# Non-conserved C-terminal regions in the histone recognition module CW, determines specificity, stability and affinity towards monomethylated histone tails


By

Øyvind Ødegård

This thesis is submitted as a partial fulfilment of the requirements for degree Master of Science

Department of molecular biology University of Bergen, Norway

November 2014

November 2014

Interactions of the CW domain with methylated histone tails: Biophysical characterization of the molecular determinants for affinity

Øyvind Ødegård

# Acknowledgements

This master thesis was carried out at the Department of Molecular Biology, University of Bergen, Norway in the period from August 2013 and December 2014 in the NucReg research program. The supervisors were Rein Aasland and Øyvind Halskau. Co-supervisors were Carol Issalene and Øyvind Strømland.

First I would like to thank Rein Aasland for his enthusiasm, support and hard questions. I would also like to thank him for introducing me to chromatin and epigenetics. Furthermore I would like to thank Øyvind Halskau for his open door policy, and for introducing me to the exiting field of NMR. I also want to thank my co- supervisors Carol Issalene and Øyvind Strømland for your never ending optimism and friendship.

This project was made possible by the external help from Nils Åge Frøystein and Olena Dobrovolska. Their help in experimental NMR setup and analysis has been highly valued. Furthermore I would like to thank Morten Vik for his help on various computer problems and expertise on C# programming, and Carina Vibe for her friendly support and her English grammar and spelling skills.

I would also like to thank my colleagues and friends at MBI for always being helpful and for making an atmosphere where science becomes fun.

Last but not least I want to thank my friends and family for their constant support and encouragement.

Bergen, November 2014

Øyvind Ødegård

# Abstract

Histone modifications have an important role in epigenetic gene regulation and contribute to the cellular memory system. It has been well documented that certain patterns of histone modifications are associated with different states of gene expression. The mechanisms underlying these associations are, however, still poorly understood. Many nuclear proteins have histone recognition modules that can serve to recruit the proteins to relevant sites in chromatin. The CW domain is a family of histone recognition modules found in chromatin-related proteins. While all known CW domains have specificity for histone H3 methylated on lysine 4, different subfamilies of CW domains have different specificities for different methylation states. The C-terminal “lid regions" of the CW domains are highly variable and they have been proposed to contribute to the ligand specificity. To explore how the C-terminal "lid regions" contribute to ligand specificity, this project aims to measure the relative mobility of these regions in *A. thaliana* ASHH2 CW domain using NMR experiments. To support our claims, intrinsic tryptophan fluorescence thermal denaturation and ITC have been used. We have engineered three ASHH2 CW domain variants, which differ in length in their N- and C-terminal regions. We show that the lid regions are important for the stability of the domain, and that the protein adopts a more stable confirmation upon ligand binding. The C-terminal helix is mobile in the unbound state but becomes less mobile upon binding. We also show that K9 has a small effect on binding since doubly modified H3K4me1K9ac has a higher $K_D$ than H3K4me1. Furthermore, the implications for the binding mechanisms are discussed.

# Table of contents

# Selected abbreviations

- PTM – Post translational modification
- DC – Dynamics Centre
- H3K4 – Histone 3 lysine four. A detailed nomenclature of PTMs on H3 is found in Table 2.17
- HRM - histone recognition modules
- HSQC / ($^{1}H\,^{15}N$- HSQC) -- heteronuclear single quantum correlation NMR experiment that observe $^{1}H\,^{15}N$ nuclei.
- ITC – Isothermal Titration Calorimetry
- NMR – Nuclear Magnetic Resonance spectroscopy
- PHD finger - Plant Homeo Domain (PfamID: PF00628)
- $S^2$ – Order parameter
- $S_f^2$- Order parameter for fast movement
- SET domain - lysine methyltransferase enzyme (PfamID: PF00856)
- $\tau_{c/e/s}$ – Correlation times

# 1 Introduction

A central question in biology has been how eukaryotic organisms obtain their large cell-type diversity, despite the fact that the same genetic content is the template for gene expression in all cell types. The key to understanding a given cell phenotype lies in differential gene expression. The classic way of imagining gene transcription is that a signal, communicated by transcription factors, triggers the recruitment of RNA polymerase pre-initiation complexes, resulting in transcription and eventually translation. However, this signal-product line of thought is not sufficient to explain how cells in the same organism obtain the wide variety of phenotypes that is observed in nature.

Walter Flemming, who studied cell division by mitosis in the 1880's, noticed that cells in interphase had a non-homogeneous threadlike substance in the nucleus (Flemming 1882; reviewed by Olins & Olins 2003). This was later identified as a protein-DNA complex that was coiled up in higher order structures, eventually termed chromatin. In chromatin, DNA is coiled around a protein component which consists of histones (Kossel 1911; reviewed by Olins & Olins 2003). This structure is in turn associated with a number of non-histone proteins such as histone recognition modules (HRM), acetyl-, and methyl-transferases. Acetylation and methylation of the N-terminal histone "tail" region on the ε-nitrogen on lysine residues has now been widely characterized, and shown to affect RNA synthesis (first discovered by Allfrey et al. 1964; Murray 1964; Phillips 1963).

Highly packed regions in chromatin, known as heterochromatin, are associated with a low concentration of genes that are expressed at low levels. The less tightly packed regions are termed euchromatin and have more genes that are associated with higher transcription activity (Armstrong 2014). It is important to note that these regions are not static. For instance, transitions between euchromatin and heterochromatin occur during the cell cycle and upon cellular differentiation (Armstrong 2014). However, regions that are compact in the early interphase in one cell cycle often retain this "compactness" in later cycles. The primary example of this is found in human females, where one of the two X chromosomes is tightly packed in the so-called called Barr body, remaining this way in subsequent cell divisions. This is one of many

mechanisms that require an additional layer of memory above that which is implemented by the DNA code.

If the full length of DNA from a typical human cell was stretched out, it would measure about 2 m. Keeping in mind the diameter of a typical human cell, about 5 µm – 100 µm (Nelson et al. 2008), this large amount of DNA requires an extensive degree of compaction in order to fit into such a small nucleus. In order to understand how this compaction is achieved, we have to look at chromatin structure. In 1973, it was suggested that histones assemble onto inactive chromatin as repeating subunits (Woodcock et al. 1976), a complex that was later called nucleosomes (Oudet et al. 1975). Extensive research has been done on nucleosomes, including solving the crystal structure at 2.8 Å resolution (Luger et al. 1997). While it is widely accepted that the DNA is compacted by associating with nucleosomes, the mechanism of further compaction and how this process is regulated is still to a large extent unknown. A structure of possible relevance to regulation and compaction, known as the 30 nm fibre from its characteristic dimensions, is readily observed in electron density micrographs (Everid et al. 1970). A cryogenic electron microscopy structure of 12 successive nucleosomes at 11 Å has been solved (*in vitro*) suggesting that the fibre can form (Song et al. 2014). However, this structure has not been observed with cryo-electron microscopy on frozen mitotic HeLa cells (Eltsov et al. 2008), which suggests that this fibre is not very abundant *in vivo.* A proposed explanation for this has been that it may only be present during active interphase (Maeshima et al. 2010).

It has also been suggested that the length of the linker DNA between histones may be the essential component that regulates the dynamic transition between highly compact and less compact DNA morphologies (Wong et al. 2007; Wu et al. 2007). However, in order to validate these theories, a better set of tools to study chromatin regulation has to be developed, and the functions of chromatin associated proteins have to be identified. Of particular interest is the nucleosome, its protein components and their post translational modifications (PTMs).

## 1.1 The nucleosome

In Nucleosomes, DNA is successively wrapped around the histone protein octamer as can be seen in Figure 1.1. Each nucleosome consists of eight histones, where the four core histones (H2A, H2B, H3 and H4), form heterotetramers, two of which assemble into a disc-like template that have a high affinity towards DNA. DNA is coiled 1.6 times around the core histones in a left-handed orientation, so that each nucleosome is associated with 147 base pairs. There is also a variable amount of DNA between each nucleosome, known as linker DNA. Nucleosome assembly is not specified by the DNA sequence (Prunell & Kornberg 1978) and most of the hydrogen bonds that stabilize the complex occur between the core histones and the phosphate backbone of DNA. The few hydrogen bonds between the DNA bases and the complex all occur within the sequence-unspecific minor groove. However, the GC content in the sequence affects the flexibility of DNA. This is relevant because of the high degree of bending that is required in order to wrap DNA around the histone octamer, making certain sequences more prone to be part of nucleosomes (Segal et al. 2006; Brogaard et al. 2012). The N-terminals of the four core histones, called histone tails, extend beyond the core itself, presenting their lysine- and arginine-rich tails to the nucleoplasm. These amino acids have an overall positive charge in their native state, and are known to be sticky towards DNA. Highlighting their importance, the amino acid sequences of histones are highly conserved in eukaryotes. However, the type and location of the histone tail modifications have high variations amongst species (Munks et al. 1991). Besides the core histones, an additional histone called H1 has been shown to mediate formation of chromatin fibres. When histone H1 is present, an additional 20 bp are coiled onto the core histones. How H1 affect chromatin organization is not fully understood, but it has been suggested that H1 could serve as the regulator of chromatin compaction by changing the length of linker DNA (Wong et al. 2007; Wu

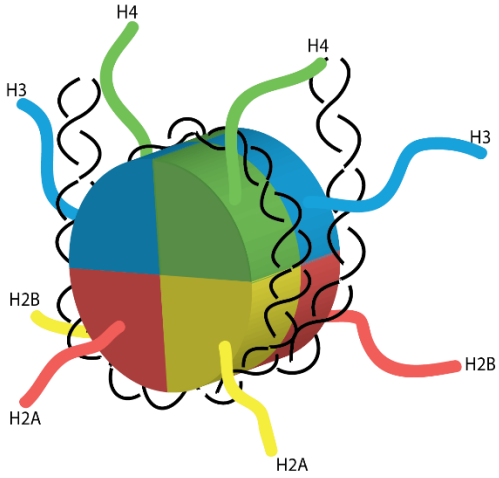


***Figure 1.1: Schematic view of the nucleosome.*** *DNA is wrapped around the histone octameric core complex with the H2A (yellow), H2B (red), H3 (blue) and H4 (green)) histone tails protruding out of the complex.*

et al. 2007). However, more experimental evidence is required in order to confirm this hypothesis.

Epigenetics has emerged as field in biology that views PTMs as an additional layer of heritable information. This offers an explanation on how the cell "remembers" the traits that cannot be explained by the information that is stored in the DNA sequence alone, such as cell differentiation (Olins & Olins 2003). There are many formal definitions of epigenetics, and in this thesis a broad definition is utilized.

''An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence'' (Berger et al. 2009).

Under this definition, stably heritable phenotypes are not necessarily trans-generational, but are at least conserved through several cell divisions. The information in epigenetics is believed to be stored mainly in methylation of DNA, and in PTMs on histone tails. The focus in this work has been on PTMs on histone tails, which are readily called the histone code.

## 1.2 The histone code

Since the discovery of methylation and acetylation, a plethora of additional PTMs have been discovered at the N-terminal tails of histones. The most prominent among these are phosphorylation, ubiquitylation and sumoylation. However, this is a progressively growing list (Kouzarides 2007; Tan et al. 2011).

Based on the amount and frequency of histone PTMs, researchers started to think of the large amount of information that these could potentially store (Turner 1993). Patterns of PTMs have been shown to be conserved during multiple cell divisions and it is thought to function as a means of epigenetic memory, often referred to as the histone code. The existence of such a code implies the presence of machinery capable of reading, writing, copying and editing the information it contains. Several writers and editors have been identified, including histone acetyl- and methyl- transferases. The first domain that was shown to bind specifically to modified histones was the bromodomain (Haynes et al. 1992), that bind to acetylated lysine 9

on H3 (Dhalluin et al. 1999). Furthermore, PHD fingers (Aasland et al. 1995; Schindler et al. 1993), where shown to bind to H3K4me3 (Shi et al. 2006; Li et al. 2006; Peña et al. 2006; Martin et al. 2006; Wysocka et al. 2006) (reviewed by Mellor 2006), and chromodomains (Paro & Hogness 1991) that bind to methylated histones (Bannister et al. 2001; Lachner et al. 2001). These bind to specific modifications of histone tails, and thereby fulfil the role of histone code readers. Multi-modular proteins with combinations of "readers" and "writers" can be characterized as copiers. Furthermore, proteins like histone deacetylases and demethylases, among others, can act as "erasers" and thus all the predicted components are present.

Modification of residues in the histone tail changes their chemical properties, such as shape and charge (Figure 1.2). Modified tails thereby become distinguishable from unmodified tails. In this work we focus mainly on methylation of lysine residues. Methylation of lysine introduces steric differences, the extent of which depends on if the modification occurs as a monomethylation (me1), dimethylation (me2), or trimethylation (me3). In contrast to acetylation, the positive charge of the ε nitrogen is preserved after methylation. However, the accessibility to this charge is reduced as methyl groups are added. In all methylation states, the nitrogen has four covalent bonds, none of which are chiral. Incremental addition of methyl groups on the ε nitrogen on lysine reduces its ability to form hydrogen bonds and increases the hydrophobicity and cation radius (reviewed by Taverna et al. 2007).

