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

Zhang, H.

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-δ1-[$^{13}CH_3$], Leu, Val-[$^{13}CH_3$, $^{12}CD_3$]-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 ^{1}H 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-^{1}H 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 ^{2}H, ^{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, Cα, Cβ, and CO have given final structural ensembles with unfixed positions of the H2B αC-helix and H2A(.B) α1-helix. Similar observations were reported for H2A-H2B dimer using the same approach where additional Hα chemical shifts were used [1]. 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 αC-helix and H2A α1-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 T_1/T_2 ratios measured for residues in the H2A(.B) α1-helix and H2B αC-helix and the other helices in the folded core. Moreover, clear intermolecular NOE signals were observed between H2B αC-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 αC-helix and H2A(.B) α1-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 αC-helix and H2A α1-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 Cα, Cβ 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 H2A [5]. 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 ~180° 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 αN-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 αN-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 factors [8]. An in vitro RNA-pulldown assay showed the RNA binding ability of the arginine rich elongated N-terminal tail of H2A.B [8]. 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 data [9] and SANS data [10]. 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.