Figure 4.2a). This suggests that residues F101-E113 have strongly reduced mobility within the context of the nucleosome and are thus most likely folded in a defined conformation.

To further examine the conformation of the docking domain, we used the methyl-TROSY approach²² to allow site-specific investigation of the histone octamer core²³. H2A.B nucleosomes were prepared with Ile- δ 1- $^{13}\text{CH}_3$, Leu, Val- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$]-H2A.B and perdeuterated H2B, H3, and H4 with unlabelled 601-DNA. The scarcity of protons in such sample dramatically slows the transverse relaxation of the methyl ^1H spins, thereby allowing to probe the Ile, Val, Leu methyl groups in such large system as the nucleosome. The methyl-TROSY spectrum of ILV-labelled H2A.B within the nucleosome, displayed in Figure 4.2b, shows overall decent chemical shift dispersion. Out of the 47 expected resonances for the 3 Ile, 13 Leu and 9 Val residues in H2A.B, 45 peaks could be identified, including three signals in the Ile-region. The signal intensities appear rather heterogenous, in particular for the Leu/Val region where several resonances have very low intensities.

Assignment of methyl-TROSY spectra is typically based on a combination of mutagenesis and NOESY spectra, by comparing the experimental connectivities in the proton network to that expected from the structure. In the case of H2A.B nucleosomes, NOESY experiments were of low quality, showing only few NOEs, thereby preventing complete assignment of the methyl signals. Instead, we focussed our analysis on the three isoleucine's in H2A.B, which give rise to well-dispersed and intense cross peaks in the nucleosome spectra. Since two of the isoleucine's are located at approximately opposite ends of the histone fold core (I59 and I82) and one (I106) is part of the docking domain, assignment of these three signals offers the opportunity to probe the structural and dynamical properties of H2A.B in the nucleosome at key locations.

In a straightforward approach, we compared the H2A.B nucleosome spectrum to that of the canonical nucleosome with known H2A assignments²³ (see Figure S4.3). Two isoleucines in H2A.B are conserved in canonical H2A; H2A.B I82 corresponds to H2A I77 and H2A.B I106 to H2A I101. The inset in Figure 4.2b shows that each of

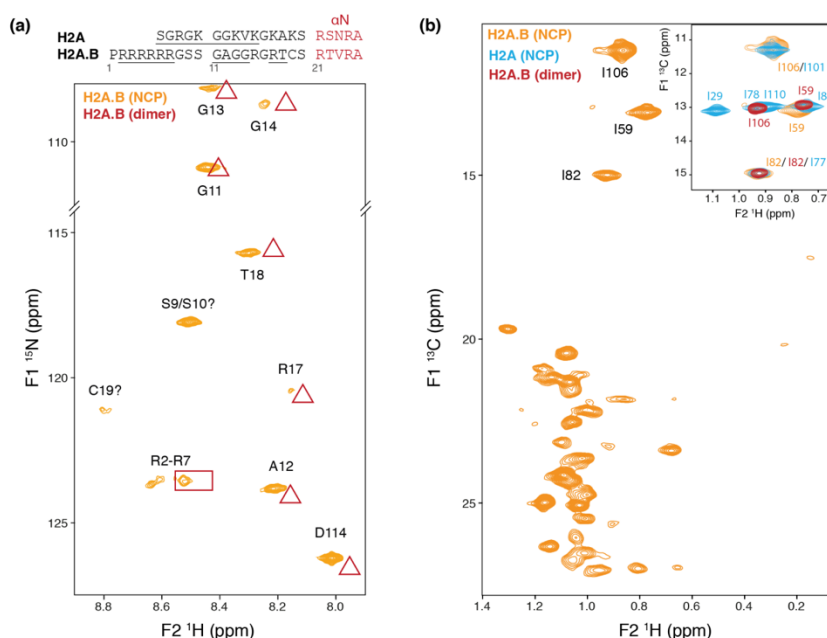


Figure 4.2. Amide and methyl-TROSY NMR spectra of H2A.B in the nucleosome. (a) Amide HN-TROSY spectrum of uniformly ^{15}N -labeled H2A.B in otherwise protonated nucleosomes. Assignments for tail residues in the H2A.B-H2B heterodimer are indicated with triangles for well resolved individual assignments and as a rectangle for overlapped group assignments. Tentative assignments indicated with ? were based on chemical shift correspondence to the dimer (S10) or based on chemical shift statistics of tail residues (S9 and C19, unassigned in the dimer). Amino acid sequences of H2A.B and H2A N-terminal tails are indicated on top, aligned with respect to the αN helix. Residues visible in canonical H2A nucleosomes are underlined. (b) Methyl-TROSY spectrum of ILV-labelled H2A.B with otherwise perdeuterated core histones and protonated DNA in the nucleosome. The inset shows the overlay of methyl-TROSY spectra Ile region of H2A.B in the nucleosome (orange) and dimer (red) and canonical H2A in nucleosome (cyan). Spectra were processed with the same parameters and overlaid by matching I82 of H2A.B to the corresponding I77 of H2A.

the H2A I77 and I101 overlaps with a H2A.B resonance, thus allowing to transfer their assignments to the corresponding residues in H2A.B. By exclusion, the remaining H2A.B isoleucine can be assigned to I59. This assignment is further supported by comparison to methyl-TROSY spectra of ILV-labelled H2A.B in H2A.B-H2B heterodimers (see Figure S4.3). There are two (nearly) overlapping signals, which match the histone fold core resonances for which little changes in chemical

environment are expected upon formation of the nucleosome. One resonance has very different chemical shifts in the two systems, corresponding to the docking domain isoleucine I106 (see inset Figure 4.2b).

Conserved docking domain structure of H2A.B in the variant nucleosome

The chemical shift difference of I106 between H2A.B in dimers and in nucleosomes suggests that the docking domain of H2A.B becomes folded upon incorporation of the variant dimer into the nucleosome. The close chemical shift correspondence to H2A I101 in canonical nucleosomes suggests the variant docking domain obtains the same fold as the canonical domain. To further investigate the conformation of the docking domain, we analysed the NOESY spectrum of H2A.B in the nucleosome. Among the few NOE cross peaks, a clear NOE was observed between I106 and a leucine or valine methyl group (Figure 4.3a). While the leucine and valine methyl signals are as of yet unassigned, the ^{13}C chemical shift (14.7 ppm) indicates the resonance belongs most likely to a valine. In addition, the distinct up-field ^1H chemical shift (0.138 ppm) indicates proximity of the corresponding methyl group to an aromatic ring. We next checked whether this NOE is consistent with folding of the H2A.B docking domain as in the canonical nucleosome structure. Using canonical H2A docking domain as a template for homology modelling, we find four methyl groups, from residues L87, V91, V109, that are within 6 Å ^{13}C - ^{13}C distance from the I106 δ 1-methyl group. Among these, the closest methyl group is from V91 at 4.1 Å, which is the only methyl within 5 Å, and the only one that is also close to an aromatic ring (see Figure 4.3b). We thus assigned the unknown resonances to V91, which is located on α 3 helix of H2A.B. This I106-V91 NOE cross peak indicates that the docking domain of H2A.B is folded in a stable conformation and suggests that the H2A.B interface to H3-H4 is formed in the canonical manner, including the β -sheet with H4.

H2A.B docking domain dynamics

Having established evidence for stable folding of the docking domain within the context of the H2A.B variant nucleosome, we next turned to characterize the H2A.B dynamics within the nucleosome. Given the

partially unwrapped nature of the H2A.B nucleosome, one could expect that the DNA unwrapping also results in a destabilization of the histone octamer and thus an increase in histone dynamics. Indeed, the prevalence of low intensity peaks in the methyl-TROSY spectrum could result from line broadening due to conformational exchange on a micro- to milli-second time scale. Alternatively, the line broadening could reflect the density of the local proton network which will have a larger impact on the line width upon increasing overall rotational tumbling time.

Comparison to the peak intensity distributions of H2A.B methyl resonances in either nucleosomes or H2A.B-H2B dimers and of H2A resonances in canonical nucleosomes indicate that the preponderance of broadened signals is specific to H2A.B in the nucleosomal context (Figure 4.4a). To quantify this more rigorously, we measured both ^1H and ^{13}C - ^1H multiple quantum transverse relaxation rates (R_2) of H2A.B isoleucine methyl groups in the variant nucleosome and compared these to relaxation rates of H2B core isoleucine methyls in the canonical nucleosome (unpublished data from Hugo van Ingen, Hidenori Kato, Yawen Bai, Lewis Kay). Histone H2B contains 9 isoleucines, of which 8 are part of the H2A-H2B histone fold core. As such, we expect the H2B isoleucine methyls to serve as equally valid reporters of the dimer dynamics as H2A. As shown in Figure 4.4b, while methyls in the variant nucleosome have similar multiple quantum R_2 values as in the canonical nucleosome, they have much higher ^1H R_2 values than in the canonical nucleosome, on average 82 s^{-1} vs. 47 s^{-1} . The elevated ^1H R_2 values in the variant nucleosome explain the relatively poor sensitivity of the NOESY experiment. Strikingly, the average ^1H R_2 of H2A.B methyl groups in buffer without any added salt matches well with rates obtained for the canonical nucleosome upon addition of 75 mM NaCl , conditions that are known to promote inter-nucleosome interactions in particular at NMR concentrations ²⁴⁻²⁵. The higher R_2 for the variant could thus reflect a larger effective particle size, for example due to DNA unwrapping or inter-nucleosome interactions. However, due to the lack of acidic patch of H2A.B, it is unlikely that inter-nucleosome interactions are happening in the sample. Nevertheless, since a similar trend for the multiple quantum R_2 values is not observed, line

broadening due to conformational dynamics affecting predominantly the ^1H chemical shift cannot be ruled out.

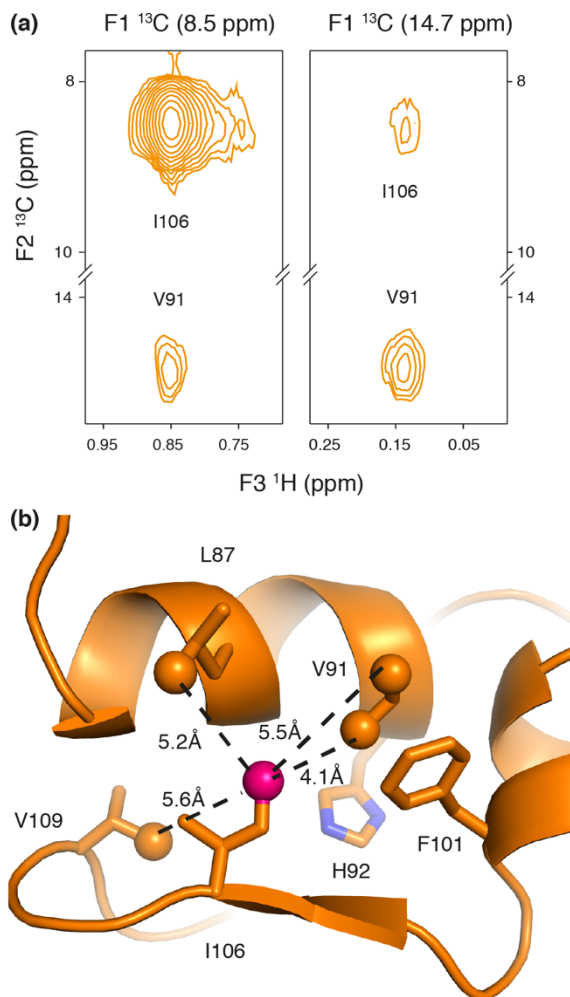


Figure 4.3. H2A.B docking domain structure. (a) Strips from the 3D CCH Me-NOESY showing NOE cross peaks between I106 to V91. (b) H2A.B structural model with docking domain modelled according to H2A structure in the canonical nucleosome. Methyl carbon of I106 is shown as pink sphere and H2A.B methyl carbons within 6 Å are shown as orange spheres. Residues and inter-methyl distances are marked. Side chains from aromatic residues (H92 and F101) are shown as sticks.

To investigate whether this specific broadening is due to slow local conformational changes, we recorded both ^1H single quantum and ^{13}C - ^1H multiple quantum (MQ) CPMG relaxation dispersion experiments. No dispersion in ^1H R_2 rates could be observed up to a CPMG pulsing rate of 1000 Hz (Figure S4.4). While small dispersions due to conformational changes ($\Delta R_2 < 10 \text{ s}^{-1}$) cannot be excluded due to the limited sensitivity, these results speak against large conformational motions in the millisecond time scale regime. In the ^{13}C - ^1H MQ experiment, similar flat dispersion profiles were observed for I59 and I82, indicating the lack of conformational changes for the two residues, whereas a slight dispersion of $R_{2,\text{eff}}$ value was observed for I106 (Figure 4.4c). Comparison against data recorded on H2A.B-H2B dimers shows that a similar slight dispersion is already present for I106 in the heterodimer (see inset Figure 4.4c). Importantly, comparison of the dimer and nucleosome data clearly show that the average $R_{2,\text{eff}}$ for I106 is similar to that for core isoleucines in the context of the nucleosome, while it was much lower within the dimer. This further supports stable folding of the H2A.B docking domain upon incorporation into the nucleosome.

Together, the NMR data on H2A.B nucleosomes point to a canonical-like histone octamer conformation. The H2A.B N-terminal tail is highly flexible as the H2A N-terminal tail, the truncated docking domain of H2A.B is folded as in H2A and there are no large-scale conformation motions present within the variant nucleosome. Yet, the elevated relaxation rates point to increased effective size of the variant nucleosome, likely as a result of partial DNA unwrapping.

Increased H3 tail dynamics in the H2A.B nucleosome

To further map the impact of H2A.B incorporation on histone octamer, we examined the H3 N-terminal tail since unwrapping of the nucleosomal DNA is likely to affect this tail foremost (see Figure 4.1). Using the same approach as for the H2A.B tail, we reconstituted the variant and canonical nucleosome using uniformly ^{15}N and ^{13}C labelled H3. Both amide NH-TROSY spectra showed well resolved peaks that were assigned to residues T3 to K36 based on previously published assignments²⁰ and standard triple resonance experiments (Figure 4.5a). Crucially, no additional peaks were observed for H3 in the H2A.B nucleosome, indicating that unwrapping of the DNA does not result in

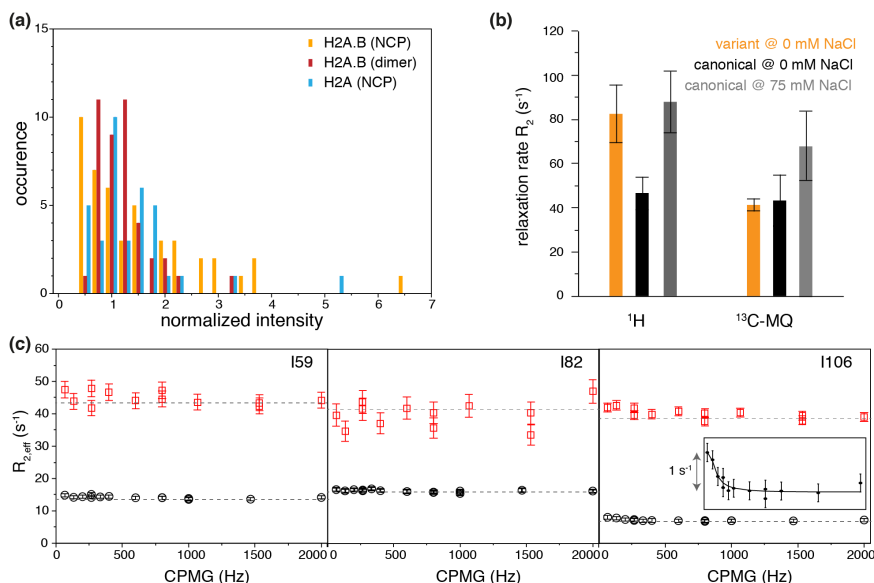


Figure 4.4. NMR relaxation data of H2A.B in NCP support stable folding of docking domain in an overall opened nucleosome conformation. (a) Histogram of normalized methyl-TROSY peak intensities for H2A.B in dimers and nucleosomes and H2A in nucleosomes. Peak intensities are normalized to the median intensity of each spectrum. **(b)** Average relaxation rates R_2 from ¹H (¹H) or ¹³C-¹H multiple quantum (¹³C-MQ) relaxation experiments for core isoleucine of H2A.B in variant nucleosome at no salt condition (orange bar), H2B in canonical nucleosome at no salt (black bar) or with salt (grey bar) conditions. **(c)** ¹³C-¹H multiple quantum CPMG relaxation dispersion curves for H2A.B Ile residues in nucleosomes (red squares) and dimers (black circles). The inset highlights the small dispersion of $R_{2,eff}$ values obtained for I106 in the variant dimer.

extension of the tail. In particular, resonances for K38-P44, which pass in between the two DNA gyres in the canonical nucleosomes, are not observed for either of the nucleosomes. This indicates that this region remains immobile, likely attached to only a single DNA helix in the variant nucleosome.

More detailed inspection of the two H3 spectra reveals that the majority of resonances are shifted upfield in the canonical compared to the variant nucleosome and this effect is present throughout the tail (Figure 4.5a,b). An elegant study from the Selenko lab has shown that the H3 tail resonances generally shift upfield upon incorporation into the

nucleosome as a result of transient DNA binding ²⁶. The general trend of the peak displacements here indicates that on average the H3 tail is less bound to the nucleosomal DNA in the variant nucleosome. More detailed comparison of the H2A.B-nucleosome spectra to those of the free H3 peptide recorded by the Selenko lab ²⁶ shows that several residues, including A29, A31 and V35 have very similar chemical shifts as in the free peptide, indicating there are likely unattached to DNA. This is in agreement with the simplistic model shown in Figure 4.1c, where the stretch of A29-V35 is far from the DNA ends and thus may not bind to it and represent more like the free peptide chemical environment. Other resonances, such as Q19-A21 and A25 have chemical shifts somewhere in between that of canonical nucleosomes and the free H3 peptide, suggesting a weakened DNA interaction. Yet a third group of residues, including K14 and A15, have chemical shifts distinct from either canonical nucleosome and free peptide. Together, the chemical shifts pattern observed for the H3 tail in the H2A.B nucleosome implies a distinctly altered DNA interaction that is overall less intimate. Notably, the glycine resonances, which were shown to be least involved in DNA binding, also have the smallest chemical shift differences between the two systems.

The reduced DNA binding of the H3 tail is further supported by a general decrease in local effective rotational correlation times (τ_c), derived from measured ¹⁵N R_1 and R_2 relaxation rates ²⁶, for the H3 tail in the variant nucleosome (Figure 4.5c). Overall, the local τ_c is reduced by ~1.5 ns for most of the H3 tail residues in the H2A.B nucleosome. Strikingly, the τ_c profile is highly similar between the two nucleosome types. This argues against full release of the H3 tail from the DNA, in which case a gradual decrease in τ_c would be expected from the nucleosome core towards the N-terminus. Rather these data speak to a maintained, but reduced DNA binding in the variant.