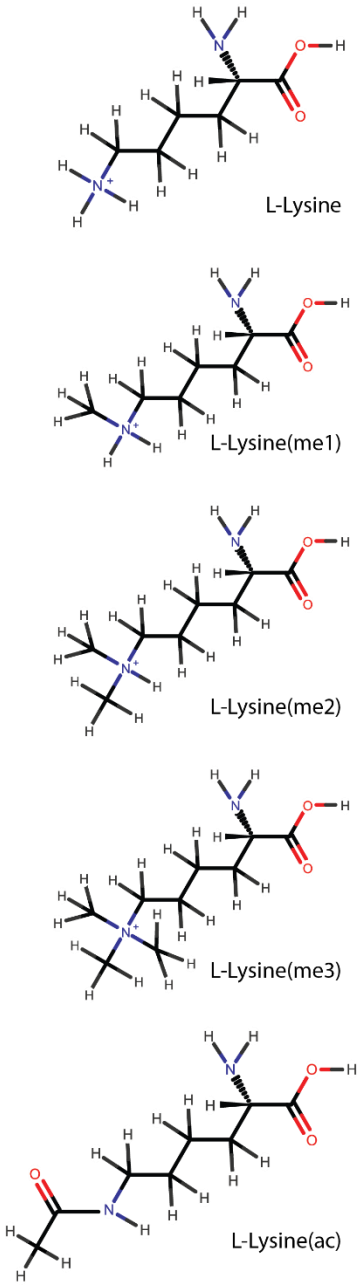


***Figure 1.2: Representation of modified lysine residues**. The monomethylted (me1), dimethylated (me2), trimethylated (me3), and acetylated (ac) states are depicted.*

As mentioned above, there are several proteins that bind to specific modifications of histone tails. These HRMs are typically found in multi-modular proteins, and multiple HRMs are often

present on the same protein (e.g. p300, UniprotID: Q09472). The HRM are further divided into a few structural folds that include the royal family, Bromodomains, 14-3-3 domains, WD40r domains, tandem BRCT domains, PHD fingers (Ruthenburg et al. 2007) and CW domains (Hoppmann et al. 2011).

As HRMs are part of multi-modular proteins they also direct their adjacent domains not only to chromatin, but also to specific sites in chromatin. This is at the core of what the histone code is, since the adaptors (such as SET domains, histone methyltransferases, or histone acetyltransferases) are directed to specific chromatin regions. By being associated with another HRM any of these adaptors could easily be directed to another region (Turner 2007).

## 1.3 The CW domain family

A group of histone recognition modules belonging to the CW domain family has been identified by means of their conserved cysteine and tryptophan consensus sequence (Perry & Zhao 2003). Several groups have reported that the members of the CW domain family have a preference for binding methylated histone H3K4, but differ in their methylation state preference (He et al. 2010; Hoppmann et al. 2011; Strømland 2012). Other domains that occur together with CW domains varies, but most belong to either writers, erasers or remodelers. In ASHH2 the CW domain is found in conjunction with AWS, SET and post SET domains (UniprotKB ID: Q2LAE1). The SET domain is a histone-lysine N-methyltransferase domain showing that the ASSH2 protein contains both a reader (CW), and a writer (SET).

In order to study the binding properties of the ASHH2 CW domain, several variants with varying sequence length were used (Figure 1.3). The longest protein is called CWs in this thesis, and

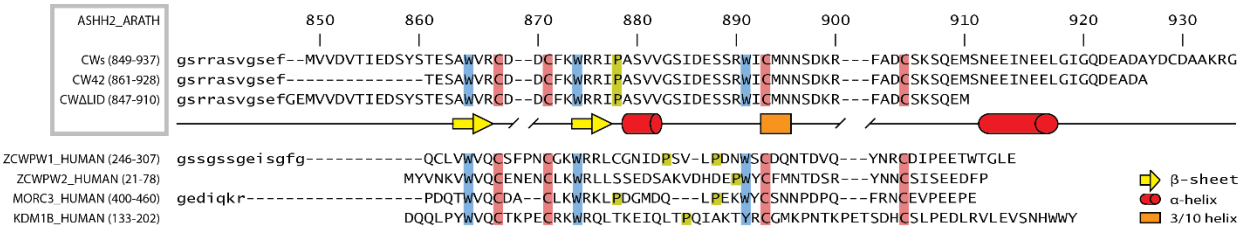


***Figure 1.3: Amino acid sequence of CW domains in this study.*** *ASHH2 CW domain variants CW42, CWs and CWΔLID compared to the amino acid sequence of ASHH2 are shown in the upper part of the figure. The secondary structure from 2L7P is shown. The sequences of three human CW domains are aligned and shown below the secondary structure of the ASHH2 CW domain. The residues included in the alignment are shown next to the accession numbers (left in the figure). Conserved W, C, and P are highlighted.*

comprise of the amino acids from 849 to 937[1]. A shorter version called CW42 861 to 928, are truncated at both termini, however all secondary structure elements are conserved. Furthermore, a version where the C-terminal helix is removed is called CWΔLID (847 – 910). Three human CW domains ZCWPW1, ZCWPW2, and MORC3 were used for comparison purposes. Their sequence is shown in the lower part of Figure 1.3. Furthermore several H3 ligands are used in this study. The ligand nomenclature is defined in Table 2.17.

At the time of writing, there are only 5 different CW domain family members where the structure is known (Protein Data Bank search: PF0856), and only ZCWPW1 is solved both with and without ligand (Table 1.1). The CW domain fold resembles the fold of PHD fingers and both have little secondary structure content. The secondary structure elements of the CW domains

***Table 1.1: PDB structures annotated as CW-type Zinc Fingers (PF07496)***

| UniProt ID | PDB structure ID without ligand | PDB structure ID with ligand | Binding preference ($K_d$) | $K_d$ reference |
|---|---|---|---|---|
| **MORC3_HUMAN** | - | 4QQ4 | H3K4me2 (3µM)<br>*H3K4me3 (8.5µM)<br>H3K4me1 (19µM) | (Strømland 2012) |
| **ZCWPW1_HUMAN** | 2E61 | 2RR4 | *H3K4me3(16.7µM)<br>H3K4me2(41.8µM)<br>H3K4me1(91.7µM)<br>H3K4me0(234.7µM) | (He et al. 2010) |
| **ZCWPW2_HUMAN** | - | 4O62 | *H3K4me3 | |
| **KDM1B_HUMAN** | 4FWE, 4FWF, 4FWJ, 4GU0, 4GU1, 4GUR, 4GUS, 4GUT, 4GUU, 4HSU. | - | - | - |
| **ASHH2_HUMAN** | 2L7P | This work** | **H3K4me1(1.15µM)<br>H3K4me2(2.1µM)<br>H3K4me3(4µM)<br>H3K4me0(~no bind) | (Hoppmann et al. 2011) |

**Indicates which histone tail peptide that was used to make the solution structure. $K_d$ values are based on published and unpublished binding data from our group and others* (Hoppmann et al. 2011; Strømland 2012; He et al. 2010). *** The ASHH2 structure with H3K4me1 has been initiated. It should be noted that KDM1B protein is solved with ligand, but not in the CW domain binding site.*

[1] In this thesis, amino acid numbering starts at the first amino acid in the particular protein variant, i.e. CWs, CW42 and CWΔLID. Figure 1.3 gives the relation between these residue numbers and the numbering used in the UniprotID identifier of the protein they were derived from (ASHH2_HUMAN).

consists of two small beta strands and two small alpha helixes / helix like structures. One tryptophan residue is present on each of these beta strands, and both are important for binding. Both CW domains and PHD fingers bind $Zn^{2+}$, PHD fingers bind two zinc ions, whereas CW domains can only bind one $Zn^{2+}$ ion, which is coordinated by four cysteine residues. More is known about PHD fingers, and at the time of writing there are structures for more than 50 different PHD domain architectures of PHD fingers (Protein Data Bank search: PF00628). Most of these PHD fingers also have two tryptophan residues on their beta sheet that are involved in binding. However, since this motif is conserved in both CW domains and PHD fingers and both can bind a variety of methylation states of H3 lysine residues (Some also bind me0 or non-histone proteins), thus another methylation state determining factor must be present. For PHD fingers the same fold with minor sequence changes is known to affect their binding preference (Yap & Zhou 2010). More interestingly, the methylation state determinants in CW domains remain elusive.

Clues concerning methylation state preference can be gathered from CW domain structures with ligand present, and to date the structures of ZCWPW1 (PDBID: 2E61 (He et al. 2010)) and (MORC3 PDBID: 4QQ4 (Liu n.d.) with H3K4me3 peptides bound has been solved (Figure 1.4 A and B respectively). In ZCWPW1 two of the methyl groups are oriented towards the two beta sheet tryptophan residues, whereas the third is protected from the hydrophilic solvent by a third tryptophan residue situated in the non-conserved C-terminal. In MORC3, the third methyl

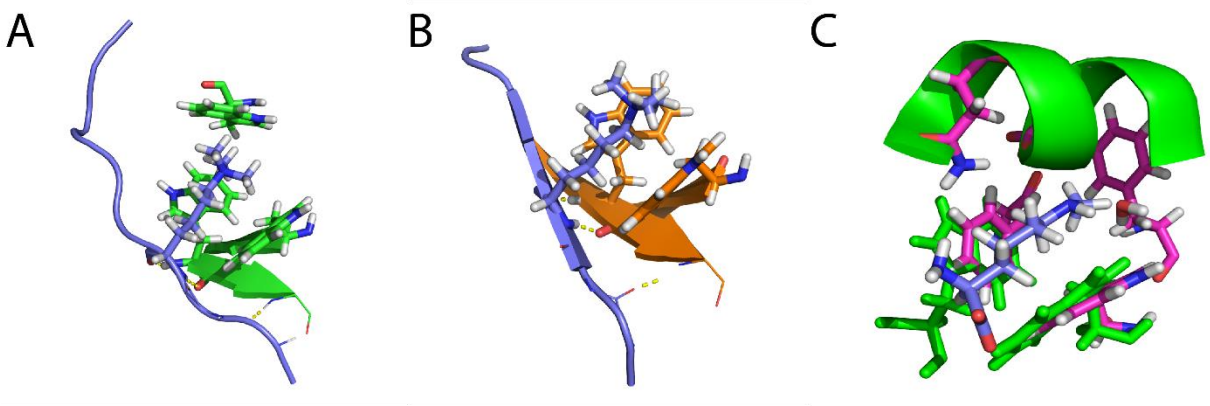


***Figure 1.4: Binding sites of methyllysine recognition modules***. *(A) CW domain in ZCWPW1 in complex with H3K4me3 peptide. (B) CW domain in MORC3 in complex with H3K4me3 peptide. Note that our $K_d$ show a stronger binding for H3K4me2 for this CW domain (Table 1.1). (C) ASHH2 CW (green) manually aligned to SCML2_HUMAN (magenta) in complex with methyllysine (blue). Note that SCML2 is not a structural homolog of the CW family, however the binding sites are similar.*

group on K4 is not enclosed by the C-terminal and the hydrophobic methyl group is thus exposed to bulk solvent. Our group has previously shown that H3K4me2 have the lowest $K_D$ towards this CW domain (Strømland 2012). This is interesting since lower methylation states have been suggested to be favoured by H3K4 recognition modules by means of a hydrogen bond acceptor situated on top of the two aromatic residues (Ruthenburg et al. 2007).

Binding to H3K4me1 is less documented than binding to e.g. H3K4me3, and upon searching PDBsum no entries with H3K4me1 in complex with a HRM was found. However, there are 60 structures of proteins in complex with monomethyllysine (including duplicates). These include several Malignant Brain Tumor-like domains (e.g. L3MBTL1_HUMAN, or SCML2_HUMAN). These monomethyllysine binding domains are not structural homologs of CW domains. Regardless, the binding site has a striking structural resemblance to CW domains as seen in the overlay in Figure 1.4 C. SCML2 has a monomethyllysine specificity with 20 fold higher $K_D$ for dimethyl lysine, and no observed binding for trimethyl lysine (Santiveri et al. 2008). A tryptophan and two phenylalanine residues provide a similar aromatic cage around the methyl group of the methyllysine. Further, both an aspartate residue and a water molecule form a hydrogen bond with the positively charged ε Nitrogen. The binding pocked is not much larger than the monomethyllysine and this suggests steric exclusion of higher methylation states, unless the binding site undergoes allosteric changes. Taken together these observations can been generalized and possibly explained by the following model for binding.

## 1.4 The "lid hypothesis" – a model for methylation state determinants

A given HRM must have several determinants of specificity in order to bind its ligand, in the presence of the different PTMs. Different CW domain family members have specificity towards different methylation states of H3K4, the molecular mechanisms governing this remains unknown. However, the C-terminal region is not conserved between these families and our group has proposed that this region determines methylation state specificity (Hoppmann et al. 2011). There are structures for only five CW domains (Table 1.1), something that makes the identification of a general mechanism difficult. However, by visual comparison of CW domain structures with PHD fingers, several of the described mechanisms for H3K4 selectivity in PHD fingers appear to be similar to the ones seen in CW domains. A conserved determinant of specificity between CW domains and PHD fingers is hydrogen bonding between the N terminal -NH3+ group of H3A1 and the domain. Furthermore, hydrogen bondage between various combinations of the side chains of R2, R3, N5, and R8 to the domain have been implicated in binding (Hoppmann et al. 2011; He et al. 2010; Liu n.d.). In addition, histone H3 tails have been shown to form an additional beta strand on top of the beta sheet already present in the domain structure, known as beta augmentation (Remaut & Waksman 2006). For the three CW domains structures with ligand, similar results are found (Figure 1.4 A, B; Table 1.1).

