How electrostatic interactions drive nucleosome binding of RNF168 & PSIP : structural studies and their implications for rational drug design

Horn, V.

5

General Conclusions & Perspectives

JOINING FORCES IN STRUCTURAL BIOLOGY

Understanding the organization and activity of DNA has come a long way since Walther Flemming first introduced the term chromatin in the 1880s. A milestone herein is the discovery and subsequent determination of an atomic resolution structure of the fundamental basis of chromatin, the nucleosome. This facilitated the understanding of how effector proteins operate on the nucleosome and drive chromatin biology. Since a growing number of these proteins is attributed important roles in pathologies, an increasing interest in inhibitors for protein-nucleosome interactions is emerging. To direct the efforts to find modulators for epigenetic interactions, an understanding of the underlaying mechanisms of binding and recognition is vital. This raises the need to resolve the 3D structures of nucleosome-protein complexes. The main techniques used in the structural biology study of these complexes are X-ray crystallography, NMR spectroscopy and cryo-electron microscopy, all of them with their own advantages and limitations. While X-ray crystallography is able to yield high-resolution structures, its general limitation is the crystallization process and possible molecular rearrangements upon crystallization. Even though cryo-electron microscopy leads to increasingly well resolved structures and freezes the molecule in its native fold, it has the same limitation as X-ray crystallography: the resulting structures are isolated snapshots that lack dynamics and pathways between conformational states. However, dynamics are of high importance in the interaction of biomolecules. Here, NMR spectroscopy offers insights into the dynamic nature of protein complexes and is especially well suited for flexible proteins. Additionally, NMR offers a clean robust interrogation of protein interactions, at atomic resolution. In return however, NMR is limited by the size of the molecular assembly and studying large complexes requires sophisticated isotope labelling schemes. Even though the first crystal structure of the nucleosome was published in 1997, the limitations of the above techniques caused the number of atomistic resolution structures of nucleosome-bound effector proteins or protein complexes to be limited compared to the entirety of protein-nucleosome interactions. This sparked ever more successful applications of the integrative use of biophysical data on different aspects of protein binding in data-driven modelling to resolve otherwise inaccessible structures. The results of these remarkable studies are reviewed in **Chapter 1**.

In this thesis, I aimed to show the immense potential of integrative modelling to direct the development of inhibitors of epigenetic effector protein activity.

The Power and Challenges of Structural Modelling as Seen by RNF168

In **Chapter 2**, we investigated the molecular basis of how the ubiquitin E3 ligase RNF168 recognizes its nucleosomal substrate and directs its corresponding E2 subunit towards its target residue for specific ubiquitination. Besides valuable previously reported insights into the activity and identification of crucial residues of this epigenetic writer protein, a structure of a nucleosome-bound state was yet to be shown. We used NMR spectroscopy to show that RNF168 binds the acidic patch on the interface between histone H2A and H2B. On the RNF168 side, we showed with mutagenesis that besides the previously reported residue R57, various arginine residues distributed over an α-helix compose the interaction interface with the nucleosome. With the use of this data in structural modelling alone, two energetically similar conformations of the RNF168-nucleosome complex were obtained that differed by 180° in their RNF168 orientation on the nucleosome surface. This is of special interest in the light of substrate specificity. While the addition of the corresponding E2 subunit UbcH5c to one conformation positioned the catalytic center in good agreement with ubiquitination of lysine on the N-terminal tail of H2A, the second directed the ubiquitination complex towards the C-terminal tail. Interestingly, another E3 ligase, Ring1B, is reported to ubiquitinate lysine residues in the latter region. This is however not the case for RNF168 and posed an important challenge to address. The information accessible by NMR and mutagenesis allowed in this case for the mapping of interaction surfaces on both nucleosome and RNF168 but did not provide information on their relative orientation or on intermolecular distances in this case. This can potentially lead to additional ambiguity in the resulting structures. Additionally, potential incomplete mapping of the interface may bias the outcome. Another possible issue is that conformational changes of subunits upon binding cannot be reliably modelled. It thus could be conceived that structural rearrangements within the H2A/H2B dimer upon binding would favor the target specific positioning of RNF168. To discriminate between the two possible orientations of RNF168 on the nucleosome, cross-linking mass spectrometry was employed to get additional distance restraints. Indeed, a cross-link specific for the N-terminal orientation allowed for a proper discrimination. Furthermore, we validated our model by engineering a H2B E110A mutant. This acidic patch residue is in the interaction interface of only the conformation that directs for N-terminal ubiquitination. Indeed, the mutant construct was observed to silence RNF168 activity. In contrast, Ring1B activity was not affected. This serves as a prime example of how different biophysical techniques can be employed to derive a structural model and highlights an important additional point.