Register shift of a minor groove arginine anchor in the H2A.B nucleosome

With a canonical-like H2A.B conformation and no evidence for large conformational changes, the underlying molecular basis for the opened nucleosome structure remains unclear. In an attempt to identify a direct (contributing) factor, we carefully inspected the molecular model for the H2A.B variant nucleosome as shown in Figure 4.1. This revealed a

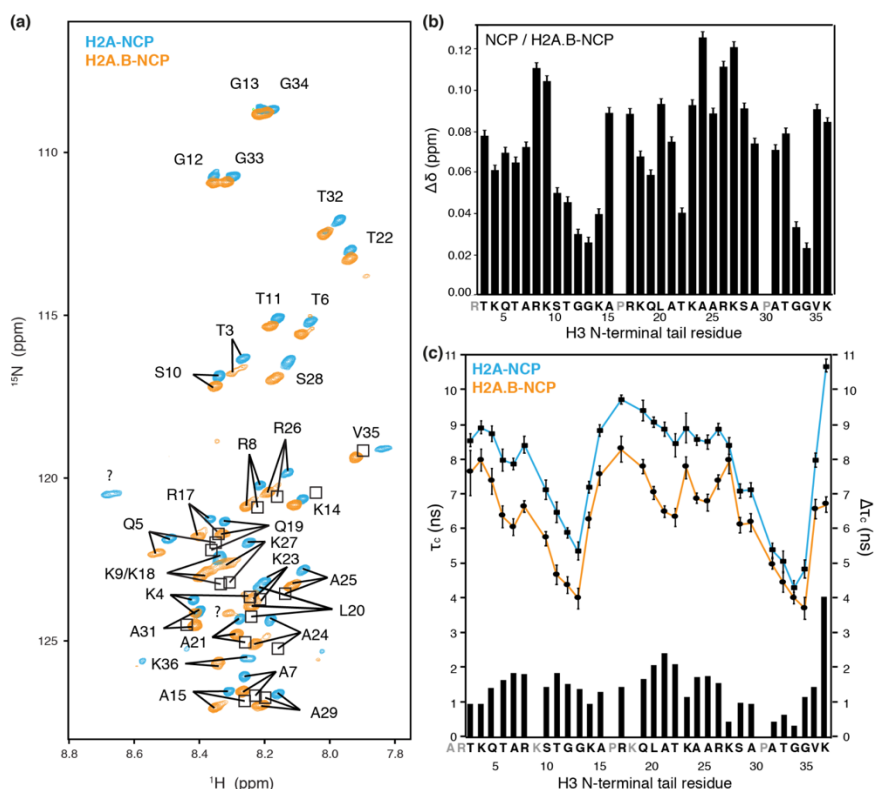


Figure 4.5. The H3 N-terminal tail has increased flexibility and reduced DNA-binding in Bbd-nucleosomes. (a) Overlay of TROSY spectra of nucleosomes containing ^{15}N -labelled H3 for both canonical (cyan) and H2A.B (orange) nucleosomes. Only highly flexible residues can be observed corresponding to the H3-tail. Peak positions for H3 tail peptide recorded by the Selenko lab ²⁶ are marked as black boxes. (b) Chemical shift perturbation ($\Delta\delta$) between canonical and variant nucleosomes for the H3 N-terminal residues. (c) Effective local rotational correlation times (τ_c) of H3 tail residues in canonical (squares/cyan) and variant (circles/orange) nucleosomes, together with differences in τ_c for each residue shown as bars ($\Delta\tau_c$).

register shift of an arginine that anchors to the DNA minor groove in the canonical nucleosome. This residue is in the L2 loop that connects the $\alpha 2$ and $\alpha 3$ helices. While in the canonical H2A sequence this is part of a sequence K74-T75-R76, the arginine is shifted by one position in the corresponding H2A.B sequence: E79-R80-N81. This register shift of the minor groove anchoring arginine causes a major clash between H2A.B R80 and the nucleosomal DNA backbone when in the

canonical, wrapped conformation (Figure 4.6a). Given the position of R80 in the L2 loop which is close to the DNA at super helix location (SHL) ± 5.5 , this clash could promote unwrapping of the nucleosomal DNA down to the next anchoring point at SHL ± 4.5 ²⁷, thus resulting in effective wrapping of 100-120 bp. Of note, the register shift of this arginine is conserved in H2A.B across most species (see Figure 4.1a and S1). To verify whether the arginine side chain conformation could adapt to the register shift and form a favourable interaction with the DNA nonetheless, two models for the H2A.B variant nucleosome were subjected to energy-minimization. While the H2A.B conformation is modelled based on the canonical H2A structure in the first model (Figure 4.6b), the other is based on the solution structure of the H2A.B-H2B histone fold core (Figure 4.6c), extended with its N- and C- tail region (including the docking domain) modelled based on canonical H2A. For both models and both copies of H2A.B in each model, R80 is not involved favourable interactions to the DNA after energy minimization (Figure 4.6b,c).

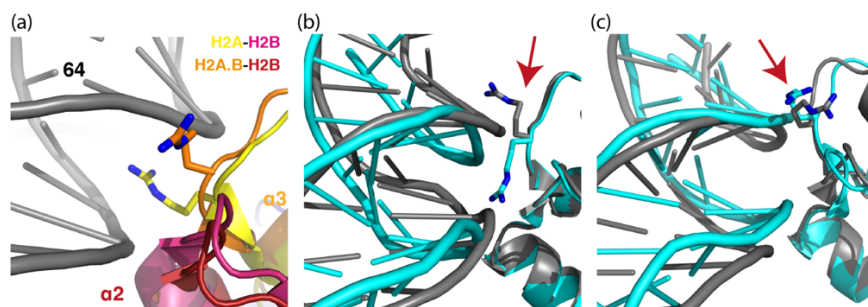


Figure 4.6. Register shift of the minor groove anchoring arginine in H2A.B. (a) Zoom on the $\alpha 2$ - $\alpha 3$ loop of H2A.B superimposed with the canonical nucleosome structure highlight the clash of Arg80 with DNA. (b, c) Energy-minimized structural model zoomed on the $\alpha 2$ - $\alpha 3$ loop showing an ‘escaped’ conformation of the Arg80. The two H2A.B $\alpha 2$ - $\alpha 3$ loop regions in each model are superimposed, coloured cyan and grey respectively. The energy minimized model starting with canonical homology model is shown in (b), and starting with H2A.B-H2B dimer solution structure shown in (c).

H2A.B L2 loop destabilizes nucleosomal DNA wrapping

To test our hypothesis that the arginine register shift in the H2A.B L2 loop disrupts the local interaction between DNA and histone octamer, we compared electrophoretic mobilities and micrococcal nuclease

(MNase) DNA digestion patterns of nucleosomes reconstituted from an H2A mutant containing the arginine register shift to that of H2A.B nucleosomes. The H2A mutant (H2A_ERN) was constructed by replacing the L2 sequence KTR in H2A to the corresponding sequence ERN from H2A.B.

Nucleosomes containing H2A_ERN could successfully be reconstituted using both purified histone octamers and mixtures of H2A_ERN-H2B dimers and H3-H4 tetramers with DNA (Figure 4.7a). This indicates that the mutant can form stable octamers with other core histones as expected from the presence of the complete docking domain. Thanks to its opened conformation, reconstituted H2A.B nucleosomes show a reduced electrophoretic mobility and results in a wider band appearance than the canonical nucleosome, as was reported before ¹⁴. The H2A_ERN nucleosome showed a markedly reduced mobility, similar to the variant, indicating the mutation resulted in a similar open conformation as the variant nucleosome. Nevertheless, the bands for the mutant nucleosomes are sharper than for the H2A.B nucleosome, indicating a more homogenous conformation for the mutant (Figure 4.7a).

We next subjected the three different types of nucleosomes to MNase treatment to assess their DNA accessibility. Samples taken at increasing lengths of MNase treatment show a progressively increased mobility and decreased intensity of the nucleosome band (Figure 4.7b). The increase in mobility reflects the trimming of the exposed DNA in the nucleosome particles, while the gradual disappearance of the band suggests disassembly of the nucleosome after excessive trimming of the nucleosomal DNA, possibly in combination with destabilization of the nucleosome during the electrophoresis and high salt exposure in the MNase quenching buffer. The lower part of most lanes shows the resulting bands from digested or undigested nucleosomal DNA fragments that are striped off from the octamer. The size of the digested DNA fragments indicates the number of base pairs that are protected in the nucleosome particle.

The canonical nucleosomes are the most resistant to MNase digestion, as shown by the modest loss of nucleosome band intensity (Figure 4.7b, lanes 1-6). As expected, these nucleosomes also protect the most of the nucleosomal DNA, as evidenced by the only gradual and modest increase in nucleosome mobility upon MNase treatment and the

gradual shift in size of the free DNA fragments from 167 to ~140 bp after 30 min treatment. Analysis of the DNA extracted after the protein digestion treatment shows fragments in the range of 130-150 bp (Figure 4.7c), which is typical for a properly folded, canonical nucleosome.

For H2A.B, MNase treatment resulted in rapid disappearance of the nucleosome band under our conditions (Figure 4.7b, lanes 13-18). Almost immediately after start of the DNA digestion, a free DNA band appeared at ~130 bp (estimated by comparison to free DNA band of canonical nucleosomes after 5 min MNase treatment), in line with the exposed nature of the DNA ends. Continued MNase digestion resulted to smaller fragments to around ~120 bp after 5 min of MNase digestion, as has been observed also in a previous study ¹⁵. The DNA extracted at the end of the MNase treatment shows a broad size range, with most fragments around 80-120 bp.

For the H2A_ERN mutant nucleosomes, like for the variant, a free DNA band at ~130 bp appeared quickly after addition of the enzyme, indicating the DNA ends are readily accessible also in the mutant. After 10 min of MNase treatment the digested DNA band shifted towards ~110-120 bp, which was also the size of majority of DNA fragments extracted at the end of the treatment. Thus, these data demonstrate that the ERN substitution in the L2 loop increases DNA accessibility and decreases the protection of the nucleosomal DNA. While the digestion pattern of the mutant does not recapitulate the high sensitivity of the variant towards MNase, we conclude that the arginine register shift contributes to the opening of the DNA in H2A.B nucleosomes.

Discussion

Histone variant proteins can alter nucleosome functionality by their different amino acid sequences compared to canonical histones, ranging from substitution of few amino acids as in H3.3 to addition of a folded protein domain as in macroH2A. Resolving the impact of these sequence changes on the nucleosome is essential to understand the function of these variants in chromatin biology. The histone variant H2A.B present a particularly interesting case as one of the most divergent histone variants, and as a variant that affects nucleosome conformation in a profound way, causing partial unwrapping of the

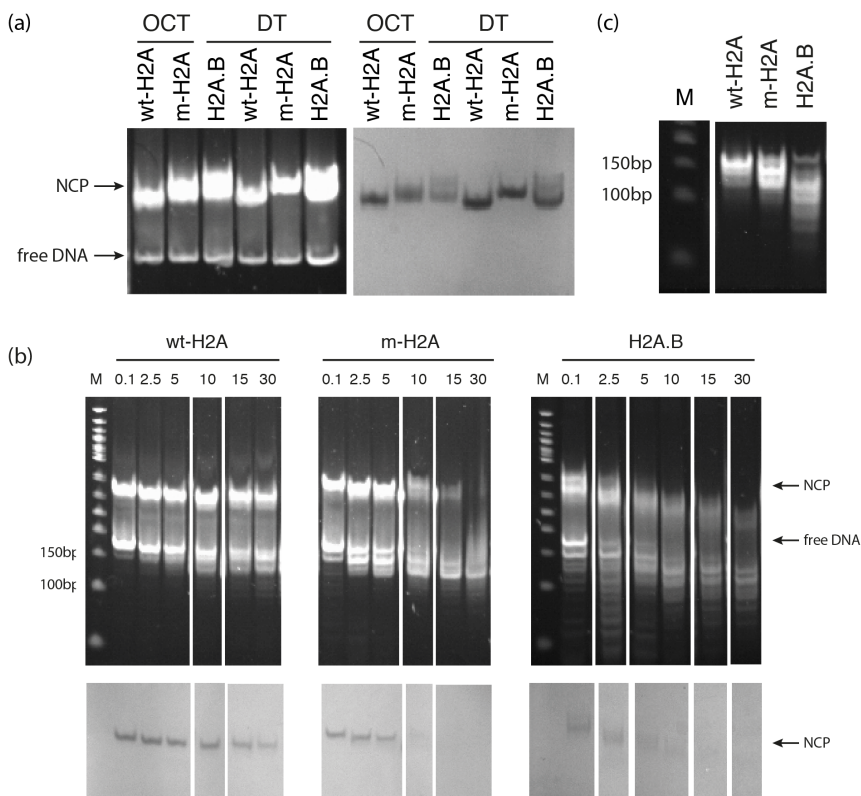


Figure 4.7. A register shift of a minor groove anchoring Arg causes destabilization of canonical nucleosomes. (a) Nucleosomes reconstituted using octamer and DNA (OCT) or dimer, tetramer and DNA (DT) were compared for canonical H2A (wt-H2A) and H2A_ERN (m-H2A) nucleosomes on the 5% native gel, stained for DNA using GelRed (left) and for proteins using Coomassie (right). (b) 5 μ g of each H2A (wt-H2A), H2A_ERN (m-H2A) and H2A.B nucleosomes reconstituted by DT were digested with 0.6U MNase for the indicated amount of time (0.1, 2.5, 5, 10, 15, 30 min). MNase were deactivated before loading the sample on the 5% native gel, stained for DNA (GelRed, top) and protein (Coomassie, bottom). Lanes are reordered according to time of treatment for clarity. (c) Samples were deproteinized after 30 min and the extracted DNA was loaded on the 5% native gel.

nucleosomal DNA. Yet, how the sequence changes in H2A.B affect the histone octamer conformation is largely unaddressed. Undoubtedly, this is in part related to the dynamic nature of the particle, complicating both crystallography and cryo-EM efforts. Here, we used state-of-the-art solution NMR to map the structure and

dynamics of the key sites in the H2A.B nucleosome at atomic resolution. In addition, we showed that a shift of critical arginine residue by one position can directly affect DNA wrapping.

The first part of our study focussed on assessing the nature of H2A.B in the variant nucleosome, in a way to verify the simplistic model of the H2A.B nucleosome in Figure 4.1. Our previous study of the H2A.B-H2B heterodimer showed that the histone core maintains the canonical fold in the variant dimer, and demonstrated that the docking domain in H2A.B is unstructured and highly flexible after the α C helix, which is incompatible with nucleosome formation. Using methyl TROSY, we show here that I106 within the H2A.B docking domain experiences a large change in chemical environment upon incorporation into nucleosome. Relaxation and NOE data further showed that the I106 δ 1 methyl group has relaxation rates comparable to histone core residues and show an NOE that is fully consistent with folding of the docking domain in the canonical fold. Only the very C-terminal residue of the H2A.B docking domain, which is predicted to protrude from the H3-H4 surface, remains highly flexible in the variant nucleosome. Taken together our NMR data indicate that, very much like canonical nucleosomes, the H2A.B docking domain is folded in the variant nucleosome in the same manner as H2A in canonical nucleosomes.

Moreover, while the H2A.B nucleosome particle is seen as a dynamic system due to the (transient) unwrapping of DNA giving rise to flexible DNA protrusions, this does not seem to be mirrored in increased dynamics for the histone octamer core. The variant dimer has been shown to have high mobility in cells suggesting a low barrier to exchange out of the nucleosome¹¹⁻¹². This may be caused by the reduced interactions between H2A.B and H3 because of the truncation of the docking domain and reduced net negative charge for the remaining part (see Chapter 3). Beautiful work from the Luger and Kay labs showed that destabilization of the dimer-tetramer interface can result in conformational changes that generate very large effects in CPMG relaxation dispersion studies²⁸. None such effects were detected for the H2A.B nucleosome, suggesting no large conformational changes either in the H2A.B core region or the docking domain. The increased ^1H transverse relaxation rates are likely rather

due to increase in effective size of the H2A.B nucleosome due to the partial DNA unwrapping.

We found significant changes in conformational dynamics for the H3 N-terminal tail, most likely directly related to the unwrapping of the DNA. The overall downfield shift, indicative of changes in the local electrostatic environment ²⁹, for the H3 tail resonances suggests that the DNA binding equilibrium is shifted towards the unbound state in the variant nucleosome. In particular, some resonances in the tail are have chemical shifts very similar to that free peptide, suggesting that some stretches of the tail may be free from the DNA in the H2A.B-nucleosome. However, since the profile of the residue-specific local effective correlation times was almost unchanged from the canonical nucleosome, we conclude that the tail remains associated to the DNA, rapidly sampling many different conformations. Given the length of the tail and the local nature of the DNA-tail interactions, this likely means that the tail as a whole remains DNA bound at all times, but the precise location of the contacts points switches rapidly in time and each contact spends significant time in an unbound state. The region where the chemical shifts most strongly resemble that of the free peptide, A29-V35, may be free at all times, but nevertheless indirectly linked to the DNA, thus limiting the flexibility. We thus amend the cartoon of the H2A.B nucleosome of Figure 4.1 by noting that the N-terminal part of the H3 tail is in fact bound to the DNA, likely limiting the opening angle between the DNA ends.

The second part of this study addressed the molecular basis for the partial DNA unwrapping in the H2A.B nucleosome. Previous work showed that truncation of the docking domain itself is not a main causative factor for DNA unwrapping ¹⁴, but rather that sequence differences in the α 3-helix to C-terminal are responsible ¹⁷. Our work establishes that this region of H2A.B is stably incorporated into nucleosomes, most likely in the same conformation as canonical H2A. This points to a rather subtle and indirect effect as this region is not involved in any DNA contacts directly. We found that the shifted arginine position in the L2 loop of H2A.B contributes directly to the open nucleosomal conformation as it is incompatible with anchoring to the DNA minor groove in the wrapped state. A H2A mutant containing the ERN sequence from the H2A.B L2 loop forms a nucleosome that is more vulnerable to MNase digestion and protects

118 bp DNA just like H2A.B nucleosomes. Nevertheless, the H2A_ERN nucleosome does not recapitulate the reduced stability of the H2A.B nucleosomes, in accordance for an additional and major role for sequence differences within the docking domain. We previously pointed to a reduced electrostatic match between the H2A.B docking domain and H3-H4 that may subtly destabilize the dimer/tetramer interface (see Chapter 3). We hypothesize that the reduced grip of H2A.B on both DNA and H3 may permit the H3 α N helix to transiently detach from the octamer core and thereby promote opening of the DNA. Yet, this process cannot involve a highly dynamic state of the α N helix is as that should have resulted in observable resonances in the amide backbone based NMR experiments. Interestingly, the arginine shift identified here is also present in other truncated histone variants, suggesting that the underlying DNA unwrapping mechanism is conserved.

In summary, we found that the histone octamer in the H2A.B nucleosome is remarkably like the canonical nucleosome. We provide evidence that that the docking domain is folded in the canonical configuration and as rigid as the histone core, pointing to defined interaction to H3. We find that the opened state of the variant nucleosome correlates with increased flexibility and reduced DNA binding for H3 N-terminal. Finally, we here show that a register shift of a minor groove anchoring arginine in the H2A.B L2 loop is an important and direct contributing cause for DNA unwrapping in H2A.B nucleosomes. With these results we have provided new insights in the structure of the H2A.B nucleosome. Furthermore, we look forward to combining these data with the outcome of ongoing MD simulations to further dissect the molecular nature of this peculiar nucleosome.

Material and methods

Site-directed mutagenesis. The plasmid containing H2A_ERN was produced by site-direct mutagenesis on plasmid pET-21b containing the *Drosophila* H2A_K74E made in Chapter 3. A pair of complementary mutagenic primers was designed to introduce 2 mutations T75R and R76N at once. The reaction was performed using

Site-Direct Mutagenesis Kit (Novagen) according to the following thermocycling: 12 cycles of 95 °C for 30 s, 52 °C for 30 s, 68 °C for 8 min. After the reaction, the 1 µL of DpnI was added and incubated at 37 °C for 1 hr to remove the original template. The generated nicked circular DNA was used to transform *E.coli* DH5α cells. Plasmids were extracted from successful transformations and the mutations were verified by sequencing.

Histone protein production. *Drosophila melanogaster* (Dm.) canonical histones H2A (Uniprot-id: P84051), H2B (Uniprot-id: P02283), H3 (Uniprot-id: P02299), H4 (Uniprot-id: P84040), *Homo sapiens* (Hs.) H2A.B (Uniprot: P0C5Z0), and mutant H2A_ERN were expressed in *E. coli* BL21 Rosetta 2 (DE3) cells (Novagen) and purified under denaturing conditions from inclusion bodies by extraction in 6 M guanidine chloride, followed by size-exclusion chromatography in buffer A (7 M urea, 50 mM NaPi pH 7.5, 1 mM EDTA, 150 mM NaCl, 5mM BME) using a Superdex 75 column (GE) and ion exchange with a salt gradient from buffer A to buffer A supplemented with 1 M NaCl. Histones used for NMR studies were produced in D₂O or H₂O-based M9 minimal medium containing desired isotopes and amino acid precursors in case of ILV methyl-labelling ³⁰. Histones used for MNase assays were produced in LB medium.