The lid model suggests that the C-termini contains major determinants for consolidation of binding (Hoppmann et al. 2011). For example, it can contain an aromatic residue as in ZCWPW1 (He et al. 2010), or a helix as described for ASHH2 (Hoppmann et al. 2011). Increased methylation of H3K4 may lead to increased contributions from van der Waal's interactions in the binding pocket from the tryptophan residues. In addition, increased methylation may lead to more optimal restriction (with respect to distance) and orientation (with respect to angle) of cation- $\pi$ interactions that are major determinants for specificity. This is similar to what has been described for PHD fingers (Ruthenburg et al. 2007), which highlights another aspect that is important for binding: the angle and distance between the tryptophan residues and the ligand.

Predictions of cation-π interaction angles and distances indicate that $NH_4^+$ cations situated 3 Å directly above the ring are more energetically favourable (Marshall et al. 2009).

ASHH2 CW is the only known domain with a preference for H3K4me1 and for which a structure has been determined. (Table 1.1), and this is solved without the ligand (Hoppmann et al. 2011). In order to obtain insight into the binding mechamism, the structure without ligand can be compared to other methyllysine binding modules. Figure 1.4 C shows the unbound structure of the ASHH2 CW domain compared to the structure of human SCML2. SCML2 has an affinity towards monomethyllysine with a similar binding site, but has no preference for histones compared to free methyllysine (Santiveri et al. 2008). There are several differences between these structures. Notable among these is the low amount of space between the two tryptophan residues and the C-terminal helix in ASHH2 that occupies the binding site. As shown in Figure 1.4 C there is more space for the monomethylated lysine in SCML2 compared to the CW domain, but it remains uncertain it there is enough space to fit the monomethylated lysine. Further, there is no charged residue in the C-terminal helix near the two tryptophan residues which could act as a proton acceptor. This raises the question as to why and how the ASHH2 domain can bind H3K4me1.

Our group has identified 6 mutants that disrupt binding. These include individual mutations of all the three tryptophan residues, Q908, E909, and the truncation of the whole C-terminal helix (Hoppmann et al. 2011). These mutants all show complete disruption of binding. This is interesting because removal of the C-terminal helix, in the absence of other evidence, could have been predicted to make the pocket more accessible to binding.

In order to further pin down which residues which are involved in binding of methylated histone tails, Hoppmann et al. 2011 investigated changes in residue-specific chemical shifts upon binding using NMR. Prominent changes occurred at R30 and R53, in what was termed the central variable region and in the C-terminal helix upon binding to H3K4me1 peptide. These changes in the central region, and C-terminal region were suggested to be a result of conformational changes upon binding. On the other hand, the shifts close to the tryptophan residues were suggested to be caused by a more direct change in molecular environment due to binding (Hoppmann et al. 2011).

Chimeric CW domains, where the C-terminal ends of the ASHH2 and the ZCWPW1 CW domain were swapped, have been made and studied (Holmedal 2013). Although all chimeras bind the methylated H3K4 peptides with very low affinity, ASHH2 CW with the C-terminal region from ZCWPW1 shows the highest affinity towards H3K4me2, More interestingly, ZCWPW1 with the C-terminal helix from ASHH2, shows the highest affinity towards H3K4me1. This shows that there is a correlation with the sequence of the C-terminal of the domains and the domain specificity. Holmedal proposed that, in order to bind H3K4me3, a third aromatic residue is required and that binding to H4K4me2 and H3K4me1 requires a proton hydrogen bond acceptor in the C-terminal region (Holmedal 2013).

An assumption in this lid hypothesis is that the C-terminus is sufficiently flexible to facilitate the incorporation of the ligand. This mobility is clearly present for ZCWPW1 and MORC3, where the C-terminal is unstructured. However for ASHH2 the C-terminal is structured and situated in the binding pocket (Hoppmann et al. 2011), suggesting lower mobility. Holmedal further developed the lid model and proposed a model of binding shown in Figure 1.5 (Holmedal 2013). In the model the C-terminal helix of the ASHH2 CW domain function as a lid that has to be opened to allow ligand binding. Upon binding, the C-terminal lid contributes to the ligand specificity and the central domain contributes to β-sheet formation between itself and the ligand, as well as cation-π interactions between the two tryptophan residues and the ligand's lysine. This development of the model leads to several interesting questions, such as which residues function as a proton acceptor, and whether the C-terminal can move as proposed by this model.

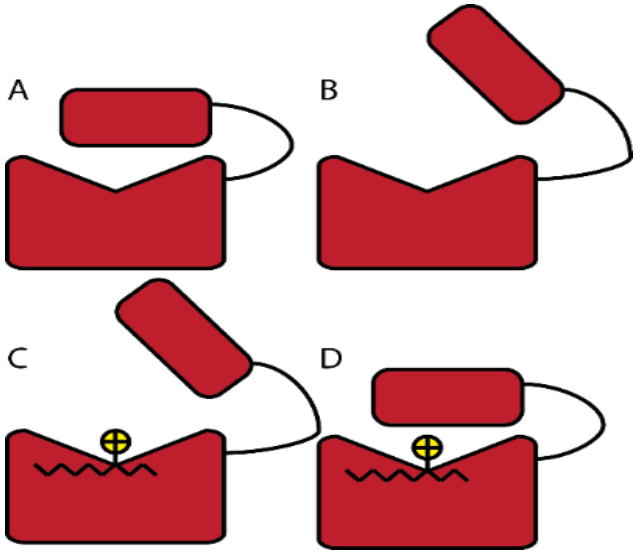


***Figure 1.5: Model for ligand binding of the CW domain.*** *A: Closed, No ligand. B: Open No ligand. C: Open, with ligand. D: Closed with ligand*

## 1.5 Aim and strategy of the study

The aim of this study was to look further into the ligand binding mechanisms for the ASHH2 CW domain. More specifically, to measure the molecular movement of the C-terminal α-helix upon binding to its ligand and thus test the lid-hypothesis outlined above.

To accomplish this, a strategy that involves the use of NMR relaxation data and heteronuclear NOEs to observe the molecular movements in the presence and absence of peptide binding was chosen. Additionally, the process of determining the solution structure of the domain bound to the ligand was initiated.

In this study, several variants and mutants of ASHH2 were examined. To ensure that the proteins where comparatively stable and to find out how ligand binding affected the domains, the stability of the protein following exposure to different denaturants was examined.

Several of the first 8 amino acids in the N-terminal tail of histone H3 are tightly associated with the CW domain upon binding. Experimental data from the collaborating lab of prof. Reidunn B. Aalen, University of Oslo, had indicated that K9 is also involved in binding to the ASHH2 CW domain. To find out if K9 affect binding, combinations of H3K4me1 and K9 modifications are of interest. Lastly, to further elucidate the contribution of K9, isothermal titration calorimetry (ITC) was used on histone-mimicking peptides to find their dissociation constants.

# 2 Materials

*Table 2.1: Chemicals*

| Chemical name | Abbreviation / Formula | Supplier |
|---|---|---|
| Tris(hydroxymethyl)aminomethane | Tris | Sigma |
| 2,2',2'',2'''-(ethane-1,2-diyldinitrilo)tetraacetic acid | EDTA | Merck |
| 2-Sulfanylethan-1-ol | β-ME | Merck |
| Acrylamide/bis-acrylamide | - | BioRad |
| Ammonium persulfate | APS | Merck |
| Ampicillin | Amp | Vitus Apotek |
| Agar-agar | - | Merck |
| Bromphenol Blue | BPB | Merck |
| Calcium chloride | CaCl | Sigma |
| Disodium hydrogen phosphate | $Na_2HPO_4 * 2\ H_2O$ | Sigma |
| 96 % Ethanol | EtOH | Kemetyl |
| Absolute ethanol | EtOH | Kemetyl |
| isopropanol | - | Kemetyl |
| Glycerol | - | Merck |
| Glycine | - | Merck |
| Hydrochloric acid | HCL | Sigma |
| Isopropyl-β-D-thiogalactoside | IPTG | Promega |
| Magnesium chloride hexahydrate | $MgCl_2 * 6H_2O$ | Merck |
| N,N,N',N'-Tetramethyl ethylenediamine | TEMED | BioRad |
| Methanol | - | VWR |
| Peptone | - | Merck |
| Polyoxyethylene (20) sorbitan monolaurate | Tween 20 | Merck |
| Potassium chloride | KCl | Merck |
| Potassium dihydrogen phosphate | $KH_2PO_4$ | Merck |
| Propan-2-ol | - | Roth |
| Sodium chloride | NaCl | Merck |

| | | |
|---|---|---|
| Triton X-100 | - | Sigma |
| Tris(2-carboxyetyl)phosphine | TCEP | Fisher Scientific |
| Zinc acetate | ZnAc | Merck |
| Biotinylated thrombin | - | Novagen |
| EDTA free Protease inhibitor | - | Roche Diagnostics |
| Streptavidin Sepharose | - | GE Healthcare |
| Potassium hydrogen phosphate | $K_2HPO_4$ | Merck |
| Dithiothreitol | DTT | Sigma |
| Sodium dodecyl sulfate | SDS | BioRad |
| Sodium Azide | $NaN_3$ | Sigma |
| Deuterium | $^2H_2O$ | Sigma |
| Biotin | - | Sigma |
| Folic acid | - | Sigma |
| para-aminobenzoic acid | PABA | Sigma |
| riboflavin | - | Sigma |
| pantothenic acid | | Sigma |
| pyridoxine HCl | - | Sigma |
| Thiamine HCl | - | Sigma |
| Nicotinamide | - | Sigma |
| Iron(II)Chloride | $FeCl_2$ | Sigma |
| $K_2SO_4$ | - | Sigma |
| Ammonium chloride | $NH_4Cl$ | Sigma |
| Ammonium chloride ($^{15}$N labelled) | - | Cambridge Isotope Lab. |
| Glucose monohydrate | - | Sigma |
| Glucose ($^{13}$C labelled) | - | Isotec |

*Table 2.2: SDS gel (2x 1.5mm)*

| Chemicals | 12 % | 20 % | Stacking gel | Units |
|---|---|---|---|---|
| **MilliQ water** | 4.98 | 0.98 | 4.54 | ml |
| **30 % Acrylamide-Bis 37.5:1** | 6 | 10 | 1.3 | ml |
| **1,5 M Tris pH 8.8** | 3.8 | 3.8 | | ml |
| **0,5 M Tris pH 6.8** | | | 2 | ml |
| **20 % SDS** | 75 | 75 | 40 | µl |
| **10 % APS** | 150 | 150 | 80 | µl |
| **TEMED** | 6 | 6 | 8 | µl |

*Table 2.3: 5x Protein Loading buffer*

| Chemicals | 10 ml | 20 ml | Units |
|---|---|---|---|
| **20 % SDS** | 2.5 | 5 | ml |
| **87 % glycerol** | 2.3 | 4.6 | ml |
| **1 M DTT** | 0.25 | 0.5 | ml |
| **1 % Bromphenol blue** | 20 | 40 | µl |
| **0.5 M Tris-HCL pH 6.8** | 1.3 | 2.6 | ml |
| **MilliQ water** | 3.85 | 7.7 | ml |

*Table 2.4: Coomassie staining solution*

| Chemicals | 0.5 l | Units |
|---|---|---|
| **Coomassie blue (R-250)** | 0.5 | g |
| **96 % Ethanol** | 200 | ml |
| **100 % Acetic acid** | 50 | ml |
| **MilliQ water** | 250 | ml |

*Table 2.5: Coomassie destaining solution*

| Chemicals | 1 l | Units |
|---|---|---|
| **96 % Ethanol** | 50 | ml |
| **100 % Acetic acid** | 70 | ml |
| **MilliQ water** | 880 | ml |

*Table 2.6: NMR buffer*

| Chemicals | 1 l | Units |
|---|---|---|
| **1 M $K_2HPO_4$** | 7.62 | ml |
| **1 M $KH_2PO_4$** | 12.38 | ml |
| **5 M NaCl** | 10 | ml |
| **1 M DTT** | 1 | ml |
| **MilliQ water** | 970 | ml |

*This result in a final concentration of 20 mM Phosphate buffer ph 6.4, 50 mM NaCl, and 1 mM DTT*

*Table 2.7: T7 buffer*

| Chemicals | 50 ml | 500 ml | Units |
|---|---|---|---|
| **1 M Tris-HCL pH 7.0** | 1.25 | 12.5 | ml |
| **5 M NaCl** | 1.5 | 15 | ml |
| **1 M TCEP*** | 50 | 500 | µl |
| **MilliQ water** | 47.2 | 472 | ml |

**0.5 g TCEP was solved in 0.5 ml 10 N NaOH and stored at 4°C*

*Table 2.8: 5x TZNK buffer*

| Chemicals | 2 l | Units |
|---|---|---|
| **1 M Tris-HCl pH 8.5** | 100 | ml |
| **5 M NaCl** | 4.8 | ml |
| **0.5 M ZnAc** | 400 | µl |
| **3 M KCl** | 100 | ml |
| **1 M $MgCl_2$** | 4 | ml |
| **MilliQ water** | 1790 | ml |