Structural modeling will always result in a structure and its quality crucially depends on the input data, thus requiring careful biochemical validation.

The Shape of the Basic Patch Potentially Dictates Specific Acidic Patch Binding

While we were able to verify the orientation of nucleosome-bound RNF168-UbcH5c, the fundamental molecular interactions that are causing the drastic change in orientation on the nucleosome between RNF168 and Ring1B still have to be proven. One hypothesis is that the RNF168 basic helix and a surface groove on the acidic patch align especially well for specific insertion. This is supported by the importance of all arginine residues in the basic anchor helix. Interestingly, acidic patch binding domains often show these additional basic residues that interact with the acidic patch (Figure 5-1). While the "arginine anchor" is generally attributed with the binding to the acidic patch, the relative position of secondary positively charged residues may favor a specific orientation by a well-defined network of electrostatic interactions with the acidic patch. This is possibly true for RNF168 as well. It cannot be excluded that mutants that lack Arg residues, other than the arginine anchor, in the anchor helix still bind the acidic patch, but are not positioned properly for effective ubiquitination. While this is also suggested for Arg mutants in our study, additional studies investigating the specific binding modes of these mutants will yield significant information on the precise impact of the overall shape and size of acidic patch binding epitopes. This is of particular interest, since we showed that structure-guided interference with the electrostatic interaction surface allows for selective *in vitro* silencing of RNF168, as demonstrated by the H2B E110A nucleosome mutant. Following this rational, it might be possible to select specific histone mutations for a large variety of acidic patch binders essentially giving control over selective activity in vitro. Furthermore, identification of specific basic patch interaction sites in combination with structural models can help to identify binding epitopes to be targeted by small molecule inhibitors.