DNA production. The same protocol as described in a previous publication ²⁰ was followed to produce the 167 bp ‘601’-Widom DNA used in this study. In short, a pUC19 plasmid containing 12 repeats of a 167 bp ‘601’-Widom sequence was amplified in *E. coli* DH5α and extracted by alkaline lysis followed by isopropanol and ethanol precipitations prior to dissolve in TE buffer (20 mM Tris-HCl, 5 mM EDTA, 100 mM NaCl, pH 7.5) and purify by anion exchange chromatography. The purified plasmid was restricted using *ScaI* (ThermoFisher) and the restricted fragment was purified by anion exchange chromatography.

Nucleosome reconstitution. Histone dimers, tetramers or octamers were refolded from equimolar mixes of denatured purified histones by dialysis to 2 M NaCl at 4°C ³¹. Histone dimers, tetramers or octamers

were subsequently purified by size-exclusion chromatography over a Superdex 200 column (GE) in 2 M NaCl buffer with an addition of 5 mM BME (see also Figure S4.2). H2A.B nucleosomes were reconstituted from a mix of H2A.B-H2B dimer, H3-H4 tetramer, and 167 bp ‘601’ DNA at ratio 2:1:1. H2A nucleosomes and H2A_ERN nucleosomes were either made the same way as H2A.B nucleosomes or directly from a mix of octamer and 167 bp ‘601’ DNA at ratio 1:1. The mixes were dialysed at 4 °C from high salt buffer (10 mM KPi pH 6.5, 2 M KCl) to low salt buffer (10 mM KPi pH 6.5, 250 mM KCl) over 18 hours using a pump. Efficiency of the reconstitutions was analysed with 15% SDS-PAGE and 5% native-PAGE (see also Figure 4.2 and Figure S4.2). Nucleosomes were buffer exchanged to lower salt concentration for NMR or MNase studies (see below).

MNase study. Nucleosomes containing H2A, H2A_ERN, or H2A.B reconstituted from dimer, tetramer and DNA were first dialyzed to buffer with 10mM NaPi pH 6.5, 50 mM KCl. 4.5µg of each nucleosome was mixed with MNase reaction buffer (10 mM Tris pH 7.5, 5 mM Ca²⁺) to a final volume of 30 µL per reaction. Reactions were incubated at 37 °C with 0.6 U MNase. At 0.1, 2.5, 5, 10, 15, and 30 min, 4 µL of each reaction mix was sampled, and the reaction stopped by adding 1 µL MNase stop buffer (100 mM EDTA pH 8.0). Reaction mixes after 30 min were deproteinized by 0.6 U proteinase K at 55°C for 90 min. DNA was extracted by precipitation with Na-acetate and 96% ethanol and re-suspended with 0.2×TBE buffer. All samples were analysed with 5% native-PAGE.

Molecular modelling. The simplistic model of H2A.B nucleosome as shown in Figure 4.1c was created based on the canonical nucleosome structure (PDB id: 1KX5) as a template. The path of nucleosomal DNA entry/exit ends was constructed by replacing the terminal 13 bp with straight B-form DNA in PyMOL software³². This model is to match the reported 118 bp DNA protected by the H2A.B containing octamer¹⁴.

The H2A.B nucleosome structure used to initiate MD simulations shown in Figure 4.6b is a homology model based on the canonical nucleosome structure (PDB id: 1KX5), or, for Figure 4.6c, was built using the H2A.B solution structure for the core region with C-terminal

and N-terminal structure modelled according to canonical conformation (PDB id: 1KX5) using MODELLER³³.

NMR spectroscopy. For H3 tail dynamics study, ¹⁵N/¹³C labelled H3 was reconstituted into nucleosomes with unlabelled H2A or H2A.B, H2B, H4 and DNA. Samples of H2A.B nucleosomes (32 μM) or H2A nucleosomes (93 μM) in NMR buffer 1 (10 mM KPi pH 6.5, 10 mM KCl, 7.5% D₂O) were measured at 303 K on a Bruker Avance III spectrometer equipped with a cryo-probe and operating at 22.3 T corresponding to 950 MHz ¹H Larmor frequency. Assignments for H3 tail residues in H2A.B nucleosomes were transferred from canonical nucleosomes published before²⁰. T₁, T₂ relaxation experiments were analysed using PINT software³⁴.

For comparison of H2A.B N-terminal tail resonances between H2A.B-H2B heterodimer and H2A.B nucleosomes, ¹⁵N labelled H2A.B was reconstituted into either dimers or nucleosomes with unlabelled H2B, H3, H4 and DNA. Samples of H2A.B-H2B dimer (20 μM) or H2A.B nucleosomes (20 μM) were carried out at 303 K in buffer 2 (20 mM NaPi pH 6.5, 5 % D₂O, 0.01 % NaN₃) on a Bruker Avance III spectrometer operating at 20.0 T corresponding to 850 MHz ¹H Larmor frequency. Assignments of H2A.B were transferred from Chapter 3.

Methyl-labelled nucleosomes and methyl-labelled H2A.B-H2B dimers were prepared with Ile-δ1-[¹³CH₃], Leu/ Val-[¹³CH₃, ¹²CD₃]-H2A.B or H2A and perdeuterated H2B, H3, and H4 with unlabelled DNA. Data for methyl-labelled nucleosomes were collected in NMR buffer 3 (20 mM NaPi pH 6.2, 0.01% NaN₃, 0.5 mM PMSF, 100% D₂O), while methyl labelled H2A.B-H2B dimers were measured in NMR buffer 3 with 200mM NaCl. Methyl-TROSY spectra of H2A.B-H2B dimer (160 μM), H2A.B nucleosome (85 μM), canonical nucleosome (100 μM) were measured on a Bruker Avance III spectrometer operating at 14.1 T corresponding to 600 MHz ¹H Larmor frequency. Nucleosomes spectra were recorded at temperature of 318 K, while dimer spectrum was recorded at 303 K. Spectra were processed to equal acquisition times with identical processing parameters. For each sample, the peak intensities were normalized based on the median value of each spectrum to allow comparison of the intensity distribution between the

different systems. N-terminal tail residues were excluded in this analysis.

CPMG relaxation dispersion experiments of H2A.B nucleosomes were performed at 318 K on a Bruker Avance III spectrometer operating at 14.1 T corresponding to 600 MHz ¹H Larmor frequency. For the measurement of ¹H single quantum relaxation rates, a relaxation delay of 10 ms was used to collect 11 CPMG pulse frequencies ranging from 100-1000 Hz with three duplicate points for experimental error assessment. For the multiple-quantum CPMG relaxation experiments, a relaxation delay of 15 ms was used to collect 13 CPMG pulse frequencies ranging from 67- 2000 Hz with three duplicate points for experimental error assessment. All dispersion relaxation data were analysed using PINT ³⁴.

NOESY experiments of H2A.B nucleosomes were measured with mixing time 200 ms at 318 K on a Bruker Avance III spectrometer operating at 22.3 T corresponding to 950 MHz ¹H Larmor frequency, equipped with a cryo-probe.

Acknowledgements

We thank Lewis Kay (University of Toronto) for providing the CPMG relaxation dispersion pulse sequences.

References

1. Draizen, E. J.; Shaytan, A. K.; Marino-Ramirez, L.; Talbert, P. B.; Landsman, D.; Panchenko, A. R., HistoneDB 2.0: a histone database with variants--an integrated resource to explore histones and their variants. *Database (Oxford)* **2016**, *2016*.
2. Long, M.; Sun, X.; Shi, W.; Yanru, A.; Leung, S. T. C.; Ding, D.; Cheema, M. S.; MacPherson, N.; Nelson, C. J.; Ausio, J.; Yan, Y.; Ishibashi, T., A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res* **2019**.
3. Talbert, P. B.; Henikoff, S., Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* **2017**, *18* (2), 115-126.
4. Tachiwana, H.; Kagawa, W.; Shiga, T.; Osakabe, A.; Miya, Y.; Saito, K.; Hayashi-Takanaka, Y.; Oda, T.; Sato, M.; Park, S. Y.; Kimura, H.; Kurumizaka, H., Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **2011**, *476* (7359), 232-5.

5. Ismail, I. H.; Hendzel, M. J., The gamma-H2A.X: is it just a surrogate marker of double-strand breaks or much more? *Environ Mol Mutagen* **2008**, *49* (1), 73-82.
6. Chadwick, B. P.; Willard, H. F., A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* **2001**, *152* (2), 375-84.
7. Tolstorukov, M. Y.; Goldman, J. A.; Gilbert, C.; Ogryzko, V.; Kingston, R. E.; Park, P. J., Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol Cell* **2012**, *47* (4), 596-607.
8. Soboleva, T. A.; Nekrasov, M.; Pahwa, A.; Williams, R.; Huttley, G. A.; Tremethick, D. J., A unique H2A histone variant occupies the transcriptional start site of active genes. *Nat Struct Mol Biol* **2011**, *19* (1), 25-30.
9. Chen, Y.; Chen, Q.; McEachin, R. C.; Cavalcoti, J. D.; Yu, X., H2A.B facilitates transcription elongation at methylated CpG loci. *Genome Res* **2014**, *24* (4), 570-9.
10. Soboleva, T. A.; Parker, B. J.; Nekrasov, M.; Hart-Smith, G.; Tay, Y. J.; Tng, W. Q.; Wilkins, M.; Ryan, D.; Tremethick, D. J., A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS Genet* **2017**, *13* (2), e1006633.
11. Arimura, Y.; Kimura, H.; Oda, T.; Sato, K.; Osakabe, A.; Tachiwana, H.; Sato, Y.; Kinugasa, Y.; Ikura, T.; Sugiyama, M.; Sato, M.; Kurumizaka, H., Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci Rep* **2013**, *3*, 3510.
12. Gautier, T.; Abbott, D. W.; Molla, A.; Verdel, A.; Ausio, J.; Dimitrov, S., Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep* **2004**, *5* (7), 715-20.
13. Montel, F.; Menoni, H.; Castelnovo, M.; Bednar, J.; Dimitrov, S.; Angelov, D.; Faivre-Moskalenko, C., The dynamics of individual nucleosomes controls the chromatin condensation pathway: direct atomic force microscopy visualization of variant chromatin. *Biophys J* **2009**, *97* (2), 544-53.
14. Bao, Y.; Konesky, K.; Park, Y. J.; Rosu, S.; Dyer, P. N.; Rangasamy, D.; Tremethick, D. J.; Laybourn, P. J.; Luger, K., Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J* **2004**, *23* (16), 3314-24.
15. Doyen, C. M.; Montel, F.; Gautier, T.; Menoni, H.; Claudet, C.; Delacour-Larose, M.; Angelov, D.; Hamiche, A.; Bednar, J.; Faivre-Moskalenko, C.; Bouvet, P.;

Dimitrov, S., Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *EMBO J* **2006**, 25 (18), 4234-44.

16. Montel, F.; Fontaine, E.; St-Jean, P.; Castelnovo, M.; Faivre-Moskalenko, C., Atomic force microscopy imaging of SWI/SNF action: mapping the nucleosome remodeling and sliding. *Biophys J* **2007**, 93 (2), 566-78.

17. Shukla, M. S.; Syed, S. H.; Goutte-Gattat, D.; Richard, J. L.; Montel, F.; Hamiche, A.; Travers, A.; Faivre-Moskalenko, C.; Bednar, J.; Hayes, J. J.; Angelov, D.; Dimitrov, S., The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling. *Nucleic Acids Res* **2011**, 39 (7), 2559-70.

18. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, 389 (6648), 251-60.

19. Dai, L.; Xie, X.; Zhou, Z., Crystal structure of the histone heterodimer containing histone variant H2A.Bbd. *Biochem Biophys Res Commun* **2018**, 503 (3), 1786-1791.

20. Xiang, S.; le Paige, U. B.; Horn, V.; Houben, K.; Baldus, M.; van Ingen, H., Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **2018**, 57 (17), 4571-4575.

21. Zhou, B. R.; Feng, H.; Ghirlando, R.; Kato, H.; Gruschus, J.; Bai, Y., Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. *J Mol Biol* **2012**, 421 (1), 30-7.

22. Tugarinov, V.; Hwang, P. M.; Ollershaw, J. E.; Kay, L. E., Cross-correlated relaxation enhanced ¹H[¹³C] NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **2003**, 125 (34), 10420-8.

23. Kato, H.; van Ingen, H.; Zhou, B. R.; Feng, H.; Bustin, M.; Kay, L. E.; Bai, Y., Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A* **2011**, 108 (30), 12283-8.

24. Zheng, C.; Lu, X.; Hansen, J. C.; Hayes, J. J., Salt-dependent intra- and internucleosomal interactions of the H3 tail domain in a model oligonucleosomal array. *J Biol Chem* **2005**, 280 (39), 33552-7.

25. Korolev, N.; Allahverdi, A.; Yang, Y.; Fan, Y.; Lyubartsev, A. P.; Nordenskiöld, L., Electrostatic origin of salt-induced nucleosome array compaction. *Biophys J* **2010**, 99 (6), 1896-905.

26. Stutzer, A.; Liokatis, S.; Kiesel, A.; Schwarzer, D.; Sprangers, R.; Soding, J.; Selenko, P.; Fischle, W., Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Mol Cell* **2016**, *61* (2), 247-59.
27. Muthurajan, U. M.; Park, Y. J.; Edayathumangalam, R. S.; Suto, R. K.; Chakravarthy, S.; Dyer, P. N.; Luger, K., Structure and dynamics of nucleosomal DNA. *Biopolymers* **2003**, *68* (4), 547-556.
28. Kitevski-LeBlanc, J. L.; Yuwen, T.; Dyer, P. N.; Rudolph, J.; Luger, K.; Kay, L. E., Investigating the Dynamics of Destabilized Nucleosomes Using Methyl-TROSY NMR. *J Am Chem Soc* **2018**, *140* (14), 4774-4777.
29. Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B., Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61-66.
30. Tugarinov, V.; Kanelis, V.; Kay, L. E., Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* **2006**, *1* (2), 749-54.
31. Dyer, P. N.; Edayathumangalam, R. S.; White, C. L.; Bao, Y.; Chakravarthy, S.; Muthurajan, U. M.; Luger, K., Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **2004**, *375*, 23-44.
32. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
33. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* **2016**, *86*, 2 9 1-2 9 37.
34. Ahlner, A.; Carlsson, M.; Jonsson, B. H.; Lundstrom, P., PINT: a software for integration of peak volumes and extraction of relaxation rates. *J Biomol NMR* **2013**, *56* (3), 191-202.
35. Taguchi, H.; Horikoshi, N.; Arimura, Y.; Kurumizaka, H., A method for evaluating nucleosome stability with a protein-binding fluorescent dye. *Methods* **2014**, *70* (2-3), 119-26.

Supplements

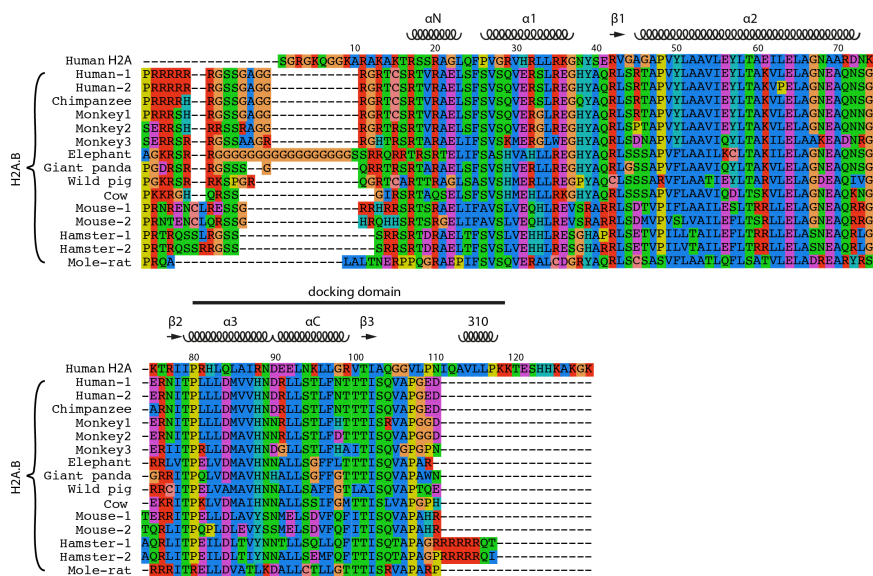


Figure S4.1. Sequence alignment of H2A.B with canonical H2A. Fifteen H2A.B sequences from different species that are available on *Histone Database 2.0*¹ are aligned and compared to human canonical H2A. Secondary structural elements of H2A in the canonical nucleosome structure are indicated above the sequences.

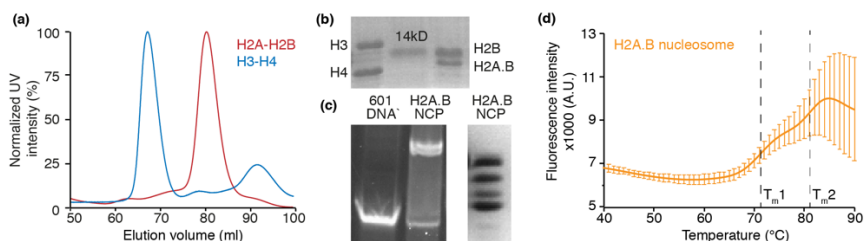


Figure S4.2. Reconstitution of H2A.B-nucleosomes. (a,b) Chromatograph and SDS-PAGE of purification of refolded H2A.B-H2B dimers and H3-H4 tetramers used in the nucleosome reconstitution. (c) Native PAGE (left) and SDS-PAGE (right) analysis of reconstituted H2A.B nucleosomes used for subsequent NMR studies. (d) Thermal melting curves of H2A.B-nucleosomes as measured in a SYPRO-orange based thermostability assay. Thermal shift assay showed the presence of two-phase melting transition, with a first transition occurring at 71 (T_{m1}) and a second at 81 °C (T_{m2}), correspondingly well to previously published

values ³⁵. These transitions correspond to stepwise release of the H2A.B-H2B dimer and H3 and H4 histones respectively.

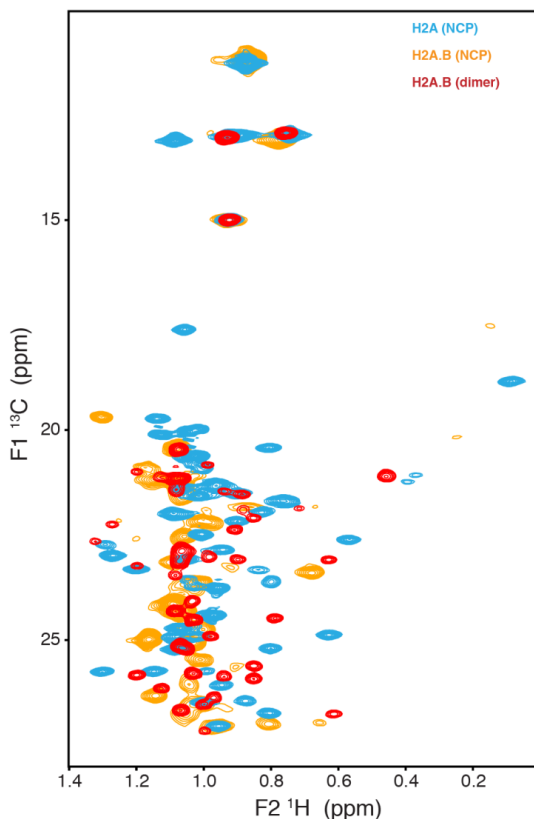


Figure S4.3. Complete Me-TROSY spectra overlay for H2A.B in either nucleosomes (orange) or H2A.B-H2B heterodimer (red), and H2A in canonical nucleosomes (cyan). The H2A.B spectra of nucleosome and heterodimer show significant changes in overall peak pattern. Similar differences between dimer and nucleosome chemical shifts were observed for canonical H2A and attributed to the formation of the large interfaces to the DNA, H3-H4 and the second copy of the H2A-H2B dimer in the nucleosome ²⁰. The changes observed here are thus consistent with formation of a well-folded nucleosome.