*Autoclaved before use and stored at 4°C*

*Table 2.9: 1x TZNKβT buffer*

| Chemicals | 1 l | Units |
|---|---|---|
| **1x TZNK buffer** | 1000 | ml |
| **100 % β-mercaptoethanol** | 0.7 | µl |
| **20 % Triton-X 100** | 5 | ml |

*Prepared and used fresh*

*Table 2.10: 2xYT media*

| Chemicals | 0.9 l | Units |
|---|---|---|
| **Tryptone** | 16 | g |
| **Yeast extract** | 10 | g |
| **NaCl** | 5 | g |
| **MilliQ water** | 900 | ml |

*Autoclaved before use*

*Table 2.11: 2xYTG media*

| Chemicals | 1 l | Units |
|---|---|---|
| **2xYT media** | 900 | ml |
| **20 % glucose** | 100 | ml |

*Preheated to room temperature when it was stored at 4°C*

*Table 2.12: Plasmids*

| Name | Origin |
|---|---|
| **pSXG** | (Ragvin et al. 2004) |
| **pSXG-ASHH2-CWs** | (Hoppmann et al. 2011) |
| **pSXG-ASHH2-CW42** | Rein Aasland |
| **pSXG-ASHH2-CWΔLID** | (Hoppmann et al. 2011) |

*Table 2.13: Bacterial strains*

| Name | Supplier |
|---|---|
| ***Escherichia coli* BL21-CodonPlus®-RIL** | Invitrogen |

*Table 2.14: Instruments*

| Name | Purpose | Manufacturer |
|---|---|---|
| **GelDoc™ EZ imager** | SDS-PAGE imaging | BioRad |
| **Nano ITC low volume** | Isothermal titration calorimetry | TA instruments |
| **NanoDrop ND-1000™** | Determination of Concentration | Saveen-Werner |
| **AV600 , with a QCI CryoProbe** | Nuclear Magnetic Resonance | Bruker BioSpin |
| **Äkta explorer with a Superdex 75 16/60 HiLoad Prepgrade column** | Protein purification | GE Healthcare |
| **LS 50B Fluorescence Spectrometer** | fluorescence spectrometry | Perkin Elmer |
| **French Press FA078 with a 40K sample cell** | Disrupting cells | Spectronic Instruments |
| **F12-ED refrigerated/heating circulator** | Control temperature in Fluorescence measurements | Julabo |

*Table 2.15: Web services*

| Name | Purpose | url | Reference |
|---|---|---|---|
| **ProtParam** | Predict protein parameters | http://web.expasy.org/protparam/ | (Gasteiger et al. 2005) |
| **The Protein Data Bank** | Search for Protein structures | http://www.rcsb.org/pdb/ | |
| **Eukaryotic linear motif resource ELM** | Guide to biology and function of linear motifs | http://elm.eu.org/ | (Dinkel et al. 2014) |
| **PDBsum** | Overview of PDB structures | http://www.ebi.ac.uk/pdbsum/ | (de Beer et al. 2014) |
| **FATCAT** | Structural alignment | http://fatcat.burnham.org/fatcat-cgi/cgi/fatcat.pl?-func=pairwise | (Ye & Godzik 2004) |
| **BMRB - Biological Magnetic Resonance Data Bank** | Obtain published Chemical shifts | http://www.bmrb.wisc.edu/ | (Ulrich et al. 2008) |
| **BRENDA** | Finding annotation of Enzymatic activity of proteins | http://www.brenda-enzymes.info/ | (Schomburg et al. 2004) |
| **Protein NMR** | Concepts of NMR assignments | http://www.protein-nmr.org.uk | (Higman 2012) |
| **MarvinSketch** | drawing, simple chemical structures | http://www.chemaxon.com/marvin/sketch/index.php | ChemAxon Ltd |
| **POSA (Partial Order Structure Alignment)** | Multiple structure alignment | http://posa.sanfordburnham.org/ | POSA (Ye & Godzik 2005) |
| **SPHERE** | Prediction of water exchange | http://www.fccc.edu/research/labs/roder/sphere/ | (Zhang 1995) |

*Table 2.16: Computer programs*

| Name and version | Purpose | Supplier |
|---|---|---|
| **TopSpin 2.1** | Conducting NMR experiments | Bruker BioSpin |
| **TopSpin 3.1** | Processing of NMR data | Bruker BioSpin |
| **Protein dynamics centre 2.2.1** | Analysing NMR Relaxation data | Bruker BioSpin |
| **SigmaPlot 12.5** | Curve fitting and data presentation | Systat Software |
| **Prism 6** | Statistical analysis and presentation | GraphPad Software |
| **Bruk2ucsfGui** | User interface for the sparky bruk2ucsf script | Made in this work |
| **Illustrator CC 15.0.0** | Making cartoons, and optimizing images and graphs | Adobe Systems Incorporated |
| **Sparky 3.114** | Visualization of NMR Data | T. D. Goddard and D. G. Kneller, University of California, SF |
| **Cara 1.9.0** | Assigning NMR data | http://cara.nmr.ch/ |
| **Visual Studio Express 2013** | Writing small programs and scripts | Microsoft Corporation |
| **SnapGene Viewer 2.4.2** | Visualize plasmid DNA sequences | SnapGene |
| **Mendeley Desktop 1.12.2** | Organize articles, books, and References | Mendeley Ltd |
| **Jalview 2.8.1** | Sequence alignment | (Waterhouse et al. 2009) |
| **PyMOL 1.3** | Molecular Graphics System | Schrödinger, LLC |
| **NanoAnalyze 2.4.1** | Analysis of ITC data | TA Instruments |
| **Screen Scales** | Measure angles and distances on the screen | Talon-Designs LLP |

*Table 2.17: Histone mimicking peptides*

| Name | Description | Sequence | Manufacturer |
|---|---|---|---|
| **H3** | Unmodified human Histone 3 mimicking peptide from A1-K9 with an additional tyrosine in the 10$^{th}$ position** | ARTKQTARY | Dr. W. J. Mawby* |
| **H3K4me1** | Modified human Histone 3 mimicking peptide from A1-K9 with an additional tyrosine in the 10$^{th}$ position, where the fourth lysine has a monomethyl group. | ARTK(me1)QTARY | LifeTein LLC |
| **H3K4me1 K9ac** | Modified human Histone 3 mimicking peptide from A1-G12 with an additional tyrosine in the 13$^{th}$ position, where the fourth lysine has a monomethyl group, and the ninth residue is acetylated. | ARTK(me1)QTARK(ac)STGY | LifeTein LLC |
| **H3K4me1 K9me2** | Modified human Histone 3 mimicking peptide from A1-G12 with an additional tyrosine in the 13$^{th}$ position, where the fourth lysine has a monomethyl group, and the ninth residue is dimethylated. | ARTK(me1)QTARK(me2)STGY | LifeTein LLC |

**Dr. W. J. Mawby, Department of Biochemistry, University of Bristol, UK, courtesy of Dr. S. J. Gamblin, NIMR, London, UK. Peptides from LifeTein LCC have a purity of 95%.** Tyrosine was introduced for concentration determination by nanodrop, and an extinction coefficient of 1200 was used.*

*Table 2.18: Consumables*

| Name | Use | Supplier |
|---|---|---|
| **Amicon Ultra-4 3000 MWCO Centrifugal filter devices** | Protein concentration | Millipore |

# 3 Methods

## 3.1 Protein expression

CW domains were expressed as GST fusion proteins from a single BL21-CodonPlus (DE3)-RIL chemicompetent cells. These cells were transformed by heat shock, where 200 ng of plasmid was put into 50 µl thawed cells and incubated for 30 minutes on ice. A 30 second heat shock at 42 °C was subsequently applied followed by 2 minutes on ice. 500 µl SOC medium was added and the cells were incubated at 37 °C/250 rpm for 1 hour. 50 µl – 100 µl of culture was plated onto LB plates with 34 µg/ml chloramphenicol and 100 µg/ml ampicillin and incubated for 10-15 hours at 37 °C. Plates were kept on 4 °C until one colony was picked and grown into a pre-culture.

Proteins used in ITC, denaturation and natural abundance NMR were purified from culture. 1000 µl pre-culture was added in 1 L 2xYTG media containing 34 µg/ml chloramphenicol and 100 µg/ml ampicillin. This was grown until it reached an $OD_{(A280)}$ of 0.6, at 37 °C/250 rpm, when it was supplemented with 400 µM IPTG, and 200 µM ZnAc, and put on 26 °C/250 rpm for 2 hours.

Proteins with uniform labelling of $^{15}N$ and/or $^{13}C$, were grown by adding 200 µl pre-culture into 25 mL M9 medium, supplemented with the preferred isotopes (1g 15N $NH_4Cl$ or 2g $^{13}C$ glucose), 34 µg/ml chloramphenicol and 100 µg/ml ampicillin. After growing this M9 pre-culture for 10-15 hours, it was inoculated in 1L M9 medium with the same supplements and grown at 37 °C/250 rpm until an $OD_{(A280)}$ of 0.6 was reached. At this point 400 µM IPTG and 200 µM ZnAc was added and the culture was incubated at 18 °C for 17 hours.

After protein expression the cells were harvested by centrifugation at 7500 xg in 1 l Beckman spin tubes and stored on -20 °C until use.

## 3.2 Protein purification

Harvested cells were thawed and resuspended in 10 mL 1x TZNKβT buffer / 1 g pellet. 1x complete, EDTA free protease inhibitor from Roche was added. Cells were disrupted two times with a French press at 1000 psi and spun down at 48500 xg in 30 ml corex tubes. The supernatant was run twice through a 10 ml column containing 5 ml 1x TZNKβT equilibrated glutathione sepharose 4B from GE healthcare. The column was washed with 15 ml 1x TZNKβT. The slurry was resuspended in 5 ml 1x TZNKβT and 7.16 U biotinylated thrombin was added and incubated at room temperature overnight. The cleaved protein was collected by washing first with 7.5 ml TZNKβT, then in 1 ml iterations until absorbance at 280 nm by NanoDrop was below 0.01 mg/ml. 50 µl streptavidin sepharose high performance was added to the sample, which was incubated at room temperature for 30 minutes and spun down at 1000 xg for 1 min. The sample was concentrated using Amicon ultra centrifugal filters with a 3 kDa cut-off until a volume of 2 ml was reached. Protein was further purified and exchanged into a buffer that was suitable for downstream applications using a GE Äkta system with a Superdex 75 16/60 HiLoad Prepgrade column. Following size exclusion chromatography, the desired fractions were, when required, concentrated using Amicon Centrifugal filter units with a 3 kDa cut-off. Purified proteins were stored at 4 °C and prior to experiments they were spun down for 17000 xg for 20 minutes to remove precipitate, if any. Protein concentration was measured using NanoDrop after this step.

## 3.3 ITC

Isothermal titration calorimetry (ITC) was used to determine the binding constants of CWs towards histone tail peptides carrying different combinations of PTMs. For all measurements, the temperature and stirring rate were kept constant at 25 °C and 300 rpm, respectively. The CWs concentration was 20 µM and the enthalpy of binding was determined by stepwise titration of 200 µM histone peptide. For each experiment, 2.03 µL of peptide was injected 22 times and the experiment was performed in triplicates. Both protein and peptide were dissolved

in T7 buffer and the heat of peptide dilution into T7 buffer was subtracted from the measurement using the average of 22 successive titrations.

Binding parameters were determined from the integrated peak areas with independent modelling, using the NanoAnalyze V 2.4.1 software.

## 3.4 Estimation of area in Cartesian coordinate systems

To determine the size of the binding pocket in ASHH2 and compare it to the binding pocket in SCML2 the area between tree selected atoms where calculated using Equation 3.1.

***Equation 3.1***

$$A = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

## 3.5 Tryptophan angle measurements

To determine the angles between the tryptophan residues in different CW domains, a multiple structure alignment was performed using the POSA web tool (Ye & Godzik 2005). After this, the structures where imported to PyMOL and the tryptophan residues were coloured black and shown as “lines”. The measurements were performed with Screen Scales from Talon-Designs LLP

## 3.6 Denaturation

Intrinsic tryptophan fluorescence spectrometry was used to monitor protein unfolding upon heat treatment or zinc depletion by EDTA. Prior to each experiment, the samples were spun down at 17000 xg for 30 minutes to ensure that any precipitates were removed. Protein concentration was measured after this step. Measurements were performed in a 100 µL quartz cuvette. The three tryptophan residues in the CW domain were specifically excited with a 295 nm ±2.5 nm light beam in a LS 50B Fluorescence Spectrometer. Fluorescence was observed in

the spectral range between 310 and 450 nm with a scanning rate of 100 nm/min. Unless otherwise stated, each measurement was repeated 5 times on each sample, and each sample was measured in triplicates with and without 5 times molar excess of H3K4me1 peptide. Also, at least two independent preparations of protein samples were used. This was done to reduce the risk of studying proteins with that by chance had obtained a random mutation in one colony. Fluorescence data measured from the blank (T7 buffer at 4 °C) was subtracted, and the fluorescence intensity at 335 nm (folded) and 355 nm (Tryptophan in solution) was plotted against denaturant (e.g. Temperature or [peptide]/[EDTA]). This is referred to as the fluorescence ratio I335/I355. A global curve fit between the parallels to a sigmoidal curve with four variables, shown in Equation 3.2, was conducted to obtain the critical point D for all fluorescence ratio graphs.