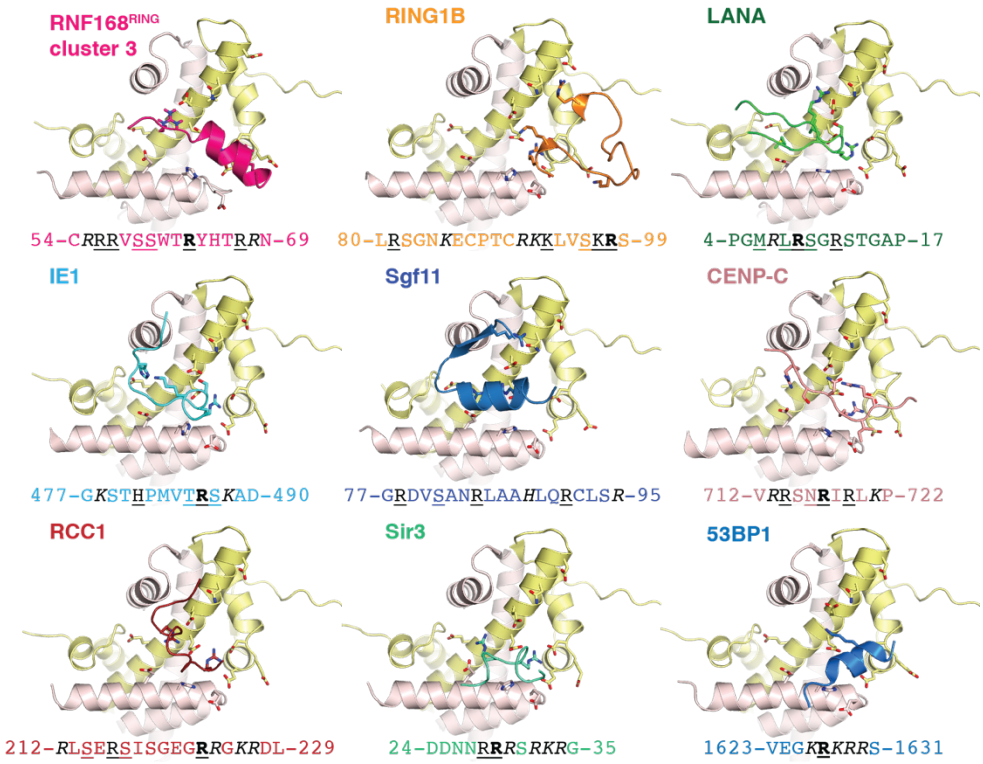

Figure 5-1. Network of electrostatic interactions for selected acidic patch binding proteins.

The acidic patch binding elements of RNF168, RING1B, LANA, IE1, Sgf11, CENP-C, RCC1, Sir3 and 53BP1 are show as cartoon with selected side chains involved in the interaction shown as sticks. The corresponding amino acid sequence is shown below each figure. Underlined residues are involved in hydrogen-bonding or hydrophobic interaction with the acidic patch. Positively charged residues shown in black, arginine anchor in bold, flanking residues in italic.

The Use of Structural Models in Rational Drug Design

In our work on RNF168, we made use of our structural model to rationally design a nucleosome mutant that selectively interferes with the activity of epigenetic effector proteins. This is of special importance since the regulation of histone modifications is a promising target for epigenetic therapeutics. An example hereof is the inhibition of acetyllysine reader domains and histone deacetylases by small molecules. While there is a growing number of drug candidates advancing to clinical trials, the number of potential inhibitors for the readout of trimethyllysine (Kme3) modifications is significantly lower. This is due to the shallow nature of the epitope that recognizes it, the aromatic cage. One of these aromatic cage-containing proteins is PSIP1 that recognizes a trimethylated lysine on the histone H3 tail via its PWWP reader

domain. It was previously shown that PSIP1-PWWP shows the lowest binding affinity for this specific modification in context of a H3 tail peptide model. Furthermore, it was established that the underlaying reason is a synergetic binding of PSIP1 to Kme3 through the aromatic cage and electrostatic interactions between positive residues on the protein surface and nucleosomal DNA. These insights were derived from a structural model based on NMR data and mutagenesis. In **Chapter 3**, we set out to use insights gained from a structural model to design H3 tail-derived peptides that incorporate both Kme3 and glutamate residues to mimic the electrostatic environment of nucleosomal DNA. We were able to show that this increases the affinity of these substrates for PSIP1-PWWP binding significantly. This is especially true for a branched peptide where an orthogonal glutamate backbone gives a more nucleosome-resembling structure to the negative charges.

NOTHING IN LIFE'S FOR FREE - TRADING SPECIFICITY FOR AFFINITY

While we observed an up to 50-times increase in affinity, investigating the protein-peptide interactions by NMR spectroscopy led to an interesting observation. We showed that the increase in affinity is mostly based on the electrostatic binding of the positive DNA interaction surface of PSIP1-PWWP while a decreasing fraction of the native H3 tail sequence is bound to the aromatic cage. This is of high interest since it shows that the electrostatic interactions lower specificity and are in competition with aromatic cage binding. In context of the native nucleosome-reader complex the question is raised how complex formation is achieved with such specificity in vivo. A possible explanation might be the static nature of nucleosomal DNA compared to highly flexible peptide tails. It might therefore be of interest to restrict the conformational space of the flexible peptide backbone and study the subsequent effect on the complex structure. We also hypothesized that a sterically non-optimized linker length between Kme3 and the negative charged backbone can cause a competition in binding their respective epitope. Therefore, a variation in linker length might deliver important insights into how the selectivity of H3 tail-derived peptides can be increased. This is of utter importance since, besides a high binding affinity, selectivity is a main criterium for a potential inhibitor.