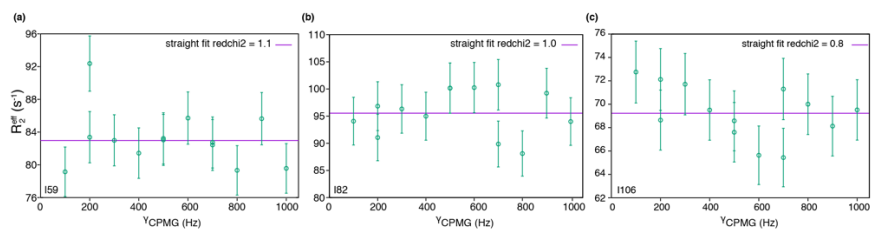


Figure S4.4 ^1H SQ CPMG relaxation dispersion profiles for H2A.B isoleucine's I59 (a), I82 (b), and I106 (c) in nucleosomes.

Chapter 5. The electrostatic potential of the nucleosome acidic patch

This chapter is based on:

Heyi Zhang, Jelmer Eerland, Velten Horn, Hugo van Ingen. The electrostatic potential of the nucleosome acidic patch. *In preparation*.

Contributions of authors:

Amide backbone titration study and histidine side chain titration study were performed by Jelmer Eerland and Velten Horn.

Abstract

The nucleosome surface contains an area with negative electrostatic potential known as the acidic patch, which functions as a binding platform for various proteins to modulate chromatin biology. The dense clustering of acidic residues in this patch may impact their effective pKa and thus the effective electrostatic potential of the acidic patch, which in turn could influence nucleosome-protein interactions. Here, we used solution NMR to determine the structure of the H2A-H2B dimer and its acidic patch free in solution, showing that its core structure, including the acidic patch, agrees well with the structure within the nucleosome. Using pH titration experiments the pKa values for acidic residues as well as histidines that compose or surround the acidic patch were experimentally determined. While for several residues elevated pKa's were predicted based on the structure, the experimental pKa values for glutamate carboxyl group at the acidic patch are all below 4.5. For residue H2A D89, part of the DEE acidic triad, a significantly elevated pKa of 5.6 was observed. Our results establish the order of protonation events in the acidic patch upon lowering pH from physiological values and establish H2B H106 as the first residue to titrate.

Introduction

As the core component of chromatin, nucleosomes serve as a major docking platform for a wide range of proteins that control chromatin biology ¹. Many of these proteins bind to nucleosomes through interaction with a distinct site on the central histone octamer surface, the acidic patch ². In addition, nucleosomes can self-interact via this surface, thereby mediating chromatin compaction ³⁻⁴. The acidic patch is formed by the close proximity of six acidic residues from histone H2A and two from H2B, resulting in a defined region of predicted negative electrostatic potential that stands out from the overall positively charged histone surface (**Figure 5.1a, 1b**). Acidic patch binding proteins invariably use one or more arginine side chains to anchor to the acidic patch via hydrogen bonding interactions (**Figure 5.1c**). The proximity of charged residues is known to affect the protonation energies of these titratable residues, possibly resulting in significant shifts in their pKa's ⁵⁻⁶. In addition, the acidic patch perimeter includes two highly conserved histidine residues (see **Figure 5.1b**) whose protonation degree may be very variable in the physiological range of pH. Thus, a priori it is not clear to what extent residues in and around the acidic patch are protonated under physiological conditions, neither to what extent they might change their protonation state due to subtle changes in local nuclear pH *in vivo*, or due to different buffer pH in experiments performed *in vitro*. Protonation or even fractional protonation of residues in or around the acidic patch may critically alter the surface charge potential and thereby impact the binding affinity of nucleosome-binding proteins. Indeed, unpublished *in vitro* experiments have shown that at least for chromatin factor HMGN2, a change in pH from 7.4 to 6.0 resulted in a ~1000-fold reduced binding affinity to the nucleosome (unpublished data from Hugo van Ingen, Hidenori Kato, Lewis E. Kay, Yawen Bai). Here, we aimed to experimentally determine the pKa values of titratable residues in and around the acidic patch using NMR spectroscopy. NMR spectroscopy is one of the few techniques that can reveal residue-specific side chain pKa's in proteins ^{5, 8-9}. Changes in the chemical shifts of the nuclear spins in and around the titratable group upon changes in pH are then used to extract pKa values for each

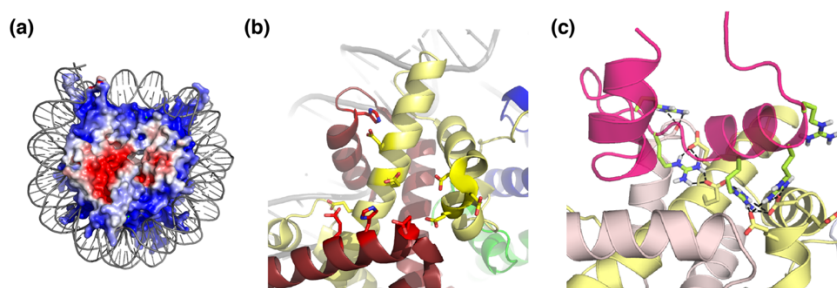


Figure 5.1. Close proximity of acidic residues from histone H2A and H2B forms negatively charged nucleosome surface and functions as arginine anchor of RNF168. (a) APBS predicted electrostatic potential ± 5 kT on solvent accessible surface of the histone octamer core in nucleosome (PDB id: 2PYO). (b) The close proximity of six acidic residues (E55, E60, E63, D89, E90, E91) from H2A and two (E102, E110) from H2B contribute to the acidic patch of the nucleosome. Histidine 46 and 106 from H2B locate around and in the acidic patch are also shown as sticks. H2A is shown in yellow, H2B in red, H3 in blue, H4 in green, and DNA in grey. (c) Acidic patch residues function as arginine anchors for the binding of RNF168 on nucleosomes ⁷.

group. Such methodology has been successfully demonstrated even for very large complexes such as the proteasome using methyl-specific labeling ¹⁰, suggesting that this is in principle also possible for the nucleosome. However, this approach relies on the proximity of a methyl-bearing residue to the titratable group of interest. Given the number and the proximity of methyl groups in and around the acidic patch, such indirect readout is unlikely to result in determination of all residue-specific pKa's. We thus resorted to the histone H2A-H2B heterodimer and applied a divide-and-conquer strategy, to be able to map the pKa's of as many titratable groups regardless of the proximity of a methyl-group. We show that in solution the isolated H2A-H2B heterodimer retains the histone-fold core as in the nucleosome and that two glutamate residues in H2A are predicted to have significantly elevated pKa's. By examining pH-dependent chemical shifts changes of both backbone amide and histidine, glutamate, and aspartate side chain groups, we determined the apparent pKa for all titratable groups in and around the acidic patch. Our results show that all acidic residues except H2A D89 have pKa values < 5 , H2A D89 has pKa value 5.6, and that pH changes around 6.5 cause large chemical shift perturbations around H2B H106 that lines the acidic patch. These results establish that the glutamate and aspartate residues in the acidic

patch are indeed deprotonated in the physiological range and thus capable to serve as hydrogen bond acceptors for arginine-anchor residues in chromatin binding proteins. Furthermore, the elevated pKa of H2B His106 suggest that it may be fractionally protonated in close-to-physiological conditions and thereby modulate the effective electrostatic potential of the acidic patch and thus the binding affinity of chromatin effector proteins.

Materials and methods

Histone protein production and dimer refolding. *Drosophila melanogaster* (Dm.) canonical histones H2A (Uniprot-id: P84051) and H2B (Uniprot-id: P02283) were expressed in *E. coli* BL21 Rosetta 2 (DE3) cells (Novagen) and purified under denaturing conditions from inclusion bodies by extraction in 6 M guanidine chloride, followed by size-exclusion chromatography in buffer A (7 M urea, 50 mM NaPi pH 7.5, 1 mM EDTA, 150 mM NaCl, 5mM BME) using a Superdex 75 column (GE) and ion exchange with a salt gradient from buffer A to buffer A supplemented with 1 M NaCl. Histones used for NMR studies were produced in D₂O or H₂O-based M9 minimal medium containing desired isotopes. Histone dimers were refolded from equimolar mixes of denatured purified histones by dialysis to 2 M NaCl at room temperature and subsequent purification using size-exclusion chromatography over a Superdex 200 column (GE) in 2M NaCl buffer¹¹. Purified dimers were stored at 4 °C before buffer exchange to lower salt concentration for NMR studies.

NMR spectroscopy. The backbone HN, N, C α , C β , C' chemical shifts of H2A and H2B in histone dimer were assigned in previous studies^{7, 12} (BMRB accession code 27547 and 27187). The H2A assignment were transferred to low pH conditions (20 mM MES pH5.8, 100 mM NaCl, 10% D₂O) with the aid of triple resonance experiments to confirm the previous assignment and to expand the assignment for the disordered N-terminal tail. Secondary structure of H2A-H2B dimer was predicted by TALOS-N using HN, N, C α , C β , C' chemical shifts¹³.

A 3D ^{15}N -edited NOESY with 200 ms mixing time was recorded on a sample containing ~ 0.5 mM dimer in NMR buffer (20 mM NaPi, 200 mM NaCl, pH 6.5, 5 mM β -mercaptoethanol). The dimer was refolded from perdeuterated ^{15}N -labeled H2B and unlabeled H2A to selectively record intermolecular NOEs.

For pH titration experiments, NMR buffers with pH range from 4.4 to 9.1 were controlled to 200 mM ionic strength with NaCl according to the prescription of the pH calculator website (Rob Beynon, University of Liverpool). 10% D_2O was added to the buffer prior to measurement. Samples were recovered after each titration point and buffer exchanged to next titration point.

For determining backbone amide pKa values, dimers refolded with $^{15}\text{N}/^{13}\text{C}$ labeled H2A and unlabeled H2B or ^{15}N labeled H2B and unlabeled H2A were used. HSQC spectra of H2A-H2B dimer (150 μM) in NMR buffers were recorded at 303K on Bruker Avance III HD spectrometers equipped with TCI cryoprobes and operating at 600 or 850 MHz Larmor frequency. Acquisition times was typically 2.5-3 hours per spectrum, with acquisition times of 150 ms for ^1H and 100 ms for ^{15}N dimension.

For determining histidine side chain amide pKa values, long-range ^1H - ^{15}N HMQC experiments¹⁴ were recorded on dimers refolded with ^{15}N -labeled H2B and unlabeled H2A (150 μM) at 303K on Bruker Avance III HD spectrometers equipped with TCI cryoprobes and operating at 600 MHz Larmor frequency.

To enhance sensitivity for detection of glutamate side chains, H2A was made in M9 medium in D_2O , with ^{13}C -protonated glucose, and ^{15}N -labeled NH_4Cl , to remove the alpha protons, and achieve about 50% deuteration on the aliphatic protons. Fractionally deuterated, fully $^{15}\text{N}/^{13}\text{C}$ labeled H2A was refolded with unlabeled H2B. The ^1H - ^{13}C carboxyl projection was recorded for 12h using H(C)CO type experiments to observe side chain carboxyl groups of glutamates and aspartates. Spectra were recorded at 303 K on a Bruker Avance III spectrometer equipped with a cryo-probe and operating at 22.3 T corresponding to 950 MHz ^1H Larmor frequency.

To assign side chains of glutamates and aspartates of H2A, several NMR experiments including ^{13}C -constant-time-HSQC, 3D HCCH, 3D CCCONH, and 3D TROSY version of HNCOCACB were recorded at 303 K on a Bruker Avance III spectrometer operating at 21.1 T

corresponding to 900 MHz ^1H Larmor frequency, equipped with a cryo-probe. H2A was either fully or fractionally deuterated and fully $^{15}\text{N}/^{13}\text{C}$ labeled in the dimer sample where H2B was unlabeled.

All NMR data was processed using Bruker Topspin, or NMRPipe and analyzed using NMRFAM-Sparky ¹⁵. Titration data were analyzed using a custom python script written by Hugo van Ingen. Figures of the dimer structure and nucleosome structure (Pdb: 2PYO) were generated in Pymol (The PyMOL Molecular Graphics System, Version 1.7, Schrödinger, LLC).

Structure calculation by CS-Rosetta. Structure calculation of the H2A-H2B dimer was first attempted using CS-Rosetta based only on backbone chemical shifts ¹⁶⁻¹⁷. The HN, N, C α , C β , C' backbone chemical shifts of the core regions of H2A (V26-S97) and H2B (Y34-K122) were used to calculate 3000 structures of the H2A-H2B histone fold core using the CS-Rosetta webserver (<https://csrosetta.bmr.b.wisc.edu/csrosetta/submit>). To allow Rosetta to fold the dimeric core, the two proteins were connected by a random coil (Gly)8 poly-glycine linker into a single chain. The 10 structures with the lowest Rosetta energy model were selected as the best models. However, the final ensemble of CS-Rosetta models was not converged enough to guide the assignment of intermolecular NOEs between H2A and H2B (see below).

Intermolecular NOE assignment by CYANA. ^{15}N -edited NOESY spectrum was recorded on a H2A-[U- ^2H , ^{15}N]-H2B sample. All H2B HN protons were assigned unambiguously based on backbone (HN, N) chemical shifts. Intermolecular NOEs were automatically assigned by CYANA using the atom and residue specific average chemical shift as deposited in the BMRB together with H2A-H2B dimer structure extracted from nucleosome crystal structure (PDB id: 2PYO) as structural reference. About 500 intermolecular NOE cross peaks were observed with chemical shift below 7 ppm. In order to unambiguously assign the large number of intermolecular NOEs observed, cross peaks were grouped and assigned individually according to peak intensities. Briefly, the most intense 35 intermolecular NOE cross peaks together with their potential network support peaks within 0.01 ppm range of each chemical shift were first assigned by CYANA using a tolerance

of 0.4 ppm, which result in 17 unambiguous assignment for the most intense NOEs and their 30 network support NOEs. The assigned atoms were removed from the prot list before the next round of CYANA assignment of median intense NOE peaks together with the unassigned most intense NOE peaks. This second round could not provide unique assignment due to multiple options available for one chemical shift.

The upper distance restraints for the 47 resulted unambiguous intermolecular NOEs were extracted from the .noa output file of CYANA assignment for the most intense peaks. To allow extra freedom of Rosetta structural sampling, an additional 0.5 Å was added to each upper distance restraints before Rosetta calculations.

Structure calculation by Rosetta with NOE distance restraints. The same protocol as described in Chapter 3 was used for calculating and selecting the final structural ensemble for H2A-H2B dimers. Briefly, the distance restraints derived from the intermolecular NOEs were used together with the chemical-shift generated fragments to calculate the structure of the H2A-H2B core region (H2A V26-S97 and H2B Y34-K122 connected by a poly-glycine linker) using the Rosetta AbInitio protocol as described previously (www.rosettacommons.org/demos/latest/public/abinitio_w_chemicals_hift_noe/README). In total, 3000 structures were generated and scored on the basis of their the full-atom and distance restraint energy (E) by Rosetta. For each model, backbone chemical shifts of $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^{13}\text{C}'$, ^1HN and ^{15}N were predicted by SPARTA+. The correspondence between predicted and experimental chemical shifts was used to rescore the initial Rosetta energy to a rescored energy (E'). The model with the lowest E' was selected as the reference model, and the 10 models with the lowest C α -RMSD from the 20 lowest energy models were selected as the final structure ensemble for the histone fold core. Structural statistics are reported in Supplementary Table S5.1. Predicted pKa's were calculated using PDB2PQR webserver (http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/). Ten structures within the ensemble for H2A-H2B dimer in this study and from previous study were used for pKa predictions to get the average value and error bars. All structure images were created using open source PyMOL.

Results

Isolated H2A-H2B dimer retains the same core structure as in the nucleosome

Previously, the solution structure of the human H2A-H2B dimer was solved by the Nishimura lab based on experimental backbone chemical shifts and the protein folding program ROSETTA, showing considerable variation in the position of H2B α C helix¹⁸. Since this helix carries two of the acidic residues that contribute to the acidic patch, a potentially dynamic position of the H2B α C helix might affect long-range electrostatic interactions and thus change the electrostatic potential of the acidic patch within the dimer compared to the nucleosome. We thus first wanted to verify that the acidic patch in our experimental system, the fruit fly H2A-H2B heterodimer, retains the same structure as in the nucleosome. Chemical shift indices (CSI) obtained from the experimental C α and C β chemical shifts confirm the presence of the histone fold core secondary structure elements for both histones (**Figure 5.2a**). Similar to the human dimer, the N-terminal α -helix and C-terminal region of H2A are unfolded in solution, while these are defined structural elements in the crystal structure of the nucleosome. Overall, these data confirm Nishimura's observation that the H2A-H2B dimer forms a well-folded histone fold domain with disordered tails in solution.

To obtain the tertiary structure of the histone fold core of the dimer that includes the acidic patch, we used a Rosetta calculation supplemented with sparse distance restraints and backbone chemical shifts. As demonstrated in **Chapter 3** for the H2A.B-H2B variant dimer, addition of distance restraints can be crucial to define the structure of the protein backbone. Indeed, a calculation of the structure purely based on HN, N, C α , C β , C' backbone chemical shifts resulted in an ill-defined histone fold for the 10 models with the lowest Rosetta energy (**Figure S5.1**). Sparse intermolecular distance restraints were extracted from a dedicated NOESY experiment designed to measure intermolecular NOEs between H2B amide protons and H2A backbone or side chain protons, using dimers refolded from unlabeled H2A and perdeuterated and ¹⁵N-labeled H2B. For nearly all H2B residues in the core (Y34-K122), including the H2B α C helix, several intermolecular NOEs were observed indicating that the dimers were well-folded (**Figure 5.2b**). Of these, 47 NOEs could be unambiguously assigned using the automated assignment procedure of CYANA¹⁹, taking the nucleosome crystal

structure (PDB id: 2PYO²⁰) as the structural reference (see Material and Methods for details) (**Figure 5.2b,c**). Among these NOEs are 7 from the H2B α 1 helix, 16 from H2B α 2 helix, 14 from H2B α C helix, 7 from H2B loop 1, and 3 from H2B loop2. The dispersion over the core region of H2B suggests that the dimer in solution indeed adopts the same structure as in the nucleosome. Inclusion of these distance restraints in the structure calculation resulted in a well-defined structural ensemble with 1.63 ± 0.29 Å RMSD to the nucleosome dimer structure. (**Figure 5.2d**, see Supplementary **Table S5.1** for Structural statistics).

Acidic patch residues H2A E55 and E60 have high predicted pKa's

Having shown that the H2A-H2B dimer retains the nucleosomal structure in solution, we next submitted the coordinates of the final structural ensemble to the PDB2PQR webserver to predict the side chain pKa values using the propka protocol. As a control, the predicted pKa from our structure were compared to those predicted from the human dimer structure and those predicted directly from the nucleosomal structure (**Figure 5.3a**). Overall, the three different structures give rise to very similar predicted pKa values. In particular, there is little difference between the values predicted for the fruit fly dimer structure determined here and the less converged human dimer structure. Notably, there are three residues with significantly different predicted pKa in the dimer and nucleosome context. Two of these are in close proximity to the nucleosomal DNA (H2A H81 and H2B D48) which increases their buriedness in the nucleosome and exposes these sidechains to the negative electrical field from the DNA. While the elevated pKa for H2B D48 can be understood in these terms, the reason for the reduced pKa predicted for H2A H81 is unclear. Finally, the predicted pKa for H2A E55 is ~1 unit higher in the nucleosome than in the dimer. Within the nucleosome, the E55 side chain is partly buried in a largely hydrophobic pocket by the H2A the α N helix, the loop to the α 1 helix and the H2B α 1 helix. The sidechain position is stabilized in this position through hydrogen-bonds to the backbone amide of H2A Q23 and F24 that are part of the loop between the α N and α 1 helices. Since the α N helix is not folded in the free dimer, the E55 side chain will be more exposed in the free dimer. We conclude that with exception of H2A E55, H81, and H2B D48, reliable estimates for the

pKa's of residues in and around the acidic patch can be made within the dimer context.