Curve fitting to Equation 3.2 was performed with SigmaPlot v12.5, followed by a one-way ANOVA statistical analysis with a p value of 0.05. This was followed by Tukey's multiple comparisons test, where significant differences between the means were identified. Prism 6.05 for Windows (Trial Edition) was used for all statistical analysis.

***Equation 3.2:***

$$f(x) = y_0 + \frac{y_f}{1+10^{\left(-\frac{EC50-X}{b}\right)}}$$

*Where F(x) is a function of temperature, $y_0$ is the slope of the pre-transition state, $y_f$ is the slope of the post transition state, EC50 is the inflection point in the curve, and b is the slope of the transition-state.*

## 3.6.1 Thermal denaturation

The melting point of the different CW domain proteins, defined as the temperature at the inflection point of I355/I355 during heat denaturation, was determined with and without bound ligand. Each sample was subjected to a temperature gradient ranging from 4 °C to 90 °C, using a. Each 5 °C – 10 °C increase in temperature was allowed 5 minutes to come to thermal equilibrium before measurements were taken. The cuvette was equipped with a lid, to prevent evaporation from the sample.

In the first parallels, a low protein concentration of about 4-6 µM was used. However, In order to reduce noise, a higher protein concentration of about 15 µM – 18 µM was used in remaining experiments.

## 3.6.2 Effect of $Zn^{2+}$ depletion

EDTA binds $Zn^{2+}$ extremely tightly and has a $K_D$ of about $10^{-16}$ M, which is about $10^5$ to $10^7$ M stronger than a typical zinc-binding motif (Nyborg & Peersen 2004), and was thus used to deplete CWs and CW42 of their $Zn^{2+}$ ions. The concentration of CWs was ~15 µM and for CW42 it was ~23 µM for all EDTA experiments. These experiments were conducted at 25 °C, and every time something was added to the sample, a 30 min equilibration time was allowed before measurements were taken.

To pinpoint the amount of EDTA required for complete denaturation by $Zn^{2+}$ depletion, CW42 was subjected to EDTA concentrations ranging from 0 to 600 µM. The spectrum of the T7 buffer was subtracted from the sample spectrums and the fluorescence ratio was plotted against [EDTA]/[CW42] and fitted to equation 3.2. In the EDTA experiments triplicates were made, however they were all from the same protein batch. Complete denaturation was determined by visual inspection of the graph.

To test if the $Zn^{2+}$ depletion from CW domains was reversible, CWs and CW42 were treated with the concentration of EDTA that was found to introduce complete denaturation, followed by

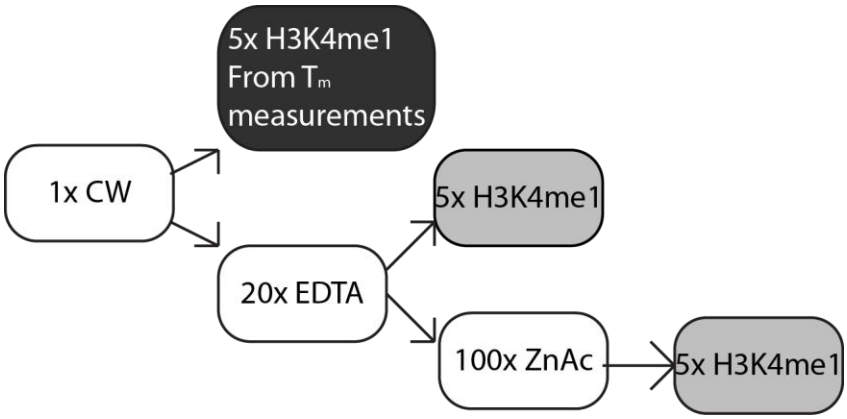


***Figure 3.1: Experimental setup of EDTA reconstitution experiment.*** *Each box represents a measurement, and each line represents the addition of either excess of EDTA, H3K4me1 or ZnAc. The excess relative to the starting CW molarity is denoted by either 5x, 20x or 100x. White boxes are measurements conducted without ligand, and grey boxes have a 5 times molar excess of H3K4me1 ligand present.*

addition of 5 times molar excess $Zn(C_2H_3O_2)_2$ (ZnAc) relative to EDTA. A sample was taken from all conditions and supplemented with 5 times molar excess of H3K4me1 and compared to the I335/I355 ratio of free CW and bound CW. A flowchart of the experimental setup is depicted in Figure 3.1.

# 3.7 NMR experiments

## 3.7.1 NMR sample, and experimental setup

Nuclear magnetic resonance (NMR) was used to evaluate and compare bound and unbound CW domain proteins, to measure the $T_1$ and $T_2$ relaxation behaviour of CWs, and solve the preliminary structure of CW42 in its peptide bound state. Unless otherwise stated, all NMR experiments were conducted with a protein concentration of 1 mM – 1.5 mM, 5% deuterium, 20 mM phosphate buffer pH 6.4, 50 mM NaCl, and 1 mM DTT. The samples that were used in relaxation experiments and the one used to make the solution structure, also had 1 mM NaAzide to prevent bacterial growth. For data-acquisition with the H3K4me1 peptide described in the materials section the sample was transferred from the shigemi tube into 2 mg peptide, resulting in a 6-fold excess of unlabelled peptide.

The NMR experiments were conducted on a Bruker Biospin AV600 NMR instrument, fitted with a QCI CryoProbe for sensitivity enhancement. Experimental setup was conducted in TopSpin 2.1. Samples were contained in BMS-003 Shigemi tubes sealed with parafilm and the temperature in the instrument was set to 25 °C. $^{15}N$ labelled samples were sufficient for relaxation studies whereas both 15N and 13C isotopes were used to obtain the spectrums for the NMR solution structure.

Prior to all NMR experiments, manual tuning and matching was performed on the $^{1}H$, $^{13}C$ (only with $^{13}C$ labelled protein) and $^{15}N$ channel. The lock phase was centred and phased manually and automatic 1D, and 3D gradshim protocols were performed. Furthermore, the 90 degree

proton pulse (typically 9.25 µs) was found by identifying the diminished 360 degree pulse (typically 37 ms) in a 1D proton experiment (pulse program: zg).

Samples prepared with natural abundance of $^{15}N$ and $^{13}C$ were used to test expression, purification, and optimize instrument parameters. In these experiments concentrations of 1.5 mM CWs and 0.6 mM CW42 were used. A higher resolution fingerprint was obtained by uniform $^{15}N$ labelling, used as a reference hub when comparing 3D experiments and to evaluate backbone amide changes during 3D experiments and between experiments.

### 3.7.2 NMR experiment strategy

In order to perform analysis on the chemical shifts of the CW domains the protein backbone had to be assigned. NMR data for CWs without ligand has previously been published (Hoppmann et al. 2011; Ulrich et al. 2008) and was used as a template for all assignments for CWs and CW42 with and without H4K4me1 peptide presence. For CWs without ligand, the backbone assignment was verified with a 3D TOCSY. CW42 without ligand was verified using the $\alpha$-carbons chemical shifts from the HNCA spectrum. Furthermore, for CW42 with ligand both HNCA and CBCANH experiments were performed. In general, the sequential assignment strategy was used by identifying amino-acid type using $\alpha$-carbon and $\beta$-carbon shifts and linking the residues through backbone correlations determined by the standard suite of triple-resonance experiments for backbone assignments (Palmer III et al. 2005)

Relaxation data and heteronuclear NOEs (hNOE) were obtained by varying the mixing time according to Table 3.1.

*Table 3.1: Mixing times in the false dimensions in pseudo 3D $T_1$, and $T_2$ experiments*

| **Experiment number** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **$T_1$ mixing time (s)** | 0.010 | 0.093 | 0.280 | 0.017 | 1.500 | 0.053 | 0.030 | 0.860 | 0.160 | 0.490 |
| **$T_2$ mixing time (s)** | 0 | 0.1232 | 0.0704 | 0.0352 | 0.088 | 0.1584 | 0.0176 | 0.0352 | 0.0176 | 0.0528 |

## 3.7.3 Analysis of NMR data

Prior to Fourier transformation, NMR data was subjected to apodisation using squared cosine window functions and zero-filling. Additionally, the real time-domain ($^{1}$H-channel) and artificial time domain(s) ($^{13}$C, $^{15}$N) were subject to a low-frequency filter to weaken the water-signal and linear prediction up to twice the experimental points, respectively. This data-processing was performed with the Bruker Biospin software v3.2. For relaxation data and hNOEs extensive (8K) zero filling was performed. Data was converted from Bruker "xrr" file format to ".ucsf" using a homemade user interface for the bruk2ucsf script that is embedded in the Sparky 3.114 NMR analysis program (Goddard & Kneller 2007). Sequence specific backbone and side chain assignment was performed in Cara Release: 1.8.4.2 (Rochus 2004). Images of the spectrums were made using Sparky 3.114 (Goddard & Kneller 2007).

The dynamics of CWs in the presence and absence of H3K4me1 was evaluated in Protein Dynamics Centre 2.2.1 from Bruker with default values according to the protein dynamics manual provided in the program. The intertwined Heteronuclear NOE experiment was divided into "NOE" and "NONOE" spectrums using the Topspin 3.2 AU program "split", before opening them in the dynamics centre.

## Using the protein Dynamics Centre software

Relaxation data was analysed according to the Dynamics Centre v 2.2.1s (DC) user manual chapter 7 and 8. NH-bond length was set to 1.02 Å, and a chemical shift anisotropy of -160.0 ppm was assumed. 1000 successive iterations with random selections of start parameters were performed several times to confirm that the same values were obtained following each run. For $T_1$ and $T_2$ the "errors were estimated by weighted fit". hNOE errors are automatically derived from an error propagation calculation based on peak intensity ratios. In order to calculate a representative global isotropic correlation time $\tau_c$, only residues with hNOE values below 0.7, $T_1$ values with less than 1 standard deviation and $T_2$ with less than 3 standard deviations were included. Calculation of the $D_{\parallel}/D_{\perp}$ ratio (parallel/perpendicular NH-vectors) was performed and proteins with $D_{\parallel}/D_{\perp} < 1.2$ were considered isotropic, whereas proteins with a larger value than 1.5 where assumed to have an anisotropic tumbling. In cases of doubt, both isotropic and anisotropic models where calculated and differences evaluated.

The "best" model for local motion was manually selected based on three criteria: first, models that after fitting predicted $T_1$, and $T_2$ values that were less than 2 % different from the experimental measured $T_1$, and $T_2$ values; secondly, the lowest AIC (isotropic) or TAIC (anisotropic) score was chosen. Thirdly, when these criteria did not return unambiguous results, the simplest model with the lowest AIC score was chosen. AIC values are automatically calculated according to Equation 3.3.

***Equation 3.3***

$$AIC = \chi^2 + 2k$$

*Where $\chi^2$ describes how good the curve fits the data, and k is the number of variables in the model.*

The curve-fitting parameters $S^2$, $S_f^2$, $\tau_c$, $\tau_e$, and $\tau_s$ from the "best" models were plotted against the protein sequence and are shown in the results section.

## Calculating δCSP

Chemical shift perturbation (δCSP) is a measure of how much a ligand (or other affector) affects the chemical environment of the backbone amide at a specific residue. When proton and amide backbone assignments are complete, the δCSP value between the bound and unbound situation ppm values can be calculated using Equation 3.4 for each amino acid (Williamson et al. 1997).

***Equation 3.4:***

$$\delta\mathrm{CSP} = (\,{}^{1}HN - {}^{1}HN_{ligand}) + \frac{({}^{15}N - {}^{15}N_{ligand})}{C}$$

*Where $^{1}$HN is the backbone amide proton ppm value, $^{15}$N is the backbone amide nitrogen ppm value, and C is the correction factor which is 7 for all amino acids except for glycine where it is 5.*

# 4 Results

## 4.1 Angle between tryptophan residues

The basis of methylation-recognition by CW-domains could involve several types of interactions. However, π-cation interactions are thought to be part of the stabilization of the ligand-bound state and the exact configuration of aromatic residues and the modified Lys in the binding pocket may be part of the recognition mechanism. To investigate this, the angle between the tryptophan residues that are involved in binding was related to different methylation states of H3K4.

As depicted in Figure 4.1, there is a clear difference between ZCWPW1 whics is known to bind trimethylated lysine strongest (He et al. 2010), MORC3 which is known to bind dimethylated lysines strongest (Strømland 2012), and ASHH2, which is known to bind monomethylated lysine residues strongest (Hoppmann et al. 2011). There is a larger difference between ASHH2 and MORC3 than between MORC3 and ZCWPW1, which is in agreement with our model which suspects that the two methyl groups in dimethylated lysine are turned towards the tryptophan residues. It is noteworthy that, for ZCWPW1, the Trp-angle changes little when ligand is bound.