IMPROVING THE SELECTIVITY OF PSIP1 BINDING H3 TAIL PEPTIDES

While we showed how the affinity of PSIP1-PWWP binding peptides can be increased, we paid for that with a loss in specificity. Even though variations in peptide

structure might result in higher affinity while retaining specificity, we additionally focused on the improvement of aromatic cage binding. In **Chapter 4**, we introduced a synthesis protocol that allows for the introduction of functional groups based on a cysteine residue in place of Kme3. This enables the screening for side chains that increase aromatic cage binding and subsequently target specificity. We showed the applicability of our protocol by synthesizing a cysteine-based Kme3 mimic in a proof-of-concept study. The suitability for other side chain structures however is still to be shown. We compared the resulting complex structures by NMR spectroscopy and found that the sulfur in place of carbon seems to have a negative effect on aromatic cage binding. We see that the negatively charged N-terminal peptide stretch encounters positive PSIP1-PWWP residues not observed for the native Kme3 equivalent of the peptide while the fraction of bound H3 tail sequence is decreased. While this seems counter intuitive, we must not forget that the mimicked Kme3 shows an initially already poor affinity for the aromatic cage. Therefore, the effect of the sulfur might be especially pronounced in this case. A more optimized engagement of the aromatic cage by a functional group other than trimethylammonium might likely render the sulfur-induced effect insignificant.

Fluorescence Screening of Potential Peptide Inhibitors

With a fast synthesis protocol for peptide substrates with varying aromatic cage-interacting side chain functional groups at hand we intended to improve the screening of these potential lead compounds. While NMR spectroscopy delivers vital insights into binding epitopes and structure of a complex, it is demanding on both time and sample consumption. We therefore set up an assay based on microscale thermophoresis (MST). For this purpose, we designed a mutant construct of PSIP1-PWWP for fluorescent labelling. With this at hand, we were able to reproduce the results obtained from NMR by fitting of specific fluorescence quenching, induced by peptide binding. The quenching showed to be specific for peptides with intact aromatic cage interaction and can therefore serve as a specific reporter for aromatic cage binding throughout screening for alternative functional side chain groups. Furthermore, we showed by NMR that peptides without aromatic cage engagement still bind to PSIP1-PWWP. This means that unalkylated peptides are likely to still show an MST effect. Though this has yet to be proven, it would make the use of an MST assay unnecessary and favor readout of initial fluorescence and fitting the data for binding dependent fluorescence quenching. Together, this provides an even more applicable approach to screen for aromatic cage binding moieties to increase the specificity and affinity of H3 tail-derived PSIP1-PWWP binders.

Closing Remarks

The studies presented in this work show the potential of the integrative use of biophysical data in defining the structural basis of protein interactions. Even if the results obtained hold a degree of ambiguity, this approach allows to iteratively refine and validate the model and interpret its meaning for the molecular basis of protein function. Often all three points at the same time. This dynamic nature makes the use of structural models in the design of therapeutic compounds especially useful since the inhibition of a certain protein function might not require a structure to be accurate down to the last atom but rather highlight key interactions or structural features that can be addressed in context of small molecule or peptide inhibitors. I therefore strongly advocate for using and sharing biophysical data, even seemingly not publishable negative results. This has the potential to allow for a synergetic rather than competitive use of techniques in structural biology to overcome their intrinsic limitations.