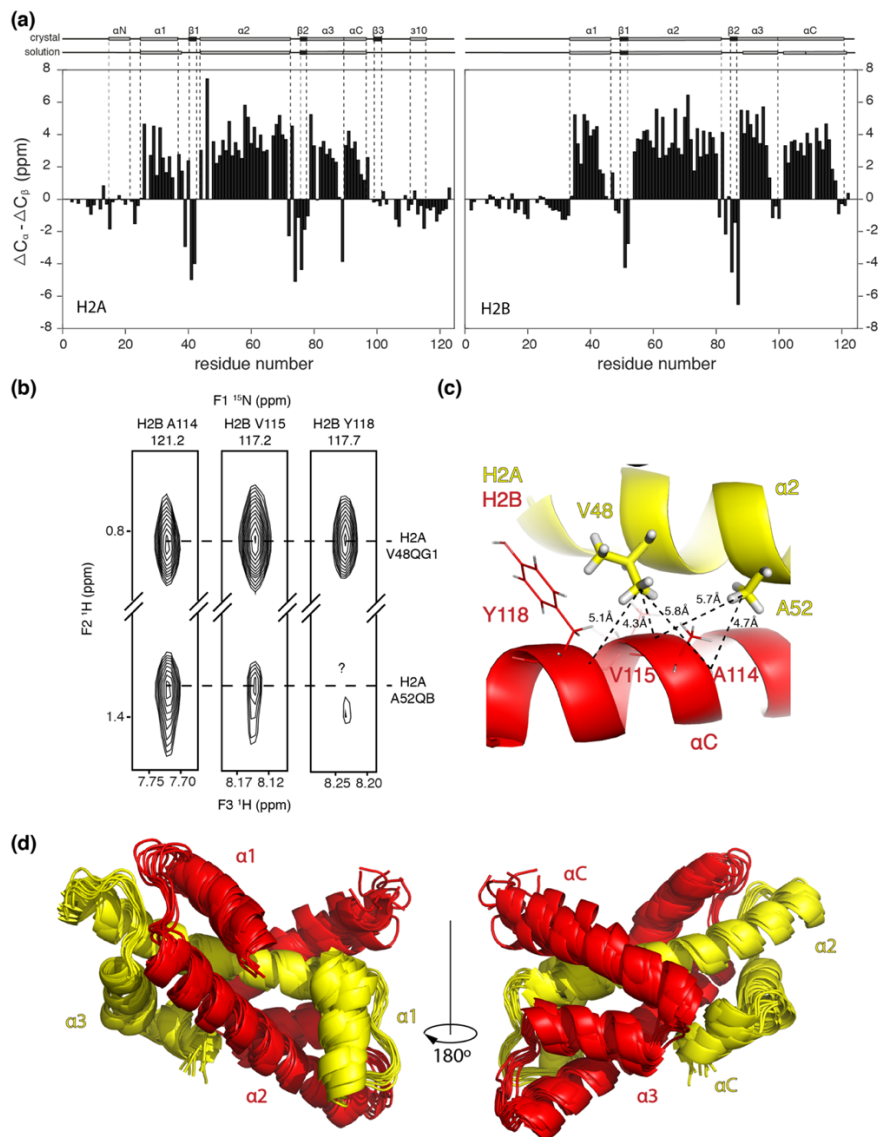


Figure 5.2, isolated H2A-H2B dimer in solution maintains canonical folding as in nucleosomes. (a) secondary structural elements of H2A and H2B in the heterodimer based on experimental chemical shift $C\alpha$ and $C\beta$. The secondary structure observed in nucleosome crystal structure (PDB id: 2PYO) and previously reported heterodimer solution structure (PDB id: 2RVQ) are indicated as bars (alpha-helices) and arrows (beta-sheets) above the plot. (b) Strips from the 3D ^{15}N -

edited NOESY recorded on heterodimers with perdeuterated ^{15}N -labeled H2B and unlabeled H2A showing intermolecular NOEs between H2B amide and H2A protons. Assignment of H2B amide and H2A resonances are indicated. NOEs shown restrain the H2B αC -helix position. (c) Zoom on the H2A-H2B structure (PDB id: 2PYO) corresponding to the intermolecular contact shown in the NOESY spectrum. (d) Superposition of best 10 Rosetta models with lowest rosetta energy and $\text{C}\alpha$ RMSD for H2A-H2B solution structure calculated by Rosetta using backbone chemical shift HN, N, $\text{C}\alpha$, $\text{C}\beta$, C' and intermolecular NOE distance restraints.

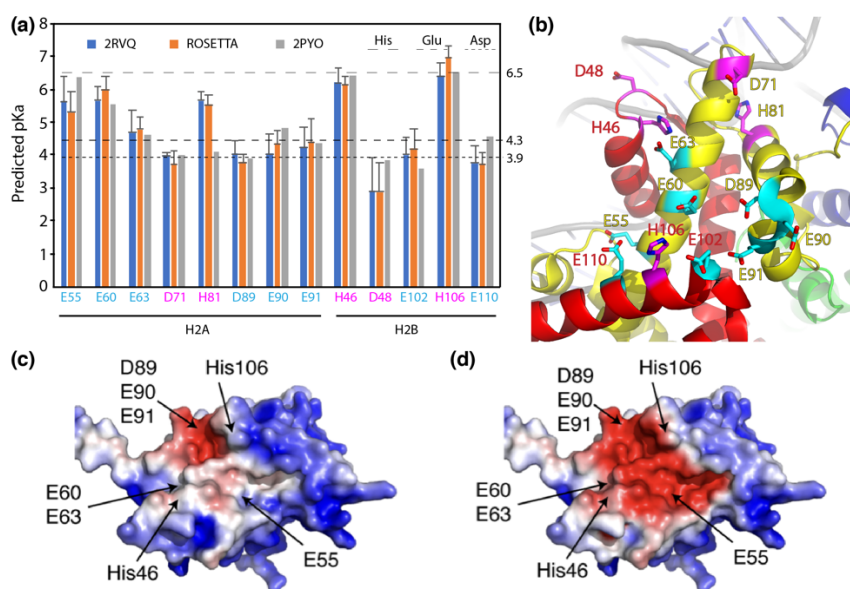


Figure 5.3. Predicted pKa for acidic patch residues represent well with corresponding ones in nucleosome. (a) the predicted pKa values for residues of acidic patch (cyan in (b)) and residues around the nucleosome surface with neutral to negative charges (magenta in (b)) are compared between Nishimura's non-converged solution structure (PDB id: 2RVQ), our converged solution structure calculated by rosetta, and nucleosome crystal structure (PDB id: 2PYO). The 10 models for solution structural ensembles were used for predictions individually and the standard deviations are plotted as error bars. The pKa values for free state Histidine, Aspartic acid and Glutamic acid side chains are indicated as dash lines. (b) the residues from plot (a) are highlighted in the nucleosome crystal structure (PDB id: 2PYO), shown as sticks. Nucleosome color coding as H2A is shown in yellow, H2B in red, H3 in blue, H4 in green, and DNA in grey. (c, d) Electrostatic surface potential plot of acidic patch in the cases when H2A E55/E60 are neutral and H2B H46/H106 are positively charged (c) and in the case when H2A E55/E60 are negatively charged and H2B H46/H106 are neutral (d). Electrostatic surface potentials were calculated for dimer extracted from PDB id 2PYO using the APBS

plugin in PyMol with the non-linear Poisson-Boltzmann equation contoured at 4 kT/e. Negatively and positively charged surface areas are colored in red and blue, respectively.

Interestingly, three residues have significantly different predicted pKa's compared to their random-coil values. Of these, two residues, H2A E55 and E60, are part of the previously established nucleosome-protein interaction surface in the acidic patch^{2,21}(see also **Figure 5.1c**). For both, significantly elevated pKa's are predicted (5.3 for E55 and 6.0 for E60). For E60, this may be due to stabilization of the protonated state by a hydrogen bonding interaction with H2A Y56, whereas for E55 this may be due to the partly buried position of this side chain also within the dimer. Within the nucleosome, the pKa of H2A E55 is predicted to even reach 6.4. Together with two histidine's in H2B (H46 and H106) that line the acidic patch, there are thus four residues in and around the acidic patch that may change protonation state depending on buffer pH for the typical pH range used *in vitro*. To illustrate the impact of such changes, we calculated the predicted electrostatic surface potential both for the “worst” case scenario in which all four residues H2A E55, E60 and H2B H46, H106 are all protonated (**Figure 5.3c**), and the “best” case scenario in which these residues are all deprotonated (**Figure 5.3d**). The clear difference in negative electrostatic potential could contribute to loss of binding for proteins that target the acidic patch.

Large changes in the acidic patch chemical environment between pH 6 and 7

We first assessed experimentally the impact of changes in pH on the heterodimer using amide backbone chemical shift perturbation mapping. Using refolded H2A-H2B dimers with uniformly ¹⁵N-labeled H2A and unlabeled H2B or vice versa, 2D NH HSQC spectra were recorded under different pH conditions with controlled ionic strength (200 mM). The buffer pH was varied between 4.5 and 9.1 (see **Table S5.2**; using acetate buffers for pH below 5.5 and MES buffers for pH between 5.5 and 6.5, and phosphate buffers for pH above 6.5, and CHES buffer at pH 9). Throughout this pH range well-dispersed spectra were obtained, indicating that the heterodimer remains folded which is in accordance with literature data²². Assignments of both H2A and H2B amide resonances were obtained earlier at pH 6.5^{7, 12}

and transferred here to all spectra (see also M&M). Resonances from the disordered histone tails were not visible above pH 8 due to fast exchange of backbone amide proton with the water resonance. Several resonances experience large chemical shift perturbations (CSPs) over the 4.5-9 pH range, including H2A Q23 and E63 as well as H2B H46 and E102. Residues with the largest CSPs cluster in the acidic patch region, also when focusing only on the changes between pH 6 and 8 (**Figure 5.4a**).

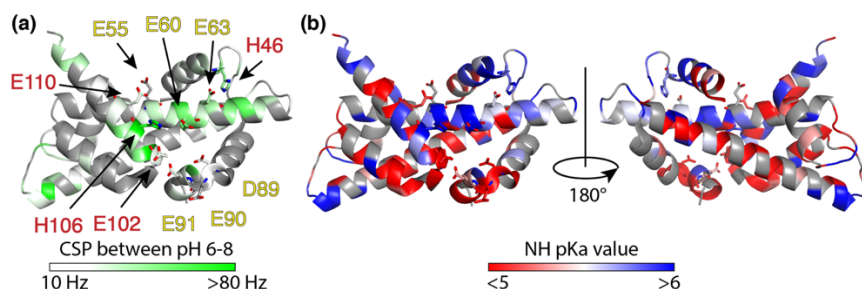


Figure 5.4. Apparent pKa of H2A-H2B dimer core region measured by chemical shift perturbations of backbone amide protons. (a) plot CSP between pH 6 to 8 on crystal structure of H2A-H2B dimer core region (PDB id: 2PYO). Combined CSPs are indicated by color gradient from white (10 Hz) to green (above 80 Hz). Residues that were too overlapped or weak, or with combined CSP less than 10 Hz are colored grey. Acidic patch residues and histidines are indicated by arrows pointing to H2A residues (yellow) and H2B residues (red). (b) plot of apparent pKa measured by backbone amide proton on crystal structure of H2A-H2B dimer core region (PDB id: 2PYO). Color coded red to white to blue, represents pKa <5 (red), 5.5 (white), >6 (blue). Residues that were too overlapped or weak are colored grey. H2A acidic patch residues and H2B acidic patch residues together with H46 and H106 are shown as sticks.

As the amide backbone chemical shifts are not only sensitive to electrostatics, but even more so to changes in hydrogen bonding strength and local conformation, the observed CSPs can only be taken as indirect reporters of changes in protonation state. Moreover, sensitivity to multiple (nearby) protonation events can result in curved peak trajectories, that can complicate the analysis. Fitting of the CSP-derived pH titration profiles to a modified version of the Henderson-Hasselbalch equation thus results in a single, residue-specific apparent pKa, even for residues without titratable group and even in the case where multiple protonation events are sensed (see **Table S5.3**). For

example, H2A residues A65 displays a clear and linear pH –dependent CSP that can be fitted to an apparent pKa value of 5.75 which may be due to protonation of the nearby H2B H46 side chain at low pH (**Figure 5.5a**). An extremely curved trajectory was observed for H2A E63 that is near to both H2A E63 and H2B H46 side chains, and thus may be sensitive to two protonation events (**Figure 5.5d**).

In cases where the amide proton is hydrogen bonded to a titratable group, the apparent pKa can be taken to reflect directly pKa of the side chain involved. For example, one of the largest CSPs is observed for H2B E102 amide which can be explained by a change in hydrogen bonding to the nearby carboxyl group of H2A E91 upon lowering the pH below ~5.3 (**Figure 5.5b**). The resulting best-fit apparent pKa of 4.28 thus likely represents the pKa of the H2A E91 carboxyl group. Similarly, the fitted value obtained for the H2B L103 amide likely reflects the E90 pKa.

Interestingly, the pKa of H2A E55 can be estimated to be 4.5, based on the apparent pKa's for the H2A Q23 and F24 amides (**Figure 5.5c**). As noted above, these amides are hydrogen bonded to the E55 carboxyl group in the nucleosome. While the preceding α N helix is not formed within the free dimer and the loop containing these amides is likely more flexible in the dimer, the strongly downfield shifted $^1\text{H}_\text{N}$ chemical shifts suggest these remain hydrogen bonded in the dimer. Furthermore, H2A E55 is the only titratable group in close proximity to these amides. This estimate for the E55 pKa value is considerably lower than the structure-based predicted value of 5.3 and more in line with the expected values for glutamate side chains.

The last acidic patch residue for which the backbone data can be used to derive the sidechain pKa is H2A D89. Its sidechain is hydrogen bonded to the E91 and L92 backbone amides. Surprisingly, these report an apparent pKa of 5.6 (**Fig 5e**, significantly higher than the random coil (3.8) or structure-predicted pKa values (3.9).

When plotted on the structure, inspection of the residue-specific pKa shows that most residues in and around the acidic patch, in particular near the H2A D89, E90, E91 triad, report apparent pKa values below 5 (**Figure 5.4b**). The opposite end of the acidic patch, near H2A E55 and H2B E102, as well as the center around H2A E60 and E63 have apparent pKa higher than 5. These are relatively close to histidine H106 and H46 from H2B, that line the acidic patch. These observations,

coupled with the large CSPs observed for the acidic patch, suggest that the effective electrostatics of the acidic patch may indeed change in the pH range of 6-8. This change is likely due to changes in protonation state of the H2B histidines H106 and H46, and H2A D89.

Direct pKa measurement of Glu and Asp by side chain carboxyl carbon chemical shifts

To verify whether the acidic patch glutamates or aspartates have indeed low pKa as suggested by the analysis above, we next wanted to observe the chemical shifts of the carboxyl carbons directly as they are more unambiguous and sensitive reporters of the protonation state. These ^{13}C chemical shifts typically move ~ 4 ppm upfield as the carboxyl group undergoes protonation. Using dimer (25 kDa) with fractionally deuterated and ^{13}C labeled H2A, 2D CH correlation spectra were recorded using a HCCO pulse sequence where each correlation connects the $^{13}\text{C}\gamma/\text{C}\delta$ carboxyl groups to the $\text{H}\beta/\text{H}\gamma$ of aspartate or glutamate ²³. Here only H2A was isotope labeled, since H2A contains most of the relevant acidic residues, including E60 which may have elevated pKa according to the structure based pKa prediction and backbone based pH titration. Histone H2A contains 7 glutamate and 2 aspartate acids together with 7 asparagines and 4 glutamines, which in principle should generate 40 observable signals in these spectra. Signals from Asn/Asp and Gln/Glu can be discriminated based on distinct chemical shifts for the carboxyl carbon (176.4/177.8 ppm for $\text{C}\gamma$ of Asn/Asp, 179.3/181.2 ppm for $\text{C}\delta$ of Gln/Glu) and for proton $\text{H}\beta/\text{H}\gamma$ (~ 2.7 ppm for Asn/Asp, ~ 2.2 ppm for Gln/Glu). The 2D HCO spectrum shows several peaks in the glutamate region, which are unfortunately heavily overlapped (**Figure 5.6c**). The Gln and Asn residues are also visible from the 2D HCO spectra and well separated from Glu residues. No CSP were observed during the titration for Gln and Asn residues (**Figure 5.6b**). Unfortunately, signals for the 2 Asp residues could not be detected, likely due to lack of more extensive line broadening for these shorter sidechains ²³. To assign the glutamate $\text{C}\delta/\text{H}\gamma$ resonance, we attempted to correlate the $\text{H}\gamma$ resonances to the assigned backbone resonances via both triple resonance backbone and sidechain resonances. However, neither strategy was successful due to poor sensitivity. The corresponding region in the CH constant time HSQC spectrum shows 7 resonance pairs that likely correspond to the

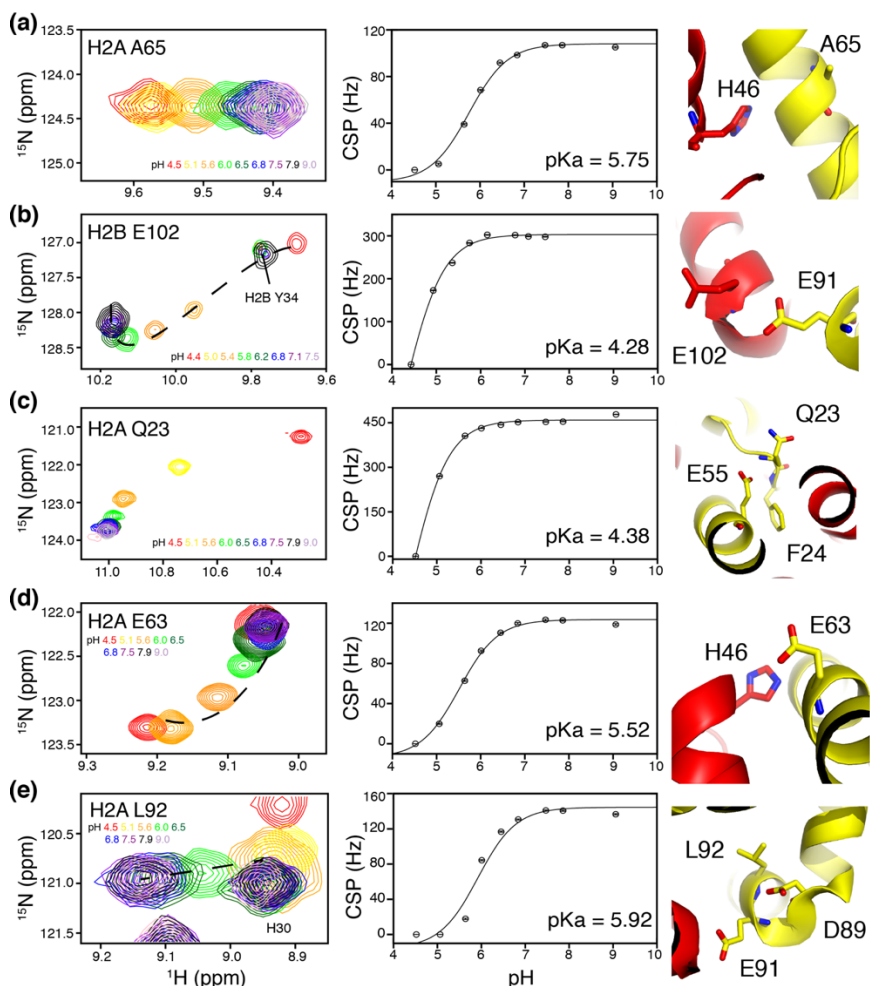


Figure 5.5. Apparent pKa of H2A-H2B dimer core region measured by chemical shift perturbations of backbone amide protons. (a-e) CSP of residues are shown on the left column, titration curves fit combined CSP to pH and the apparent pKa determined are shown in the middle column, and the structural references using the crystal structure of H2A-H2B dimer (PDB id: 2PYO) are shown on the right column.

seven Glu residues in H2A (**Figure 5.6a**). Due to the high dynamics of the C-terminal H2A tail, we could tentatively assign C-terminal E120 to the most intense pair of signals. By comparing the CH constant time HSQC spectrum with the 2D HCO spectrum at similar pH conditions (5.8 and 6.0 respectively), we could arbitrarily assign the other six

glutamates to the overlapped peaks (**Figure 5.6c**). By following these arbitrarily assigned glutamate peaks in the 2D HCO spectrum at titrated pH conditions, valuable information of the pKa values of the observed resonances was extracted (**Figure 5.6d**). During titration, all peaks shift more or less together, indicating the corresponding glutamate side chains have similar pKa values. All glutamates side chain carboxyl group pKa are below 5 (**Figure 5.6e**). Since the dimer does not fold stably in conditions with pH lower than 4.5, no data could be collected to reach the fully protonated state. Nevertheless, the limited CSPs (~1.1 ppm) and approximate fit to the Henderson-Hasselbalch equation clearly indicate their pKa value is lower than 5. These findings thus indicate that the acidic residues in the acidic patch have pKa's close to their default values and thus point to the histidines as being responsible for the large CSPs in the acidic patch between pH 6 and 8.

pKa values of H2B H106 and H46

To experimentally determine the pKa values of the histidines that line the acidic patch, we used heterodimers with ^{15}N , ^{13}C -labeled H2B and unlabeled H2A and recorded HMQC spectra probing long-range correlation between the sidechain $^1\text{H}\epsilon 1$, $^1\text{H}\delta 2$, $^{15}\text{N}\epsilon 2$ and $^{15}\text{N}\delta 1$ spins in the histidine side chains. Since H2B contains both H46 and H106 that line the acidic patch, as well as an additional histidine H79 close to the dimer-dimer interface, a titration experiment on this single sample suffices to obtain the relevant pKa's. The spectra indeed show correlations for each of the three histidines in H2B (**Figure S5.2**). Titration of the pH between 4.4 and 7.5, resulted in clear chemical shift changes (**Figure 5.7a**) that could be fitted to pKa's of 5.9, 6.5 and 6.7 (**Figure 5.7b**). Although side chain assignment of the histidines is currently lacking, we could assign the three histidines based on comparison of the side chain determined pKa data with the backbone amide proton determined pKa's. Histidine number 3 has the largest chemical shift perturbation in the proton dimension and its pKa is determined to be below 6 (**Figure 5.7a,b**). In the dimer structure, H2B H46 side chain is close to several H2A residue backbone amide protons whose pKa values were determined to be below 6 (H2A E63-N67, see **Table S5.2**), while residues around H2B H79 have backbone amide proton pKa value above 6.5 (H2A A39 6.7, H2B R76 6.8, Y80 6.5).