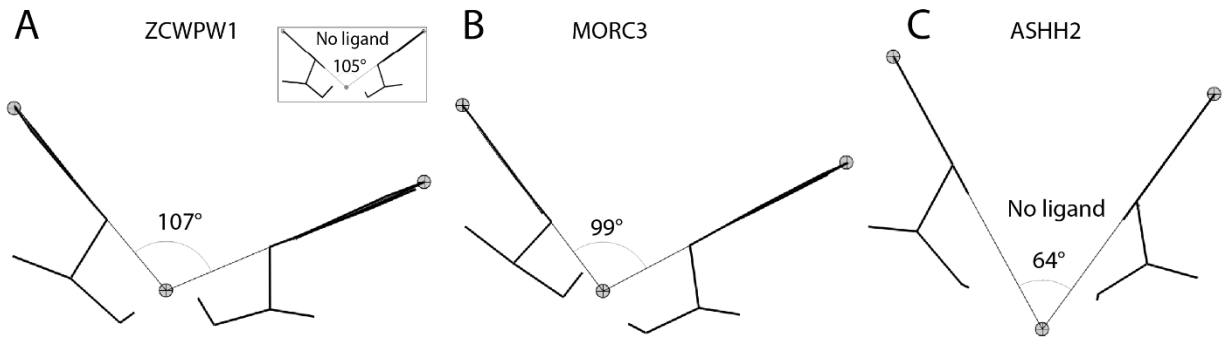


***Figure 4.1: Angles of the tryptophan residues in CW domain binding pockets.*** *(A) ZCWPW1, (B) MORC3, and (C) ASHH2 without ligand. Insert in A shows the angle of ZCWPW1 without ligand.*

## 4.2 Steric hindrance in the ASHH2 CW domain binding pocket

The lid hypothesis suggests that the C-terminal contributes to binding of the ligand with specific interactions. In ASHH2 this region is in the form of a helix that covers the binding site, raising the question if monomethyllysine would fit there. Visual comparison reveals that the binding site is small. To obtain a better understanding of the required size of the ASHH2 CW binding site, was compared to the monomethyllysine binding site in SCML2 (Figure 4.2). By drawing a triangle between selected atoms and measuring the lengths between them followed by area calculations yielding 8.6 Å$^2$ for ASHH2, and 12.6 Å$^2$ for SCML2. The angles and distances (90 ° and 3 Å from centre of an aromatic residue to a hydrogen in e.g. $NH_4^+$) suggested by Marshall *et al*. (2009) to be optimal, would be impossible to accomplish without allowing movement of the

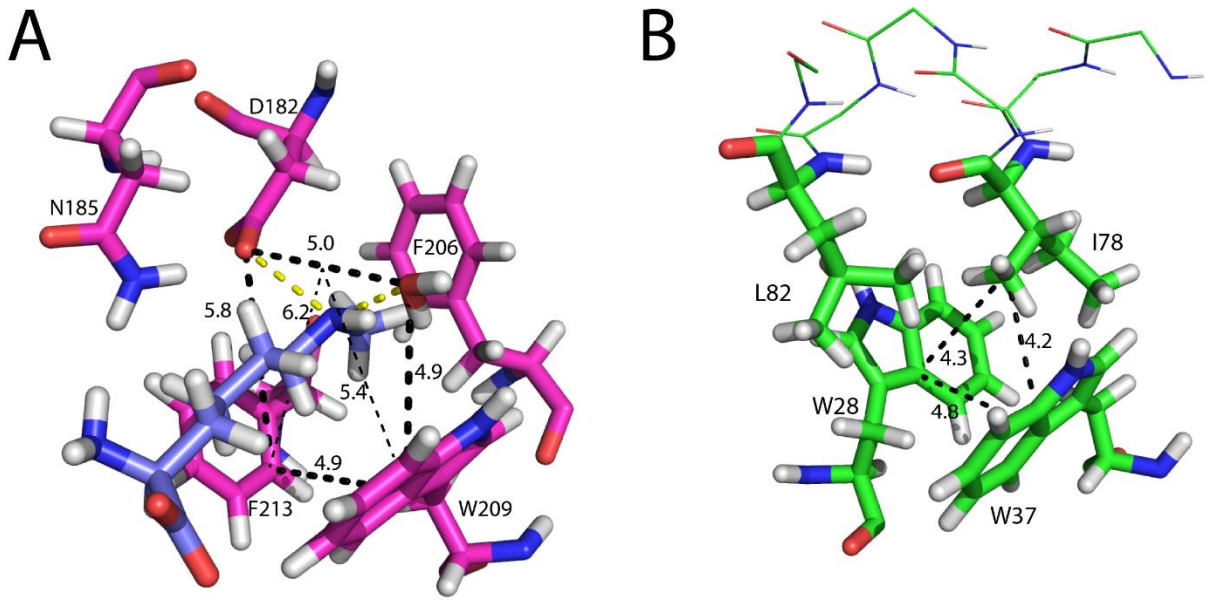


***Figure 4.2: Size comparison between SCML2 and ASHH2 CW binding pockets.*** *(A) Monomethylated lysine (grey) docked to the binding site (W209, F213) enclosed by D182 and a water molecule in human SCML2 (PDBID: 2VYT). Numbers refer to distances in Å. The available space for the ligand is shown as a long dotted triangle with sides measuring 4.9 Å, 6.2 Å, and 5.4 Å, area 12.6 Å$^2$ (The midpoint between D182 and hydrogen bonded $H_2O$ was used for comparability). B: Binding pocket of CWs (PDBID: 2L7P) consisting of W28 and W37 enclosed by I78 and L82 from the N-terminal α-helix. A triangle with its sides measured to be 4.8 Å 4.3 Å and 4.2 Å. Area 8.6 Å$^2$, show that there is less space for the ligand in CWs compared to SCML2. All hydrogen atoms are white, nitrogen atoms are dark blue, and oxygen atoms are red. Residue number refers to the amino acid sequence in the PDBs indicated here, i.e. 2VYT and 2L7P.*

helix in ASHH2. Since binding to H3K4me1 has been shown (Hoppmann et al. 2011) this is indicative of local motion in this helix, relative to the tryptophan residues.

## 4.3 Estimation of binding parameters for selected peptides with ITC

The residues from A1 to R8 of the H3 N-terminal tail are implicated in the binding of H3K4 to CW domains. The solution structure of the human MORC3 CW domain is similar to the CWs structure (backbone RMSD of 1.7 Å in a FATCAT alignment). The structure for the ligand bound situation in MORC3 is solved (PDBID: 4QQ4). In this structure, K9 is close to the unstructured loop between the two beta strands. This opens for the possibility that PTMs also at K9 can affect binding. Furthermore, a histone peptide array suggested that CWs bind stronger to the doubly modified H3K4me1K9ac and H3K4me1K9me2 compared to the "singly modified" H3K4me1 peptide (pers. comm. H. Z. Nenseth, R.B. Aalen, University of Oslo).

### 4.3.1 Testing the effect of H3K9ac or H3K9me2 on H3K4me1 specific binding of the ASHH2 CW domain

To test whether H3K4me1K9ac and H3K4me1K9me2 bind to the CW domain more strongly than H3K4me1 (used as a reference and positive control for binding), ITC was used to determine the $K_D$ of these doubly modified ligands (Figure 4.3). The unmodified H3 peptide does not bind to the CW domain (Hoppmann et al. 2011), and is used as a negative control in this experiment. All the reactions are exothermic, as can be seen from positive raw heat rates in Figure 4.3 A, C, E, and G. The values from the integration of the raw data from each titration step were plotted against the molar ratio between peptide and CWs. The binding curves of the H3K4me1 peptide, H3K4me1K9ac peptide and H3K4me1K9me2 peptide have a sigmoidal shape (Figure 4.3 B, D, and F). This was not observed for the negative control, and binding parameters were thus not calculated for the H3 peptide (Figure 4.3 H)

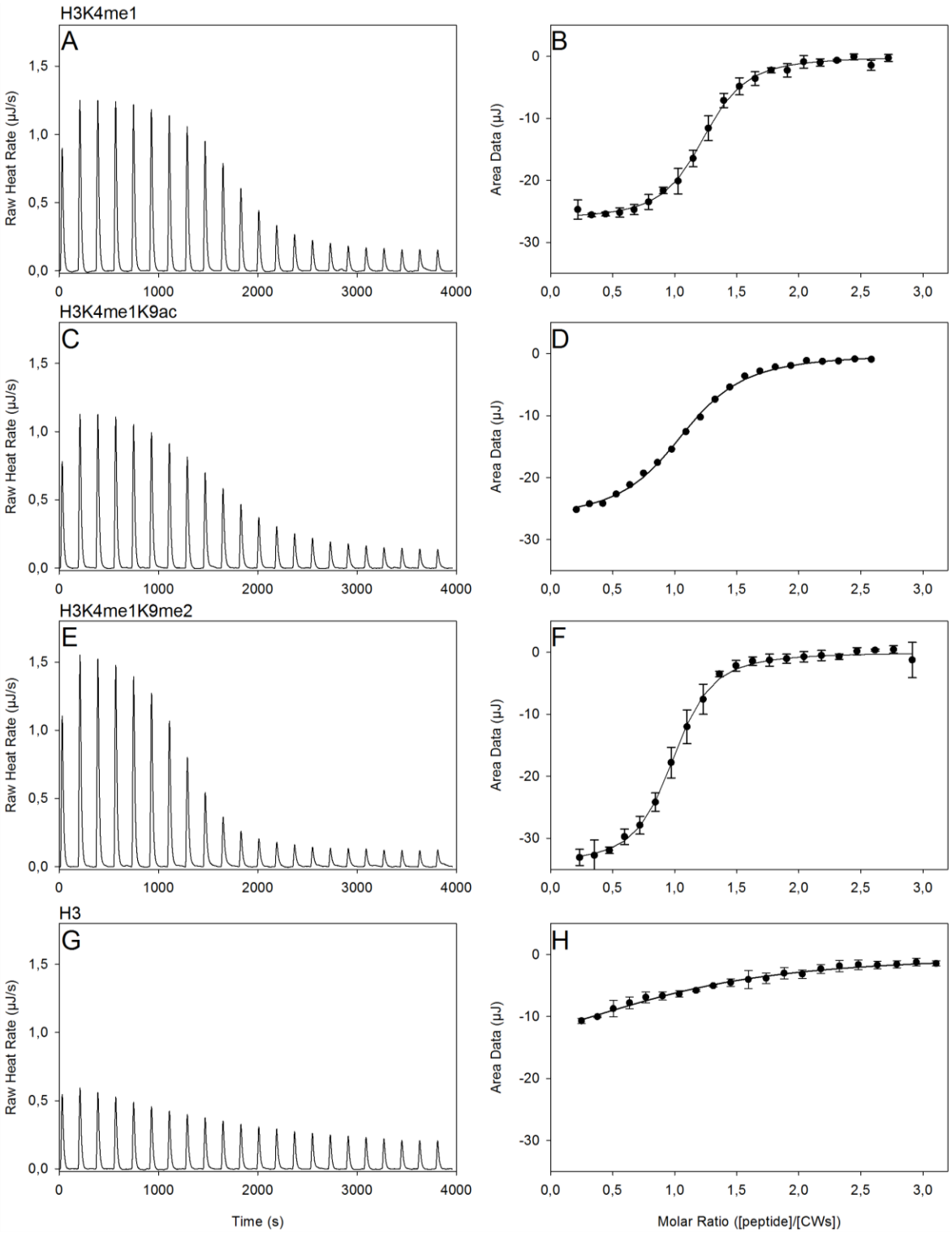


***Figure 4.3: ITC titration of double modified H3 peptides.*** *(A, C, E, F) ITC heat transfer curves as a function of peptide injections. 200 µM ligand was titrated into 20 µM CWs in T7 buffer with a stirring rate of 300 rpm. The raw heat of peptide dilution in T7 buffer has been subtracted from each titration curve. (B, D, F, H) Area under injection peaks plotted vs. [peptide]/CWs] molar ration. Data series have been fitted to an independent model where the first injection was removed from the analysis. The error bars show 1 standard deviation between technical triplicates. Note that there are not error bars in D, since one parallel was performed with a slightly different peptide concentration. (A, B) H3K4me1 peptide. (C, D) H3K4me1K9ac peptide (E, F) H3K4me1K9me2 (G, H) Unmodified H3*

As summarized in Table 4.1, and Figure 4.4, there is little difference in $K_D$ between the positive control (0.49 µM ± 0.05 µM), and H3K4me1K9me2 (0.45 µM ± 0.05). This suggests that additional dimethylation of K9 has little effect on binding. Furthermore, H3K4me1K9ac has a higher $K_D$ (1.3 µM ± 0.11) than the positive control, suggesting that the additional acetylation of K9 results in weaker binding. While the binding is significantly weaker than the reference, the data shows a clear sigmoidal shape and is not disrupted completely. When the ligand is added, there is a negative change in entropy in all cases. This is consistent with a binding event that induces structure from loops or restricts the mobility of side chains, loops, helices or other structural elements. Furthermore, the ΔS and the ΔH are different in the case of H3K4me1K9me2.

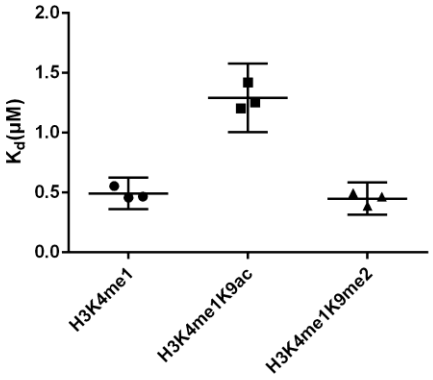


***Figure 4.4: Summary of dissociation constants ($K_D$) of CWs ligands.*** *Error bars indicate 95 % confidence interval.*

***Table 4.1: Calculated CWs binding parameters towards possible ligands at 25 °C.***

| Parameter | H3K4me1 | H3K4me1K9ac | H3K4me1K9me2 | H3 unmodified |
|---|---|---|---|---|
| **ΔH (kJ/mol)** | -65 ± 1 | -66 ± 7 | -92 ± 4 | n.d. |
| **n** | 1.20 ± 0.05 | 1.11 ± 0.07 | 0.96 ± 0.05 | n.d. |
| **Kd (µM)** | 0.49 ± 0.05 | 1.3 ± 0.11 | 0.45 ± 0.05 | n.d. |
| **ΔS (J/mol*K)** | -99 ± 2 | -108 ± 25 | -187 ± 16 | n.d. |

*Reported values are averages from technical triplicates ± 1 standard deviation. No sigmoidal curve was found for H3 and the parameters where not determined (n.d.)*

## 4.4 Stability of the ASHH2 CW domain

To investigate the relative contribution of the unstructured termini, C-terminal helix, $Zn^{2+}$, and H3K4me1 peptides, thermal denaturation and EDTA depletion of CW domains of different lengths was performed. Moreover, it was of interest to determine whether the expressed proteins were properly folded and stable in solution before extended NMR experiments on $^{15}N$ and $^{13}C$ labelled samples were attempted. To clarify these questions, the stabilities of the proteins were evaluated by intrinsic tryptophan fluorescence spectrometry. The melting temperature ($T_m$) of denaturation of a folded protein, reflects its resistance to heat denaturation and a clear transition is indicative of a folded protein (Pace & Shaw 2000). By comparing the $T_m$ between different versions of the CW domain, with and without ligand, it may be possible to suggest what elements contribute to overall stability of the protein alone and in complex with ligand.

### 4.4.1 CWs and CW42 are stable at 25 °C

The fluorescence intensity curves for CWs (Figure 4.5 A), and CW42 (Figure 4.5 E) show that both proteins without their ligands, have fluorescence maxima at about 342 nm at 4 °C. Upon heating, they undergo a transition until their fluorescence maxima approaches 355 nm, which is close to the value for free tryptophan in solution (i.e. fully solvent exposed). When the fluorescence ratio (I335/I355) is plotted against temperature, as shown for CWs in Figure 4.5 B and CW42 in Figure 4.5 F, it can be observed that at 25 °C both proteins are in the pre-transition state. CWs has a $T_m$ of 50 °C ± 1.5 °C whereas CW42s is 58 °C ± 1.4 °C in the absence of ligand (Figure 4.6).

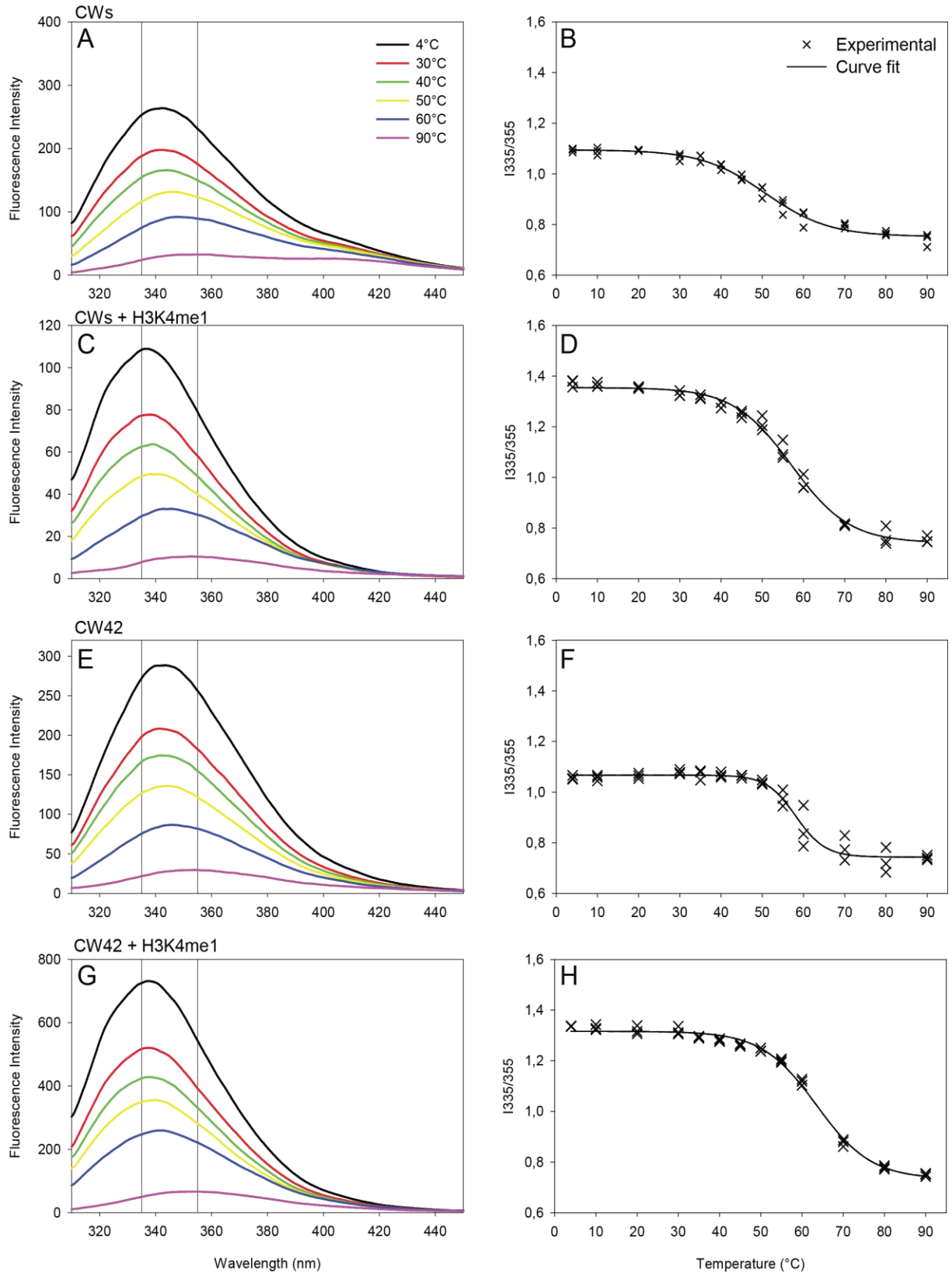


***Figure 4.5: Thermal denaturation of CWs, and CW42 with and without H3K4me1 mimicking peptide***. *(A, C, E, G) show fluorescence intensity spectrums as a function of temperature. Note that there is a y-axis mismatch between the experiments due to a difference in protein concentration of 6 µM – 18 µM. (B, D, F, H) Show the fluorescence ratios I335/I355 that provide a description of protein folding that does not rely on manual reading of emission maximum wavelengths. The approach also removes the effect of sample concentration differences. The curves are fitted to a sigmoidal model with 4 variables (Equation 3.2). Only selected temperatures are shown in raw fluorescence intensity graphs*

## 4.4.2 Protein is stabilized by addition of ligand

When a 5 molar excess H3K4me1 peptide was added, both proteins produced a lower fluorescence maxima (about 335 nm, indicative of less solvent accessibility of tryptophan residues) compared to the ligand-free construct at 4 °C (Figure 4.5 C, and Figure 4.5 G). Also, when the ligand is present, a single transition occurs upon heating, before the fluorescence maxima approaches 355nm (Figure 4.5 D, and Figure 4.5 H). The $T_m$ of CWs with ligand was measured to be 57 °C ± 1.0 °C, while CW42 had a $T_m$ of 64 °C ± 1.0 °C. (Figure 4.6).

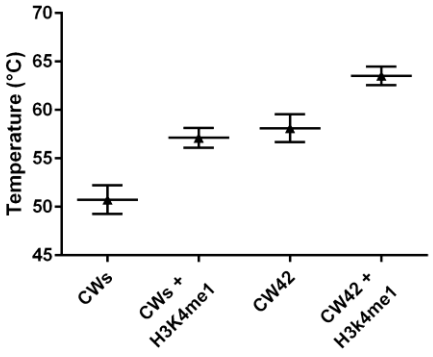


***Figure 4.6: Summary of $T_m$ for CWs and CW42 with and without ligand.*** *Error bars show 95 % confidence interval.*

## 8.4.3 Folding state and stability of the CWΔLID

ASHH2 CW domains with a truncation of the C-terminal helix (CWΔLID) has been shown to not bind histone H3K4me1 peptides (Hoppmann et al. 2011). To determine whether this loss of binding capability is due to loss of specificity, or disruption of the fold, thermal denaturation of CWΔLID was performed, and measured with intrinsic tryptophan fluorescence.

The fluorescence maxima at 4 °C is at about 350nm (Figure 4.7 A). Figure 4.7 B shows that at 4 °C, the fluorescence ratio is less than 0.9, and upon heating the CWΔLID construct has a near linear and flat trend, as it approaches the fluorescence ratio (0.6) of an unfolded protein with tryptophan residues completely exposed to the solvent at 90 °C. This small difference and lack of a sigmoidal transition between low temperature and high temperature was not indicative of a "well-behaved", folded protein, and no curve fitting was performed. This suggests that the C-terminal is required for protein stability.

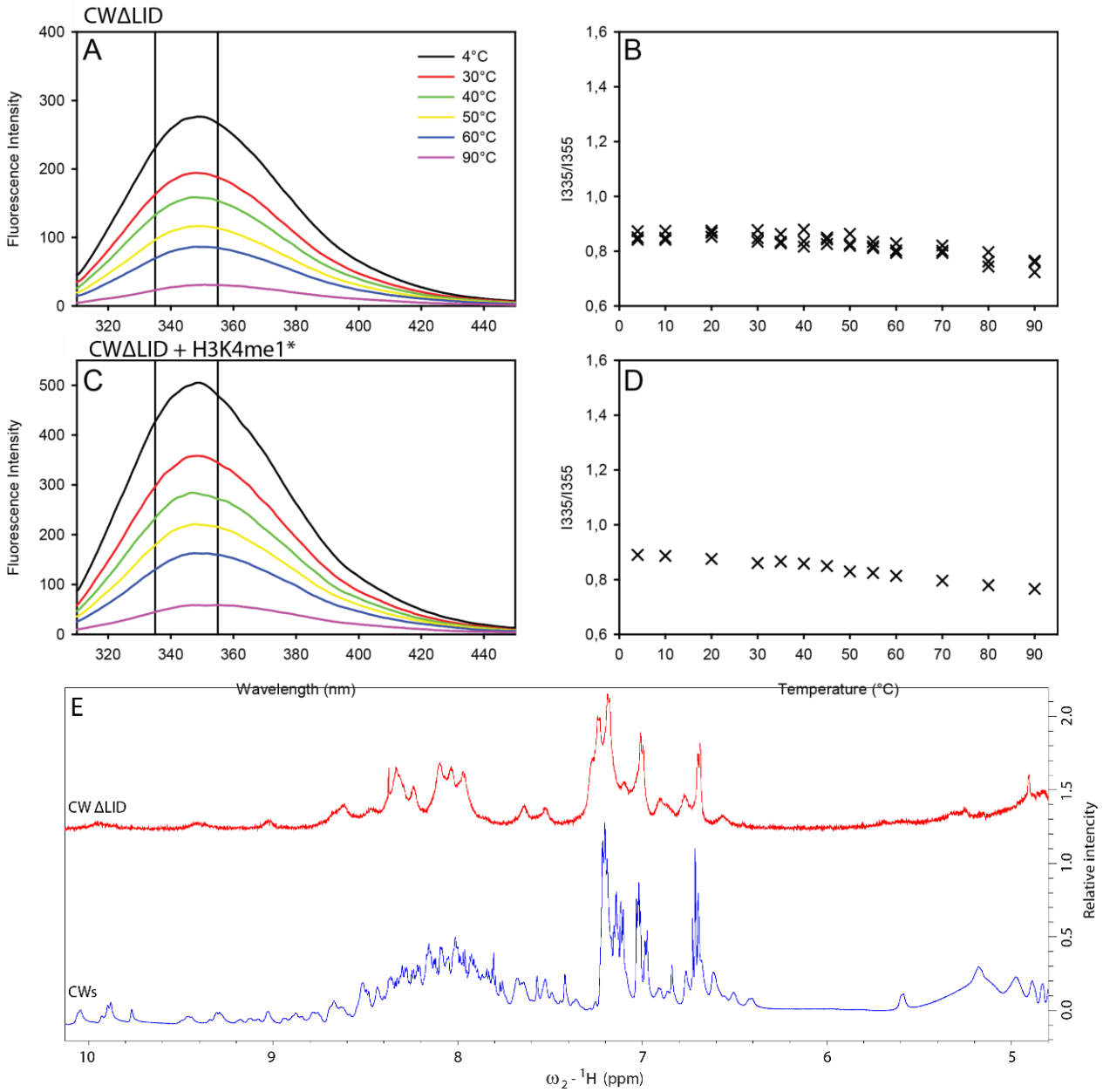


***Figure 4.7: Thermal denaturation of CWΔLID where the C-terminal helix has been removed.*** *(A, C) Fluorescence intensity, and C (with ligand), has only been measured once. No curve fitting was attempted on the fluorescence ratio 335 nm/355 nm, shown in B and D. In figure E, the proton spectrum of 0.3 mM CW ΔLID is compared to the proton spectrum of 1 mM CWs (blue trace). CWΔLID (red trace) was d in a high salt (T7 buffer) whereas CWs was, in low salt (NMR buffer). The CWΔLID trace has been shifted up 1.5 relative units to make comparison easier.*

Newly purified CWΔLID is typically soluble at concentrations up to 0.3 mM. At this concentration it precipitates when stored overnight at 4 °C, or upon further concentration, again indicative of an unfolded protein that is prone to aggregation. To check whether the protein is indeed

unfolded, a simple 1D proton NMR spectrum was acquired. As depicted in Figure 4.7 E, the NMR proton spectrum of CWΔLID has fewer and broader peaks spread across a narrower spectral region compared to a similar spectrum for CWs. This confirms that the truncation of the C-terminal helix results in an unfolded protein.