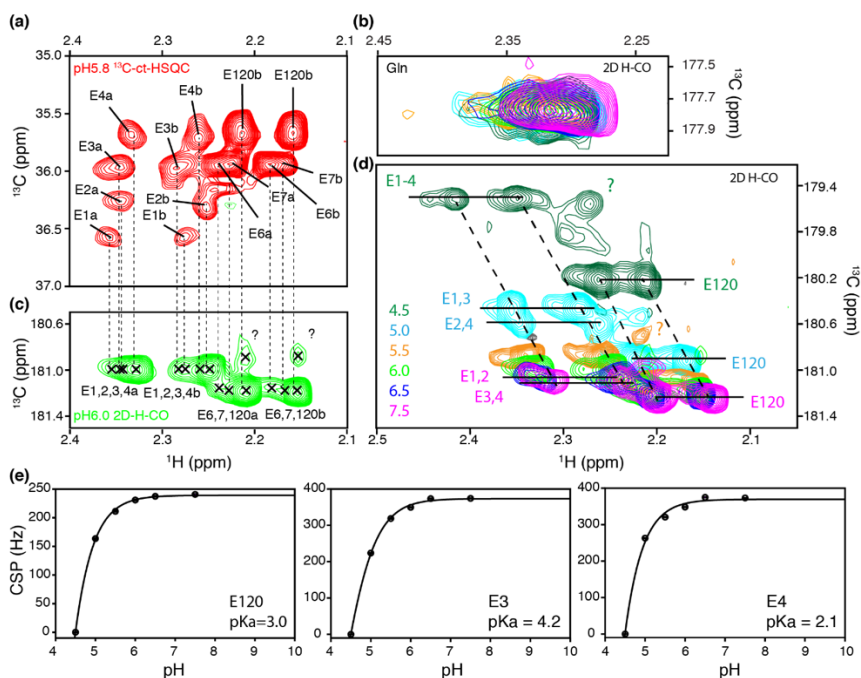


Figure 5.6. glutamates in H2A has pKa below 5 for side chain carboxyl group in dimer. (a) ^{13}C -H constant time HSQC spectrum of H2A glutamates show well resolved peaks for 7 glutamates in sequence. Each glutamate has a duplicated peak for the side chain carboxyl group. (b, c, d) H-CO 2D spectra measured using HCACO sequence centered on side chain. Glutamines (b) has different chemical shifts than glutamates (c, d) on 2D H-CO spectra. (c) Glutamates side chain carboxyl group heavily overlap with each other. Dash lines showing the assignment of 7 glutamates based on ^{13}C -H constant time HSQC spectrum. Question marked peaks are not traceable at other pH conditions. (d) CSPs of H2A glutamates shown by 2D H-CO spectra overlay at different pH conditions. During pH titration, obvious shift started to occur when pH drops to 5. Peak positions are marked at representative pH conditions. Arbitrarily assigned E6 and E7 shifts closely represent the pattern of E120. (e) titration curves fit CSP to pH are shown for H2A E120, arbitrarily assigned E3 and E4.

For histidine number 2, we have observed a change in shifting direction of 2N ϵ 2-H δ 2 and 2N δ 1-H δ 2 side chain group when the pH drops below 5 (**Figure 5.7e**). Similarly, changes in shifting direction of backbone amide proton for H2B A107 and L103 were observed (**Figure 5.7f,g**). Since H2B A107 and L103 is close to H106 in the dimer structure, we can confirm the assignment for histidine number 2 to H106. Thus, we assign the measured pKa for number 1 to H79,

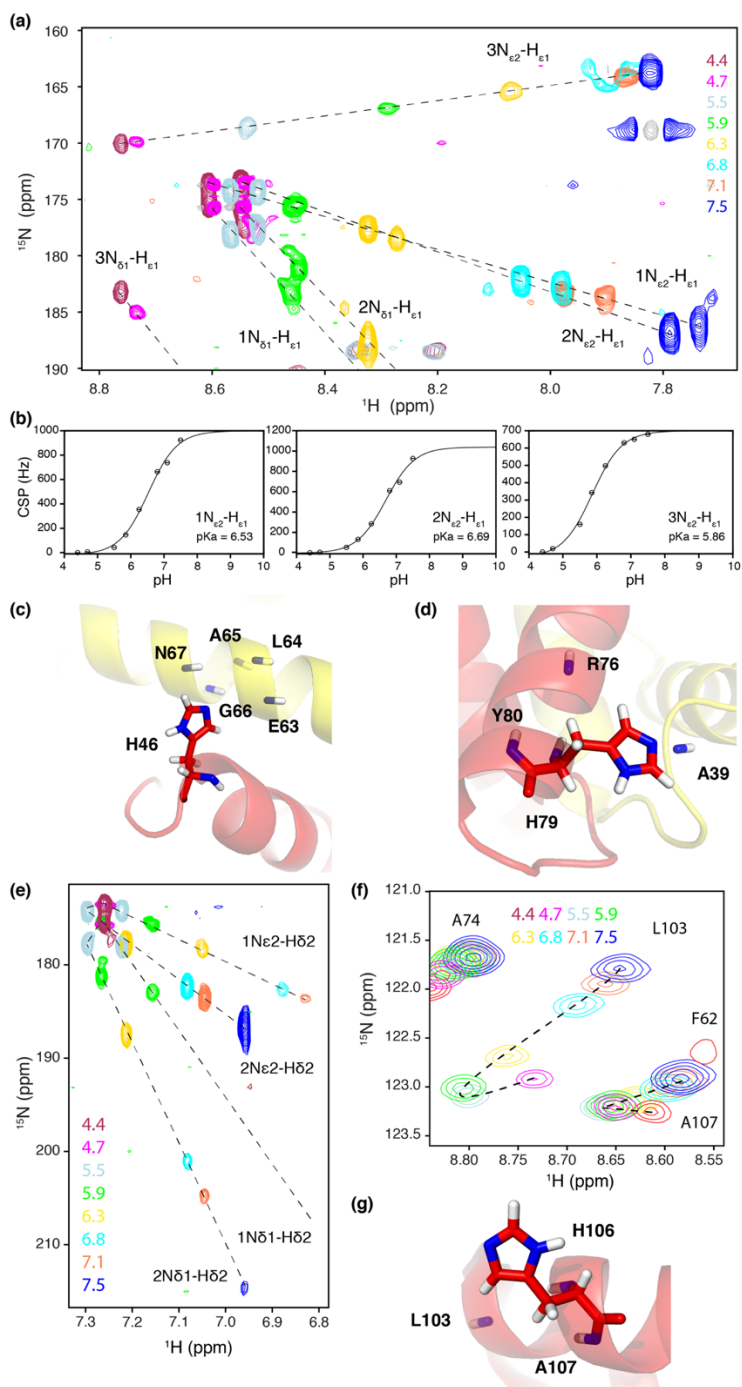


Figure 5.7. pKa value determined by histidine side chain. (a) CSP of 3 histidines side chain amide groups from H2B at different pH conditions. Peak shifting directions are marked as dashed lines. (b) titration curves fit CSP to pH are shown for each histidine. (c) Backbone amide protons from H2A E63-N67 close to H2B H46 side chain amide group report apparent pKa's below 6 (see table S5.2). (d) backbone amide groups from H2B R76 and Y80 and H2A A39 close to H2B H79 side chain amide group report apparent pKa's above 6.5 (see table S5.2). (e) change of CSP direction at pH below 5 of histidine's side chain amide groups of 2Nε2-Hδ2 and 2Nδ1-Hδ2. (f) change of CSP direction at pH below 5 of backbone amid groups of H2B L103 and A107. (g) H2B L103, H106 and A107 are close to each other in dimer structure (PDB id: 2PYO).

histidine number 2 to H2B H106, and histidine number 3 to H2B H46. To conclude, we have measured histidine side chain pKa values of H2B histidines as 5.9 for H46, 6.5 for H79 and 6.7 for H106.

Discussion

The nucleosome acidic patch is a binding surface for many nucleosome remodeling factors. The electrostatic potential of the surface is essential for binding of specific partners, and moderate changes may cause binding deficiency. For chromatin factor HMGN2 it was found that a drop in buffer pH from 7 to 6 results in a drastic decrease in binding affinity to nucleosome (unpublished data from Hugo van Ingen, Hidenori Kato, Lewis E. Kay, Yawen Bai). Such peculiar pH dependent binding affinity may directly be caused by changes in the electrostatic potential of nucleosome surface. Here, we sought to experimentally determine the pKa values of all titratable groups in and around the acidic patch within the free H2A-H2B dimer.

We first determined the H2A-H2B solution structure using a CS-ROSETTA protocol supplemented with sparse intermolecular distance restraints. The resulting structure is well defined and close to nucleosomal dimer structure. Notably, this structure does not show evidence for an intrinsic conformational heterogeneity present in the dimer as was argued in a previous study using CS-ROSETTA only¹⁸. Determination of all side chain pKa was complicated by the substantial size of the H2A-H2B dimer (25 kDa). While residue specific mapping of sidechain pKa has been demonstrated before for smaller protein domains and intrinsically disordered proteins, the enhanced relaxation in this system made it challenging to directly probe the protonation of

all sidechain carboxyl groups in an unambiguous manner. Yet, combination of these data with apparent pKa derived from backbone amide data and experimental pKa's for the histidine sidechains allows to assign pKa values or estimated values to all titratable groups in and around the acidic patch (**Table 5.1**). Our results show that all glutamates residues in the acidic patch have pKa below 5, and thus unambiguously falsify the high predicted pKa for H2A E55 and E60. Surprisingly, for the sole aspartate in the acidic patch, H2A D89, we find a best estimate of 5.6 for its pKa. This estimate is based on the backbone data and could unfortunately not be verified directly in sidechain experiments due to too low a sensitivity. For two histidines that line the acidic patch, H106 and H46, pKa values of 6.7 (H106) and 5.8 (H46) were determined, respectively. At physiological pH, the acidic patch is thus highly negatively charged. We found that reduction of the pH below the physiological value results in large CSPs in and around the acidic patch. Our data indicate that this is first due to the protonation of H2B H106, then due to protonation of H2B H46 and H2A D98. Protonation of these residues likely alters their sidechain position, which in particular for the aromatic histidine rings will result in large chemical shift changes for surrounding backbone amides.

In situ pH measurements have indicated that the nuclear pH is 7.8, slightly higher than the cytosolic pH (7.4) ²⁴. At pH of 7.8, the histidine H106 and H46 are for 7.4% and 1.0% in their protonated, positively charged form, while D89 is <0.6% in the neutral state. At pH 7.4 this increases to 16.6%, 2.5% and 1.5%, respectively. The reduced pKa for H46 may help to retain a pronounced acidic character near zone I of the acidic patch (H2A E60, L64, D89, E91, and H2B E102, L103) which is the key interaction site of nucleosome-binding proteins ^{2, 21}. Such slight changes in protonation levels are unlikely to result in significant perturbation of binding of chromatin factors to the acidic patch. In addition, burial of these residues upon protein binding may reduce their pKa's by stabilizing the deprotonated state. Yet, it is conceivable that changes in the local microenvironment could result in more pronounced acidification that would disfavor binding.

Table 5.1. Predicted pKa and experimentally measured pKa values for all titratable groups in H2A-H2B dimer.

residue	Predicted pKa (dimer / nucleosome)	Best estimated from backbone N-H apparent pKa	Best estimated from side chain pKa	Final best estimated pKa
H2A H30	5.68/1.28	4.36		
H2A E40	3.82/4.84	5.91	<5	<5
H2A E55 (AP)	5.30/6.37	4.53 (hb to Q23, F24)	<5	4.53
H2A E60 (AP)	5.99/5.56	6.15	<5	<5
H2A E63 (AP)	4.78/4.62	5.52	<5	<5
H2A D71	3.73/4.02	4.58		
H2A H81	5.53/4.10	N/A (hb to I78)		
H2A D89 (AP)	3.79/3.88	5.59 (hb to E91, L92)		5.59
H2A E90 (AP)	4.33/4.82	4.24	<5	<5
H2A E91 (AP)	4.40/4.37	4.35 (hb to H2B E102, L103)	<5	<5
H2A E120	-/5.08	4.18	<5	4.18
H2A A123 (C-termini)	-/3.24	4.08		4.08
H2B E32	-/3.87	6.10		
H2B H46	6.14/6.43	5.68 (close to H2A G66, N67)	5.86	5.77
H2B D48	2.88/3.85	5.83		
H2B D65	4.42/2.85	-		
H2B E68	4.50/6.28	-		

H2B E73	4.25/3.78	-		
H2B H79	6.55/5.23	6.91	6.53	6.53
H2B E90	3.32/4.02	3.77 (hb to T87)		3.77
H2B E102 (AP)	4.20/3.59	4.50		
H2B H106	6.96/6.53	6.61	6.69	6.69
H2B E110 (AP)	3.72/4.55	5.14		
H2B K122 (C-terminus)	2.43/3.07	4.25		4.25

Comparison of the experimental and predicted pKa values highlights the challenge in accurate prediction of pKa values and the need for experimental validation (**Table 5.1**). The predictions indicate that for residues close to the nucleosomal DNA, e.g. D48 or H81, the pKa measured may be altered in the nucleosome. This may also be the case for H2A E55, for which a higher pKa values was predicted in the nucleosome context. Assuming that the experimental value in the nucleosome is higher by one unit, as in the predictions, this would result in a pKa of ~5.5 for E55 in the nucleosome. This residue would thus be the fourth residue to titrate upon dropping the pH below physiological value, after H106, H46 and D98.

In conclusion, we have determined the solution structure of the H2A-H2B dimer and studied the pH titration of the acidic patch residues. We find that the acidic patch is indeed highly acidic at physiological pH and remains so down to pH ~6.7 at which point H2B H106 starts to titrate. Further reduction in pH will cause protonation of both H2B H46 and H2A D98 in that will significantly decrease the electronegativity near the key protein interaction site of the acidic patch. In addition, H2A E55 may also get significantly protonated around pH 5.5 when in the nucleosome context. These results are important for proper planning of *in vitro* experiments as well as for accurate molecular dynamics simulations of nucleosome-nucleosome or nucleosome-proteins interactions that are mediated through the acidic patch.

References

1. Speranzini, V.; Pilotto, S.; Sixma, T. K.; Mattevi, A., Touch, act and go: landing and operating on nucleosomes. *EMBO J* **2016**, 35 (4), 376-88.
2. McGinty, R. K.; Tan, S., Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol* **2016**, 37, 54-61.
3. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, 389 (6648), 251-60.
4. Chodaparambil, J. V.; Barbera, A. J.; Lu, X.; Kaye, K. M.; Hansen, J. C.; Luger, K., A charged and contoured surface on the nucleosome regulates chromatin compaction. *Nat Struct Mol Biol* **2007**, 14 (11), 1105-7.
5. Pahari, S.; Sun, L.; Alexov, E., PKAD: a database of experimentally measured pKa values of ionizable groups in proteins. *Database (Oxford)* **2019**, 2019.
6. Bombarda, E.; Ullmann, G. M., pH-dependent pKa values in proteins--a theoretical analysis of protonation energies with practical consequences for enzymatic reactions. *J Phys Chem B* **2010**, 114 (5), 1994-2003.
7. Horn, V.; Uckelmann, M.; Zhang, H.; Eerland, J.; Aarsman, I.; le Paige, U. B.; Davidovich, C.; Sixma, T. K.; van Ingen, H., Structural basis of specific H2A K13/K15 ubiquitination by RNF168. *Nat Commun* **2019**, 10 (1), 1751.
8. Oda, Y.; Yamazaki, T.; Nagayama, K.; Kanaya, S.; Kuroda, Y.; Nakamura, H., Individual ionization constants of all the carboxyl groups in ribonuclease HI from *Escherichia coli* determined by NMR. *Biochemistry* **1994**, 33 (17), 5275-84.
9. Platzner, G.; Okon, M.; McIntosh, L. P., pH-dependent random coil (1)H, (13)C, and (15)N chemical shifts of the ionizable amino acids: a guide for protein pK_a measurements. *J Biomol NMR* **2014**, 60 (2-3), 109-29.
10. Velyvis, A.; Kay, L. E., Measurement of active site ionization equilibria in the 670 kDa proteasome core particle using methyl-TROSY NMR. *J Am Chem Soc* **2013**, 135 (25), 9259-62.
11. Luger, K.; Rechsteiner, T. J.; Richmond, T. J., Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* **1999**, 304, 3-19.
12. Corbeski, I.; Dolinar, K.; Wienk, H.; Boelens, R.; van Ingen, H., DNA repair factor APLF acts as a H2A-H2B histone chaperone through binding its DNA interaction surface. *Nucleic Acids Res* **2018**, 46 (14), 7138-7152.

13. Shen, Y.; Bax, A., Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR* **2013**, *56* (3), 227-241.
14. Pelton, J. G.; Torchia, D. A.; Meadow, N. D.; Roseman, S., Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci* **1993**, *2* (4), 543-58.
15. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325-7.
16. Shen, Y.; Bryan, P. N.; He, Y.; Orban, J.; Baker, D.; Bax, A., De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci* **2010**, *19* (2), 349-56.
17. Shen, Y.; Vernon, R.; Baker, D.; Bax, A., De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* **2009**, *43* (2), 63-78.
18. Moriwaki, Y.; Yamane, T.; Ohtomo, H.; Ikeguchi, M.; Kurita, J.; Sato, M.; Nagadoi, A.; Shimojo, H.; Nishimura, Y., Solution structure of the isolated histone H2A-H2B heterodimer. *Sci Rep* **2016**, *6*, 24999.
19. Guntert, P.; Buchner, L., Combined automated NOE assignment and structure calculation with CYANA. *J Biomol NMR* **2015**, *62* (4), 453-71.
20. Clapier, C. R.; Chakravarthy, S.; Petosa, C.; Fernandez-Tornero, C.; Luger, K.; Muller, C. W., Structure of the *Drosophila* nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **2008**, *71* (1), 1-7.
21. Fang, Q.; Chen, P.; Wang, M.; Fang, J.; Yang, N.; Li, G.; Xu, R. M., Human cytomegalovirus IE1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome. *Elife* **2016**, *5*.
22. Karantza, V.; Baxevanis, A. D.; Freire, E.; Moudrianakis, E. N., Thermodynamic studies of the core histones: ionic strength and pH dependence of H2A-H2B dimer stability. *Biochemistry* **1995**, *34* (17), 5988-96.
23. Everill, P.; Sudmeier, J. L.; Bachovchin, W. W., Direct NMR observation and pKa determination of the Asp102 side chain in a serine protease. *J Am Chem Soc* **2012**, *134* (4), 2348-54.
24. Seksek, O.; Bolard, J., Nuclear pH gradient in mammalian cells revealed by laser microspectrofluorimetry. *J Cell Sci* **1996**, *109* (Pt 1), 257-62.

Supplements

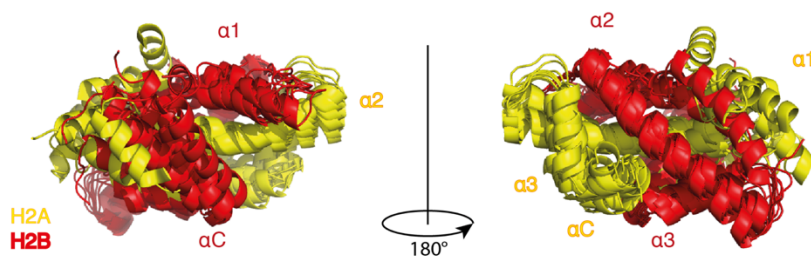


Figure S5.1. Superposition of the 10 CS-Rosetta models with the lowest Rosetta energy to align the H2A and H2B $\alpha 2$ helices, showing ill-defined positions of H2A $\alpha 1$ and H2B αC helices within the dimer structure.

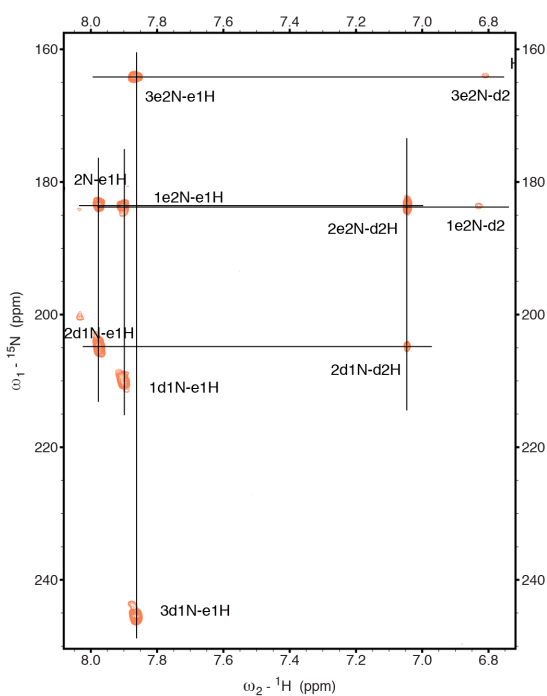


Figure S5.2. Side chain H-N spectra for three histidine's in H2B at pH 7.1.

Table S5.1. Structural statistics for the core region (H2A V26-S97 and H2B Y34-K122) of isolated H2A-H2B heterodimer

A. Restraint information	
number of intermolecular NOEs	47
B. Average RMS deviation from experimental restraints	
All experimental distance restraints (Å)	0.3 ± 0.5
C. Coordinate RMS deviation ^a (Å)	
Average RMSD to mean	
Ordered backbone atoms	1.1 ± 0.2
Ordered heavy atoms	1.5 ± 0.2
Global backbone atoms	1.2 ± 0.2
Global all heavy atoms	1.6 ± 0.2
Average RMSD to 2PYO	
Ordered backbone atoms	1.6 ± 0.3
Ordered heavy atoms	2.3 ± 0.3
Global backbone atoms	1.9 ± 0.3
Global all heavy atoms	2.5 ± 0.3
Average Pairwise RMSD	
Ordered backbone atoms	1.5 ± 0.3
Ordered all heavy atoms	2.2 ± 0.3
Global backbone atoms	1.8 ± 0.3
Global all heavy atoms	2.3 ± 0.3

^a Ordered regions correspond to all helices at the core

Table S5.2: NMR samples and buffers (with ionic strength of 200mM NaCl) used for pH titration experiments.

¹³ C/ ¹⁵ N labeled H2A-unlabeled H2B	
Buffer type (20 mM)	pH
HAC	4.5
HAC	5.1
MES	5.6
MES	6.0
MES	6.5
Pi	6.8
Pi	7.5
Pi	7.9
CHES	9.1
unlabeled labeled H2A- ¹⁵ N H2B	
HAC	4.4
HAC	5.0
MES	5.4
MES	5.8
MES	6.2
Pi	6.8
Pi	7.1
Pi	7.5

Table S5.3: all amide pKa and CSP values for H2A and H2B during the NMR pH titration. When the combined CSP is smaller than 10 Hz then no pKa value was calculated.

H2A backbone NH apparent pKa			
Residue number	Residue name	Combined CSP (Hz)	pKa
4	GLY	31	5.66
6	GLY	23	6.25
9	VAL	17	5.87
10	LYS	19	5.94
11	GLY	31	6.20
12	LYS	31	5.52
13	ALA	190	6.10
14	LYS	133	6.25
15	SER	120	6.13
17	SER	65	5.67
18	ASN	46	5.75
20	ALA	118	5.09
21	GLY	77	5.68
22	LEU	82	5.44
23	GLN	460	4.38
24	PHE	70	4.67
26	VAL	30	4.73
27	GLY	58	5.28
28	ARG	48	5.81
29	ILE	30	5.23
30	HIS	52	4.36
32	LEU	25	4.29
33	LEU	49	4.52
34	ARG	120	4.74
35	LYS	28	5.37
36	GLY	34	5.48
38	TYR	47	4.79
39	ALA	29	6.67
40	GLU	78	5.91
41	ARG	30	5.22
42	VAL	27	4.45

43	GLY	45	4.83
44	ALA	15	4.25
46	ALA	18	6.08
51	ALA	50	5.05
54	MET	62	5.72
55	GLU	60	5.63
56	TYR	23	6.50
57	LEU	32	4.47
58	ALA	65	5.86
59	ALA	118	6.14
60	GLU	63	6.15
61	VAL	110	5.80
62	LEU	80	5.54
63	GLU	120	5.52
64	LEU	55	4.96
65	ALA	110	5.75
66	GLY	145	5.64
67	ASN	150	5.72
69	ALA	85	5.55
70	ARG	79	5.60
71	ASP	13	4.58
72	ASN	18	5.58
73	LYS	24	6.41
76	ARG	33	6.02
77	ILE	23	7.11
80	ARG	14	5.53
81	HIS	13	6.14
82	LEU	27	5.30
83	GLN	17	5.62
84	LEU	22	3.53
85	ALA	25	5.69
86	ILE	42	4.71
87	ARG	41	4.59
88	ASN	30	4.85
89	ASP	87	4.96
90	GLU	15	4.24
91	GLU	45	5.36

92	LEU	140	5.82
93	ASN	110	4.91
94	LYS	140	5.01
95	LEU	50	4.74
99	VAL	25	4.11
100	THR	27	5.04
101	ILE	13	4.18
120	GLU	18	4.18
123	ALA	26	4.08
H2B backbone NH apparent pKa			
29	LYS	28	7.32
30	ARG	18	5.10
31	LYS	21	5.23
32	GLU	12	6.10
33	SER	23	5.18
35	ALA	53	4.91
36	ILE	25	5.31
37	TYR	22	4.48
40	LYS	32	6.29
41	VAL	46	5.81
42	LEU	21	5.94
43	LYS	48	6.00
44	GLN	40	5.59
45	VAL	33	6.46
46	HIS	140	5.77
48	ASP	50	5.83
49	THR	90	5.80
50	GLY	23	6.05
51	ILE	18	5.42
52	SER	17	4.67
55	ALA	90	4.96
56	MET	49	6.62
57	SER	15	4.82
58	ILE	16	5.85
59	MET	27	5.02
61	SER	40	5.52
62	PHE	45	4.38

63	VAL	40	3.02
67	PHE	20	3.75
69	ARG	46	3.42
70	ILE	28	4.75
71	ALA	13	6.44
74	ALA	33	5.40
76	ARG	39	6.79
77	LEU	16	5.88
79	HIS	100	6.91
80	TYR	22	6.52
81	ASN	35	6.42
82	LYS	52	5.81
83	ARG	16	5.23
87	THR	61	3.77
90	GLU	18	4.23
91	ILE	26	3.63
93	THR	40	6.28
94	ALA	30	5.06
95	VAL	29	5.03
97	LEU	38	5.21
98	LEU	23	4.54
102	GLU	300	4.50
103	LEU	70	4.19
104	ALA	140	5.74
105	LYS	95	5.70
106	HIS	71	6.61
108	VAL	125	5.84
109	SER	41	5.85
110	GLU	104	5.14
111	GLY	33	4.33
114	ALA	15	4.40
115	VAL	48	4.41
117	LYS	115	5.78
118	TYR	21	7.38
119	THR	114	5.81
120	SER	79	5.79
121	SER	45	5.74

122	LYS	21	4.25
-----	-----	----	------

Chapter 6. Discussion

Biological macromolecules including proteins, nucleic acids, and membranes carry out most of the cellular functions. To fully understand the biological roles of these macromolecules it is necessary to know their three-dimensional structures. This is no less true in the field of chromatin biology, where nucleosomes and nucleosome-binding proteins cooperate to regulate the basic DNA templated processes such as transcription and DNA repair. In this thesis, I aimed to characterize the structural and dynamical properties of nucleosomes containing histone variant H2A.B. This highly divergent variant is specifically incorporated into the genome to replace the canonical histone H2A in nucleosomes in order to regulate transcription and translation. The properties and functions of H2A.B incorporated nucleosomes discovered from previous studies were introduced in **Chapter 1**. One of the most peculiar features of H2A.B nucleosomes compared to canonical ones is its open conformation of DNA entry/exit ends within the nucleosome. A complete description of the structure-function relationship for the H2A.B-nucleosome requires a high-resolution structure, which is yet unknown. The loss of DNA-protein interactions for the ultimate ~15 bp in H2A.B nucleosomes is likely to interfere with crystallization as tight packing of the DNA ends within the crystal lattice has been decisive for solving the nucleosome structure. At the same time, the partially unwrapped and dynamic nucleosomal DNA ends and possible dynamics within the octamer will likely limit the attainable resolution in cryo-EM studies. NMR spectroscopy on the other hand has developed into a viable alternative to study large macromolecular assemblies, even for systems as large as the nucleosomes with a molecular mass over 200 kDa. NMR spectroscopy has the unique advantage to be able to handle dynamic systems without necessarily compromising the atomic resolution of the data. It is also able to probe dynamics as well as interactions in addition to structure. In this thesis, I have used an NMR-driven approach to study the structure and dynamics of the H2A.B nucleosome, a prime example of large, dynamic assembly.

Isotope-labeling strategies to study H2A.B nucleosomes and histone dimers.

Incorporation of stable, NMR-active isotopes of carbon and nitrogen is an absolute requirement for NMR studies. In **chapter 2** I reviewed various isotope labeling strategies that address two important challenges: strategies to enhance sensitivity and resolution of NMR spectra for large proteins, and alternative ways of producing more cost-friendly and less laborious NMR samples. The latest developments over the past six years and the recent applications of these labeling strategies were discussed to show the possibilities and capabilities of using solution NMR spectroscopy to study macromolecular assemblies. In **Chapter 4**, various labeling schemes were applied to H2A.B nucleosomes to study its structure and dynamics. Using uniform ^{15}N -labelling, the flexible N-terminal tails of H2A.B and H3 in the nucleosomes could selectively be observed as signals from the rigid core region were effectively broadened beyond detection. To observe the folded core of H2A.B, nucleosomes were prepared with Ile- $\delta 1$ - $^{13}\text{CH}_3$, Leu, Val- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$ -H2A.B and perdeuterated H2B, H3, and H4 with unlabeled 601-DNA (ILV labeling). The scarcity of protons within the sample dramatically slows the transverse relaxation of the methyl ^1H spins, which allows to probe the Ile, Val, Leu methyl groups of H2A.B within the nucleosome. Structural and dynamical properties of H2A.B docking domain in nucleosomes were thus characterized by methyl-NOESY and ^{13}C - ^1H multiple quantum CPMG relaxation dispersion experiments. Not only large systems, but also smaller protein complexes can benefit from specific isotope labeling. As described in **Chapter 3** and **Chapter 5**, uniformly ^2H , ^{15}N -labeled H2B were refolded with unlabeled H2A.B or H2A to selective observe intermolecular NOEs in a ^{15}N -NOESY experiment, with the exception of intramolecular H_N - H_N NOEs. These NOE restraints were essential to determine the solution structures of H2A-H2B and H2A.B-H2B heterodimers.

Chemical shifts vs. NOEs in to solving histone heterodimer structures by NMR.

In **Chapter 3** and **Chapter 5**, I heavily invested in solving H2A.B-H2B and H2A-H2B dimer solution structures using NMR in combination with Rosetta. The original approach based only on

backbone chemical shifts HN, N, C α , C β , and CO have given final structural ensembles with unfixed positions of the H2B α C-helix and H2A(.B) α 1-helix. Similar observations were reported for H2A-H2B dimer using the same approach where additional H α chemical shifts were used¹. Moriwaki *et al.* argued that the observed structural variations for these helices are intrinsic properties of the dimer and that only upon incorporation into nucleosomes these structures would be fully defined. To further validate the plasticity of the dimer, Moriwaki *et al.* performed H/D exchange experiments, fast hydrogen exchange experiments, and hetero-nuclear NOE experiments to probe the dynamics of the dimer. While H/D exchange rates for amide protons located in H2B α C-helix and H2A α 1-helix were relatively rapid compared to other helices, no obvious differences were detected by fast hydrogen exchange experiments or hetero-nuclear NOE experiments. In this thesis, I similarly found no obvious differences between T₁/T₂ ratios measured for residues in the H2A(.B) α 1-helix and H2B α C-helix and the other helices in the folded core. Moreover, clear intermolecular NOE signals were observed between H2B α C-helix and H2A(.B) residues indicating a well-defined local tertiary structure. By including these NOE data I obtained a well-defined structural ensemble for the dimer core, including the H2B α C-helix and H2A(.B) α 1-helix, for both the H2A.B-H2B and H2A-H2B dimer. Since the primary structures of the fruit fly histones used in our work are highly similar to the human histones used in the Moriwaki study, with most sequence differences located at flexible tail regions, I believe that also the human dimer is in fact fully structured in solution and that similar intermolecular NOEs would be observed for that system. Vice versa, their observation of faster H/D exchange rates of amide protons of H2B α C-helix and H2A α 1-helix is probably also true in our system, and such experiment would be of interest to provide more insights into the local dynamics. It should be noted that a somewhat elevated H/D exchange rate is not at odds with a defined structural fold. Increased solvent accessibility and weaker hydrogen bonding for amides in these helices is consistent with their outside and terminal position within the fold. Vice versa, the position of the helix within the fold is mostly determined by side chain interactions which are not directly probed by the backbone chemical shift based structure determination approach, but are included in the NOESY approach. The intermolecular NOEs

measured in our studies thus provide more reliable information for constructing tertiary structures of histone dimers.

Origin of DNA unwrapping of H2A.B nucleosomes.

In **Chapter 4**, I characterized the structural and dynamical properties of the truncated docking domain of H2A.B in nucleosomes. Both NOESY and dynamics data support folding of H2A.B docking domain in the canonical conformation when incorporated in the nucleosome. The ILV labeling approach unfortunately only gives access to one isoleucine, that is centrally located in this region. The absence of more ILV residues in this C-terminal region of the docking domain precludes to obtain more signals using solution state NMR. As an alternative, solid state NMR on sedimented nucleosomes, as has recently been demonstrated ², may be used to examine in more detail the structure of the H2A.B docking domain in the nucleosomes. Both backbone C α , C β chemical shift for H2A.B docking domain residues and dipolar contact may be measured to define the structure.

The H2A.B docking domain is thought to play a role in promoting the unwrapping of the last ~15 bp of nucleosomal DNA from the histone octamer in the variant nucleosome. Replacement of the H2A docking domain with that from H2A.B resulted in nucleosomal DNA ends further apart compared to the canonical conformation ³⁻⁴. Notably, previous studies also established that the truncation of the docking domain itself is not the main determinant of DNA unwrapping. Our observations of a canonical conformation and lack of large scale motions for the H2A.B docking domain in nucleosomes thus suggest that its role in unwrapping nucleosomal DNA is due to the sequence differences within an otherwise stable and canonical fold. Considering 22 out of 31 residues forming H2A.B docking domain are different from canonical H2A and 9 out of the 22 sequence differences involve neutralized or reverse charge mutations, it is likely that altered electrostatic interactions make the dimer-tetramer interface less stable in the variant nucleosome, as was highlighted in **Chapter 3**. This could allow for slight changes in position or orientation of the H2A.B-H2B dimer relative to other components of the nucleosome eventually leading to the unwrapping of nucleosomal DNA ends. Alternatively,

the destabilized interface could allow for increased dynamics in H3, thus ‘pushing out’ the DNA.

A complicating factor in deciphering the determinants of DNA unwrapping in the H2A.B nucleosome comes from a previous study in which the H2A.B docking domain was replaced with that from H2A⁵. These chimeric nucleosomes showed several of distinctive properties of H2A.B nucleosomes: both the electrophoretic mobility and DNase I foot printing pattern largely represented the one from H2A.B nucleosomes. Interestingly, also the DNA ends formed a $\sim 180^\circ$ angle as shown from cryo electron-microscopy. These data thus point to a role for the histone fold region of H2A.B in promoting unwrapping. In **Chapter 4**, I uncovered the likely origin of nucleosome unwrapping within the H2A.B histone fold domain.

Several arginine residues of histone octamer core play essential role in stabilizing the contacts between histones and DNA. One of these is H2A R76, which is shifted by one residue position to R80 in H2A.B. I have demonstrated that a H2A mutant containing this arginine register shift reduces the resistance of its nucleosome to MNase digestion. The uniqueness of this arginine shift for H2A.B within the H2A family, together with the consistence of this arginine shift for H2A.B proteins across different species (except for mouse) supports its role in promoting an opened nucleosome structure. An interesting follow-up experiment is to make a reverse H2A.B mutant, where the H2A arginine anchor position is reintroduced to the variant, and assess its impact on counteracting the unwrapped state. Next to MNase digestion, cryo electron-microscopic and/or AFM could be used to visualize the DNA ends in the nucleosome particles and measure the length of DNA wrapped around the histone octamer. Comparison of these results with that from docking domain-swapped H2A/H2B.B nucleosomes, will allow us to decipher whether the arginine register shift, or the altered dimer-tetramer interface is the larger contributing factor to the open conformation of H2A.B nucleosomes.

While writing this chapter, a molecular dynamic simulation was running to study the opening of the DNA ends in H2A.B nucleosomes. Clear opening of the nucleosomal DNA ends has already been observed for the variant nucleosomes in such simulations. As a next step, the relation of H2A.B R80 side chain and its contacted DNA is of

particular interest to dissect the direct influence of the arginine register shift in unwrapping nucleosomal DNA.

A role for the H3 N-terminal tail in DNA unwrapping for H2A.B nucleosomes?

The H3 N-terminal tail consists of 43 residues among which ~35 residues protrude from the core particle. The H3 tail interacts with nucleosomal or linker DNA through electrostatic interactions within canonical nucleosomes. It is likely that the extent of the H3-tail/DNA interactions are related or perhaps even (co)-determine the extent of DNA unwrapping. In **chapter 4**, I found that in H2A.B nucleosomes the H3 N-terminal tail is overall less intimately bound to DNA than in canonical nucleosomes. Several residues have chemical shifts close to that of the free H3 tail, suggesting that parts of the tail are not in contact to the DNA. Other parts of the tail have a distinct interaction with the DNA, as based on their unique chemical shifts. Nevertheless, dynamics data indicate that overall the tail remains bound to DNA but is more dynamic than in canonical nucleosomes. I suggest that in H2A.B only the first ~30 residues of the tail are transiently and weakly bound to DNA, with rapidly switching short contact points. The last 5 residues are not in contact with DNA but rather ‘stretch’ from the nucleosome core to the unwrapped DNA. Overall the tail has higher mobility. Since H3 N-terminal tail mobility is correlated to its modifiability ⁶, the H2A.B nucleosome is thus potentially more readily modified on its H3 tail than the canonical nucleosome. With more than 30 known PTMs in the H3 tail ⁷, this could contribute to chromatin remodeling in its unique way.

To further characterize the role of the H3 tail in the unwrapping of the DNA, it would be interesting to also measure the dynamics of H3 α N-helix. This helix is contacted by the H2A C-terminal tail in canonical nucleosomes. Since this tail is absent in H2A.B, the loss of these contacts together with the destabilization of the dimer-tetramer interface (see above) could result in increased dynamics for this region, including transient detachment from the nucleosome core, which could be transmitted to the H3-tail. The location of seven methyl groups from

V46, L48, and I51 on the H3 α N-helix would provide sufficient number of probes for reliable determination of local dynamics.

Towards a structural basis for the H2A.B–RNA splicing link.

H2A.B was found to be involved in active gene transcription and RNA splicing mechanisms and has the ability to interact with RNA processing factors, DNA as well as RNA. Both *in vivo* and *in vitro* experiments have demonstrated the RNA binding ability of H2A.B, and this negatively regulated the H2A.B interaction with RNA Pol II and other RNA processing factors⁸. An *in vitro* RNA-pulldown assay showed the RNA binding ability of the arginine rich elongated N-terminal tail of H2A.B⁸. I observed that this tail is highly flexible within the H2A.B nucleosome. Future RNA titration experiments on a H2A.B N-tail peptide and H2A.B nucleosomes could be used to gain more insights into how H2A.B interacts with RNA at the molecular level.

In addition, the H2A.B nucleosomes likely also interacts with other proteins or splicing factors through the nucleosome core. The partly neutralized acidic patch of H2A.B nucleosomes would be suitable docking site for such proteins, as is the exposed H2A.B octamer surface due to the unwrapping of the nucleosomal DNA ends. While several such RNA splicing factors were enriched in H2A.B nucleosomes in a mass spectroscopy based screening, these results still need to be validated as direct and specific interactions to the nucleosome. The data described in this thesis will be helpful to generate a refined integrative structural model of the H2A.B nucleosome, based on our NMR data and previously recorded SAXS data⁹ and SANS data¹⁰. Structure-based models of the complex could then be made to guide further experimentation to investigate the function of these H2A.B nucleosome-protein complexes.

Electrostatic contributions to protein stability and interactions.

Protein fold stability is determined by various interactions, including hydrophobic, electrostatic, hydrogen bonding, van der Waals interactions. Despite the enormous progress in understanding how these forces are responsible for protein stability, to what degree electrostatic interactions stabilize protein folding is still a subject of debate. Histones are highly charged proteins. Histone variant H2A.B has a significant amount of charge neutral or reverse mutations compared to its canonical counterpart. I showed in **Chapter 3** that the variant dimer and the canonical dimer have essentially the same folding in solution, thus histone H2A-H2B dimer can serve as a good model system to investigate the role of surface charge-charge interactions in protein stability. A significant increase in thermostability of H2A.B-H2B dimer in comparison to H2A-H2B dimer was discovered. By selectively substituting H2A.B residues with H2A residues at the sites where charge is swapped or neutralized and vice versa, an overall correlation between net charge of the dimer core and the thermostability was discovered and it was concluded that this is due to the reduced electrostatic repulsion between the two histone monomers. Further experiments would be needed to determine how this increased thermostability translates stability at physiological temperature. Notably, the identified sequence differences that responsible for the increased thermostability are widely conserved in the H2A.B family across different species, suggesting also the increased stability is conserved and may thus serve a specific biological function. A possible explanation would be that the increased stability of the H2A.B-H2B dimer promotes its exchange from and to the nucleosomes, accounting for the higher mobility of this protein in cells.

Along with stabilizing protein structures, electrostatic interactions are also important for protein-protein interactions. While H2A, H2B, H3, and H4 can fold into the histone octamer directly by salt gradient dialysis in vitro, H2A.B cannot be refolded into the octamer with the other core histones through this conventional method. Instead, H2A.B-H2B dimers and H3-H4 tetramers have to be mixed together with DNA to reconstitute the variant nucleosome. I highlighted in **Chapter 3** that the failure in refolding may in part be due to the altered electrostatics in the H3-H4 binding interface of H2A.B. This region is only partly folded in the H2A.B-H2B dimer and becomes fully folded in the

nucleosomes, as shown in **Chapter 4**. In particular the elements that form the direct interface between the dimer and tetramer subunits are unstructured in the free state and are only folded in the associated state. This raises the question as to how the dimer recognizes the H3-H4 interface in the initial stage of assembly.

Finally, in **Chapter 5** the charged state of the so-called acidic patch on the surface of the H2A-H2B dimers, which forms a key binding platform for various nucleosome-binding proteins, was investigated. I experimentally assessed the side chain pKa's for titratable residues in and around the acidic patch residues using pH titration NMR experiments. Notably, while two acidic patch residues were predicted pKa's to have pKa values above 5, I found that neither have elevated pKa's despite their clustering. On the contrary, acidic patch residue H2A D89 was determined to have an elevated apparent pKa value. I further highlighted a conserved histidine in H2B that directly lines the acidic patch, whose pKa is reasonably close to the physiological pH. This histidine may thus be able to reduce overall negative electrostatic potential of the acidic patch depending on the local environmental pH. This may further result in reduced binding of acidic-patch binding proteins, thus affecting chromatin biology.

Final remarks.

In this study, I have achieved solution structure for H2A.B-H2B dimer, uncovered the structural and dynamical properties of H2A.B C-terminal docking domain and H3 N-terminal tail in nucleosomes, and verified the importance of H2A.B R80 in opening nucleosome conformations. Altogether, I improved the H2A.B nucleosome model. This study offers fundamental knowledge of the H2A.B nucleosome and provides a solid basis for future experiments to understand its function *in vivo*.

References

1. Moriwaki, Y.; Yamane, T.; Ohtomo, H.; Ikeguchi, M.; Kurita, J.; Sato, M.; Nagadoi, A.; Shimojo, H.; Nishimura, Y., Solution structure of the isolated histone H2A-H2B heterodimer. *Sci Rep* **2016**, *6*, 24999.
2. Xiang, S.; le Paige, U. B.; Horn, V.; Houben, K.; Baldus, M.; van Ingen, H., Site-Specific Studies of Nucleosome Interactions by Solid-State NMR Spectroscopy. *Angew Chem Int Ed Engl* **2018**, *57* (17), 4571-4575.
3. Bao, Y.; Konesky, K.; Park, Y. J.; Rosu, S.; Dyer, P. N.; Rangasamy, D.; Tremethick, D. J.; Laybourn, P. J.; Luger, K., Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *EMBO J* **2004**, *23* (16), 3314-24.
4. Shukla, M. S.; Syed, S. H.; Goutte-Gattat, D.; Richard, J. L.; Montel, F.; Hamiche, A.; Travers, A.; Faivre-Moskalenko, C.; Bednar, J.; Hayes, J. J.; Angelov, D.; Dimitrov, S., The docking domain of histone H2A is required for H1 binding and RSC-mediated nucleosome remodeling. *Nucleic Acids Res* **2011**, *39* (7), 2559-70.
5. Doyen, C. M.; Montel, F.; Gautier, T.; Menoni, H.; Claudet, C.; Delacour-Larose, M.; Angelov, D.; Hamiche, A.; Bednar, J.; Faivre-Moskalenko, C.; Bouvet, P.; Dimitrov, S., Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *EMBO J* **2006**, *25* (18), 4234-44.
6. Stutzer, A.; Liokatis, S.; Kiesel, A.; Schwarzer, D.; Sprangers, R.; Soding, J.; Selenko, P.; Fischle, W., Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Mol Cell* **2016**, *61* (2), 247-59.
7. Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. *Cell Res* **2011**, *21* (3), 381-95.
8. Soboleva, T. A.; Parker, B. J.; Nekrasov, M.; Hart-Smith, G.; Tay, Y. J.; Tng, W. Q.; Wilkins, M.; Ryan, D.; Tremethick, D. J., A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLoS Genet* **2017**, *13* (2), e1006633.
9. Arimura, Y.; Kimura, H.; Oda, T.; Sato, K.; Osakabe, A.; Tachiwana, H.; Sato, Y.; Kinugasa, Y.; Ikura, T.; Sugiyama, M.; Sato, M.; Kurumizaka, H., Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Sci Rep* **2013**, *3*, 3510.
10. Sugiyama, M.; Arimura, Y.; Shirayama, K.; Fujita, R.; Oba, Y.; Sato, N.; Inoue, R.; Oda, T.; Sato, M.; Heenan, R. K.; Kurumizaka, H., Distinct features of the histone core structure in nucleosomes containing the histone H2A.B variant. *Biophys J* **2014**, *106* (10), 2206-13.

Summary

Nature uses a special class of histone proteins, histone variants, to modulate the properties of chromatin at defined genomic locations. H2A.B is one of the most divergent H2A variants and is involved in important cellular functions, such as transcription and mRNA splicing. Incorporation of H2A.B in nucleosomes causes unwrapping of ~15 bp entry/ exit nucleosomal DNA from the histone octamer core. Yet, the molecular basis of such peculiar nucleosome conformation is unclear. The work described in this thesis aimed to determine the impact of H2A.B incorporation on the structural and dynamical properties of the nucleosome, primarily using nuclear magnetic resonance (NMR) spectroscopy.

Chapter 1 summarizes the current knowledge on the structure and function of histone variant H2A.B. I highlight the sequence differences between the variant H2A.B and canonical H2A, present an overview of the studies that uncovered the opened nucleosome structure for H2A.B, and conclude by discussing its role in active transcription and mRNA splicing. **Chapter 2** provides a compact review of isotope-labeling techniques used in solution NMR for studies of macromolecular complexes, such as nucleosomes, with particular emphasis on the range of options available in terms of labeling strategy.

In the first step of our studies on H2A.B, the solution structure of the H2A.B-H2B heterodimer was determined using NMR spectroscopy and Rosetta, as described in **Chapter 3**. Through the combination of backbone chemical shifts and sparse intermolecular NOE distance restraints, a converged ensemble of structures could be obtained. The H2A.B-H2B structure shows that the dimer core is folded in the canonical histone handshake motif from which highly flexible N- and C-terminal regions protrude. The variant dimer structure is highly similar to the structure of canonical H2A-H2B dimer. I further found that the H2A.B-H2B dimer has a higher thermostability than the canonical H2A-H2B dimer. On the basis of an analysis of several mutant dimer proteins, I conclude that the lower the net charge of the H2A.B core region, resulting in reduced electrostatic repulsion with the H2B core, is responsible for the increased stability.

I next investigated the structural and dynamical properties of the H2A.B nucleosomes in **Chapter 4**. First, using an isotope-labeling strategy optimized for the highly flexible parts of the nucleosome, I studied in detail the H2A.B and H3 N-terminal histone tails within the variant nucleosome. Our data show the tails have similar length as the canonical nucleosome structure, indicating that the α N helix in H2A.B and H3 are well folded in the H2A.B nucleosome, despite its more open structure. Furthermore, our data suggest decreased DNA binding and increased flexibility for the H3 N-tail in the variant nucleosome. Second, using methyl-specific labeling, I probed the conformation and dynamics of the H2A.B core within the nucleosome. Supported by both structural and dynamical data, I found that the H2A.B docking domain, which forms the interface to H3-H4, is folded into the canonical conformation upon incorporation into the nucleosome. On the basis of the updated structural model for the H2A.B nucleosome, I identified a register shift of a DNA-binding arginine in H2A.B that would cause a clash with the nucleosomal DNA and could thus cause opening of the DNA ends. Nucleosomes formed from an H2A mutant with the arginine shifted in position as in H2A.B showed significantly increased digestion of the nucleosomal DNA by micrococcal nuclease, highlighting that the arginine register shift is a contributing factor to the opened nucleosome conformation for H2A.B.

In **chapter 5**, I studied the acidity of the acidic patch, a key protein binding site on the surface of the nucleosome, by experimentally determining the pKa values of residues in and around the acidic patch by NMR. Prediction of pKa values based on the solution structure of the canonical H2A-H2B dimer, determined following a similar approach as in Chapter 3, indicated that several acidic residues may have elevated pKa's due to the clustering of many acidic residues in this patch. However, NMR pH titration experiments showed clearly that all glutamate side chains have pKa values below 4.5, in contrast to the predicted values. I further identified H2B H106 as the group that titrates most closely to the physiological pH range. Its proximity to the binding interface of several nucleosome-binding proteins suggests that protonation of this histidine upon lowering pH to sub-physiological values may decrease the binding of acidic patch binding proteins.

Chapter 6 offers a general discussion of the results obtained in this thesis in perspective of NMR-driven integrative structural biology, the

structure-function relationship of H2A.B, and the role of protein electrostatics in general. Suggestions are made for future experiments that may add to our knowledge on the structure and function of H2A.B nucleosomes.

Samenvatting

De natuur gebruikt een speciale klasse van histon-eiwitten, de zogenaamde histon-varianten, om eigenschappen van chromatine op specifieke plekken in het genoom te veranderen. Histon variant H2A.B is een van de meest afwijkende varianten van histon H2A en is betrokken bij een aantal belangrijke cellulaire functies, zoals transcriptie en mRNA-splicing. Door incorporatie van H2A.B in nucleosomen laten circa 15 baseparen van de uiteindes van het nucleosomale DNA los van de histoneiwitkern. De onderliggende moleculaire basis van deze bijzondere, geopende nucleosoomconformatie is nog niet bekend. Het werk beschreven in dit proefschrift heeft als doel om het effect van H2A.B-integratie op de structurele en dynamische eigenschappen van het nucleosoom te bepalen, voornamelijk met behulp van kernspinresonantie (NMR).

Hoofdstuk 1 vat de huidige kennis over de structuur en functie van histonvariant H2A.B samen. Ik belicht de verschillen in aminozuurvolgorde tussen de variant H2A.B en de canonieke H2A, geef een overzicht van het experimentele bewijs voor de geopende nucleosoomstructuur en bespreek de rol van H2A.B in transcriptie en mRNA-splicing.

Hoofdstuk 2 biedt een compact overzicht van isotooplabeling technieken die worden gebruikt in NMR aan vloeistoffen met macromoleculaire complexen, zoals nucleosomen.

Als eerste stap in deze studie van H2A.B wordt de structuurbepaling van de H2A.B-H2B heterodimeer met behulp van NMR-spectroscopie en Rosetta beschreven in **Hoofdstuk 3**. Het combineren van de chemische verschuivingen van de kernen in de eiwit-backbone met intermoleculaire NOEs resulteerde in een goed gedefinieerd ensemble van structuren. De H2A.B-H2B-structuur laat zien dat de kern van de dimeer is gevouwen volgens het klassieke 'histone-handshake' motief waaruit zeer flexibele N- en C-terminale staarten steken. De structuur van de variant lijkt sterk op die van het canonieke H2A-H2B dimeer. Ik heb verder ontdekt dat de H2A.B-H2B dimeer een hogere thermostabiliteit heeft dan de canonieke H2A-H2B dimeer. Op basis van een analyse van verschillende mutanten concludeer ik dat de lagere netto lading van de H2A.B-kern resulteert in verminderde

elektrostatistische afstoting met de H2B-kern, en daardoor in verhoogde stabiliteit.

Vervolgens heb ik onderzoek gedaan naar de structurele en dynamische eigenschappen van de H2A.B-nucleosomen in **Hoofdstuk 4**. Met behulp van een isotoplabeling strategie geoptimaliseerd voor de zeer flexibele delen van het nucleosoom, heb ik eerst in detail de N-terminale staarten van H3 en H2A.B onderzocht. De gegevens laten zien dat de lengtes van de staarten zijn zoals verwacht op basis van de canonieke nucleosoomstructuur. Dit wijst erop dat de α N-helix van H2A.B of H3 stabiel gevouwen zijn, ondanks de geopende conformatie van het H2A.B nucleosoom. Daarnaast is de H3-staart flexibeler en waarschijnlijk minder sterk aan DNA gebonden in H2A.B nucleosomen. De conformatie en dynamiek van de H2A.B-kern in het nucleosoom is vervolgens onderzocht met behulp van specifieke isotoplabeling van methyl groepen. Op basis van zowel structurele als dynamische gegevens, concludeer ik dat zogenaamde ‘docking’ domein van H2A.B, dat de interface vormt met histon H3-H4, stabiel gevouwen is in het nucleosoom. Met hulp van het bijgewerkte structurele model voor het H2A.B-nucleosoom kon ik een verschuiving van een DNA-bindende arginine in de aminozuurvolgorde van H2A.B identificeren. Door deze verschuiving in register zou de zijketen kunnen botsen met het nucleosomale DNA en dus bijdragen aan de opening van de DNA-uiteinden. Nucleosomen gemaakt met een H2A-mutant waarin deze arginine is verschoven naar de positie zoals in H2A.B vertoonden een aanzienlijk verhoogde gevoeligheid voor restrictie van het nucleosomale DNA door micrococcal nuclease. Dit laat zien dat de arginine-registerverschuiving bijdraagt aan de geopende nucleosoomconformatie voor H2A.B.

In **hoofdstuk 5** heb ik de pKa-waarden van residuen in en rond de ‘acidic patch’, een bindingsplaats voor vele eiwitten op het oppervlak van het nucleosoom, experimenteel bepaald met behulp van NMR spectroscopie. Allereerst heb ik de structuur van het canonieke H2A-H2B-dimeer bepaald volgens dezelfde methode als beschreven in Hoofdstuk 3 voor de variant-dimeer. De op basis van deze structuur voorspelde pKa waarden zijn voor verschillende zure residuen verhoogd als gevolg van een clustering effect. Desalniettemin tonen de NMR pH-titratie experimenten duidelijk aan dat de pKa-waarden van

alle glutamaten lager zijn dan 4.5. Ik heb daarnaast H2B H106 geïdentificeerd als de groep die het dichtst bij de fysiologische pH titreert. Dit residu bevindt zich dicht bij de bindingsplek van verschillende nucleosoom-bindende eiwitten, wat suggereert dat protonering van deze histidine verantwoordelijk kan zijn voor vermindering van bindingsaffiniteit.

Hoofdstuk 6 bevat een brede bespreking van de resultaten verkregen in dit proefschrift vanuit het perspectief van integratieve structurele biologie, de structuur-functie relatie van H2A.B en de rol van eiwit-elektrostatica in het algemeen. Er worden suggesties gedaan voor vervolgexperimenten die kunnen bijdragen aan onze kennis over de structuur en functie van H2A.B-nucleosomen.

Curriculum Vitae

Heyi Zhang was born on April 16th 1990 in Wuxi, China. She completed her Bachelor in Bioengineering at China Pharmaceutical University. She then moved to Delft, the Netherlands to study Life Science and Technology at TU Delft, where she became a Master in Science and an Engineer. Soon after, she began her PhD research in the field of structural biology at Leiden University in the group of Dr. Hugo van Ingen with Prof. Dr. Marcellus Ubbink as promotor. From year 2017 to 2019, she followed Dr. Hugo van Ingen to Utrecht University where she completed the studies presented in this book.

Publications

- 1, **Zhang H**, Eerland J, Horn V, van Ingen H. The electrostatic potential of the nucleosome acidic patch. In preparation.
- 2, **Zhang H**, Lin Y, Lobbia V, Cojocaru V, Huertas J, van Ingen H. Structure and dynamics of the H2A.B variant nucleosome. In preparation.
- 3, **Zhang H**, Dubbeldam S, van Ingen H. Variant H2A.B-H2B histone dimer is more stable than the canonical dimer. Submitted.
- 4, Horn, V.; Uckelmann, M.; **Zhang, H.**; Eerland, J.; Aarsman, I.; le Paige, UB.; Davidovich, C.; Sixma, TK.; van Ingen, H. Structural basis of specific H2A K13/K15 ubiquitination by RNF168. *Nat Commun.* 2019, 10(1):1751.
- 5, **Zhang, H.**; van Ingen, H. Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. *Curr. Opin. Struct. Biol.* 2016, 38, 75–82.