### 4.4.3 $Zn^{2+}$ is essential for structural integrity

Since the CW domain has low a content of secondary structure, a major contribution to its stability comes from the $Zn^{2+}$ ion coordinated by four cysteine residues. Using EDTA as a chelating agent, it can be seen that the fluorescence maxima, decreased when the amount of EDTA is increased (Figure 4.8). A pre-transition state was not observed, but the post-transition state is, however, similar to that of the thermally denatured CW42 (I335/I355 of about 0.7). Not surprisingly, since the first EDTA concentration already has a higher than 1:1 molar ratio between [CW42] and [EDTA], no sigmoidal curve was observed in Figure 4.8 B. A curve fitting by Equation 3.2 was performed, suggesting that 1.3 times molar excess of EDTA is required to obtain half maximal denaturation. However, since no sigmoidal curve was observed, this value should be viewed with caution. It is however clear that EDTA treatment introduces changes in the protein that make the tryptophan residues more solvent accesible and that this transition is saturated at about 20 times molar exess of EDTA.

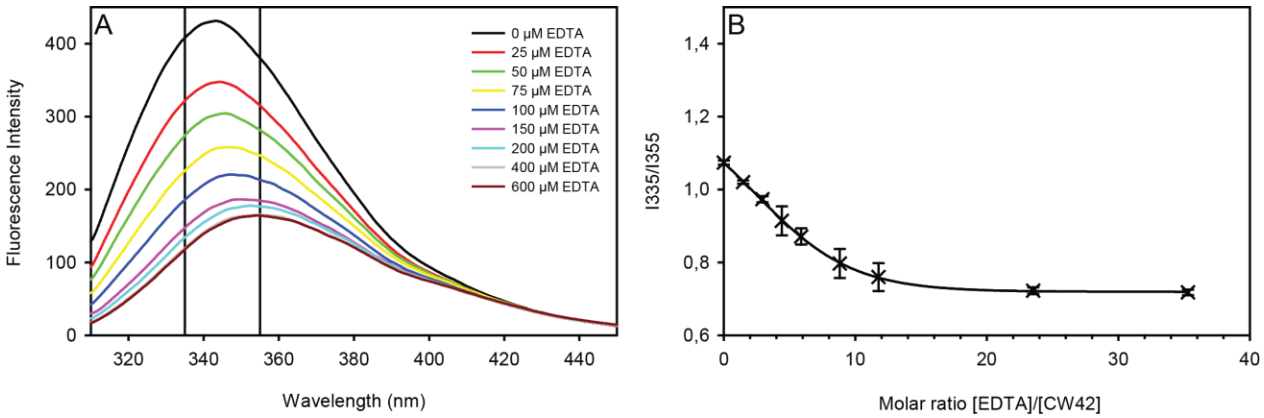


***Figure 4.8: (A) $Zn^{2+}$-depletion of CW42 monitored by fluorescence spectrometry.*** *23 µM of CW42 was titrated using EDTA. Vertical lines indicate the wavelengths that were used for calculating fluorescence ratios (I335/I355). (B) Fluorescence intensity ratio I335/355, plotted against the molar ratio of EDTA and CW42.One standard deviation of three technical replicates are shown as error bars.*

## 4.4.4 Reversibility of $Zn^{2+}$ removal

To see if the unfolding of the protein upon removal of $Zn^{2+}$ is reversible, the changes in the intrinsic tryptophan fluorescence of both CWs, and CW42 were observed upon addition of EDTA, followed by addition of excess amounts of ZnAc. Ligand binding capability was also tested by adding H3K4me1 peptide.

As shown in Figure 4.9 (Native), free CWs and CW42 have a fluorescence ratio (I335/I355) similar to CWs and CW42 proteins in the pre-transition state during the thermal denaturation experiments (I335/I355 of about 1.1). Upon $Zn^{2+}$depletion, CWs gave I335/I355 of 0.77 ± 0.023 which is comparable with the post transition state of thermal denaturation (about 80 °C in Figure 4.5 B). Surprisingly, CW42 had an I335/I355 of 0.85 ± 0.023 upon $Zn^{2+}$ depletion, which is comparable to heat denaturation at about 60 °C (Figure 4.5 F). This suggests that 20x molar excess of EDTA did not result in complete exposure of the tryptophan residues. This is not in agreement with the EDTA treatment depicted in Figure 4.6. The addition of H3K4me1 peptide to the $Zn^{2+}$ depleted sample gives a small recovery. When the $Zn^{2+}$ depleted sample was reconstituted with $Zn^{2+}$, there is an increase in fluorescence intensity to a I335/I355 ratio of 0.82 ± 0.015 for CWs and an almost complete recovery in fluorescence ratio for CW42 (I335/I355 of 1.06 ± 0.005). When the ligand was added to this "reconstituted sample", there is a further

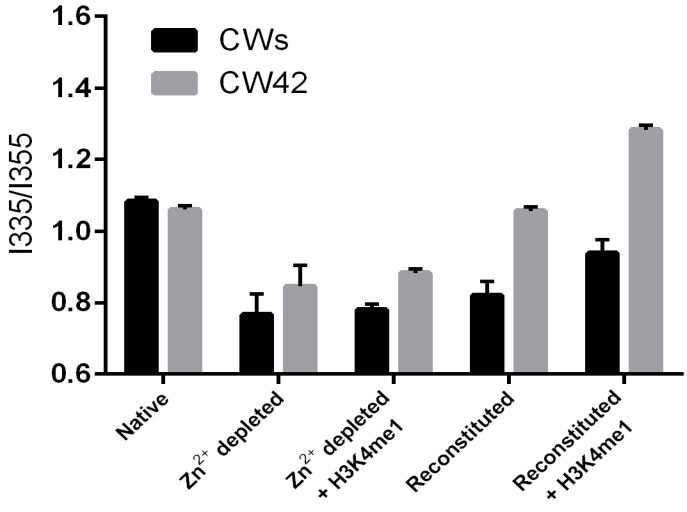


***Figure 4.9: $Zn^{2+}$ depletion and subsequent reconstitution of CWs and CW42.*** *The fluorescence ratio (I335/355) of free 15 µM CWs, and 23 µM CW42 upon treatment with EDTA, H3K4me1 peptide, and ZnAc in various combinations at 25 °C.* ***(Native)*** *CWs and CW42 before any treatment.* ***($Zn^{2+}$ depleted)*** *CWs and CW42 with 20x molar excess of EDTA.* ***($Zn^{2+}$ depleted + H3K4me1)*** *5 times molar excess peptide in comparison to protein concentration added to the $Zn^{2+}$ depleted sample.* ***(Reconstitution)*** *Reconstitution of CWs and CW42 by addition of 5 times molar excess of ZnAc compared to EDTA concentration to $Zn^{2+}$ depleted sample.* ***(Reconstitution + H3K4me1)*** *5 times molar excess peptide in comparison to protein concentration added to reconstituted CWs and CW42.* ***(All)*** *Error bars show 1 standard deviation from 3 technical replicates.*

increase in fluorescence ratio. These values are lower compared to bound CWs and CW42 in the pre transition states in the thermal denaturation experiments (Figure 4.5 D, and H).

# 4.5 NMR Mobility studies

## 4.5.1 Comparing $^{15}$N HSQC of CWs without ligand with published data

To determine whether existing backbone assignments could be transferred, or a completely new backbone assignment was required, the HSQC of CWs expressed in this project was compared to the previously published HSQC of CWs (Hoppmann et al. 2011). As shown in Figure 4.10, the two spectra showed a high degree of overlap. There are a few peaks that are only present in one of the spectra. M910 (corresponding to M73 for CWs), situated at 113.2 $^{15}$N ppm and 7.2 $^{1}$H ppm, is much weaker in our dataset, but its ppm values are conserved. R3, situated

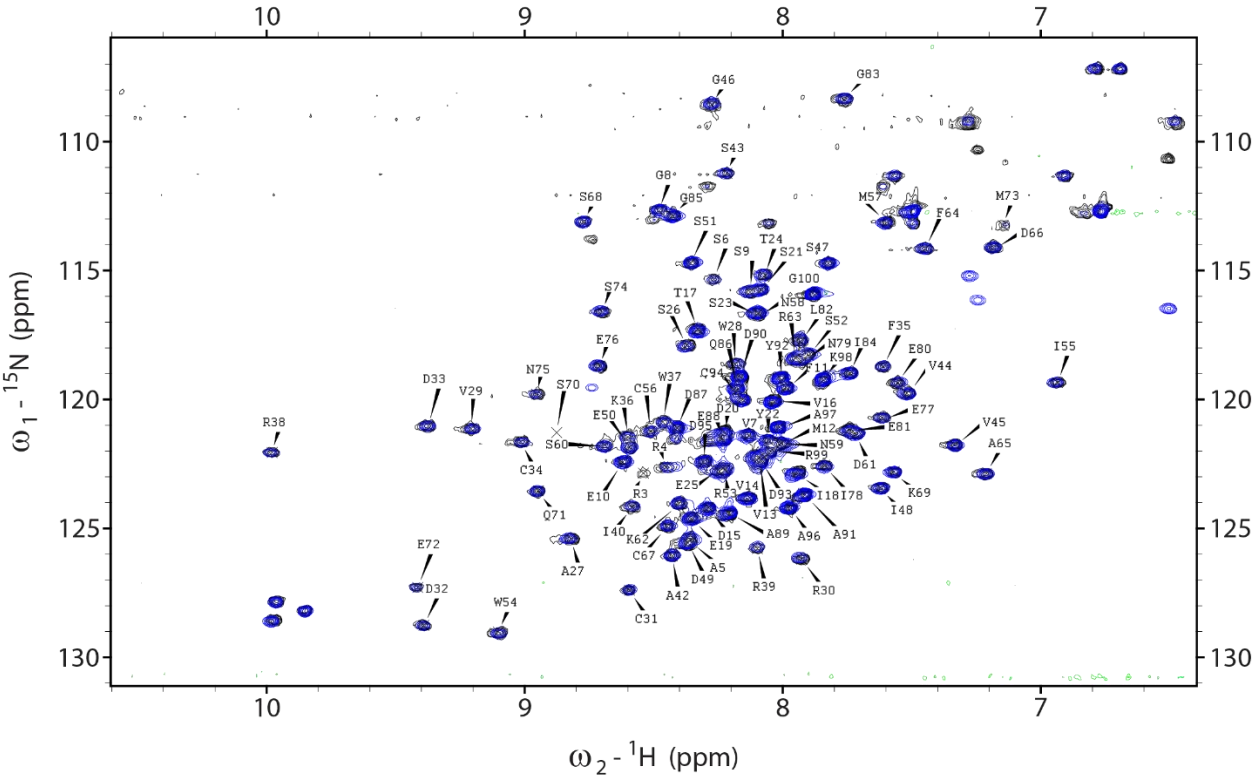


***Figure 4.10: Overlay of HSQCs of CWs without ligand.*** *Data and assignment from the Hoppmann 2011 paper (black) obtained from the BMRB database* (Ulrich et al. 2008) *are compared to NMR data generated in this study (blue). Hoppmann data was shifted -0.03 ppm in the $^{15}$N, and -0.09 ppm in $^{1}$H dimension to fit our data.*

at 122.8 $^{15}$N ppm and 8.6 $^{1}$H ppm in the Hoppmann spectrum, is missing, or very weak in our spectrum. R3 is not a part of the ASHH2 CW domain, but part of the thrombin cutting site.

In general, both fingerprints display a wide chemical shift range indicative of a folded protein, and the crosspeaks have about the same width and intensities within each spectrum. There is some overlap, especially for the amino acids in the unstructured termini, which have ppm values typical for random coils. The Hoppmann assignment was transferred to our spectrum after all peaks were shifted by - 0.03 ppm in the $^{15}$N, and - 0.09 ppm in the $^{1}$H dimensions.

## 4.5.2 Comparing HSQC of CWs with H3K4me1 with published data

A similar comparison was performed on the CWs fingerprint in the presence of H3K4me1

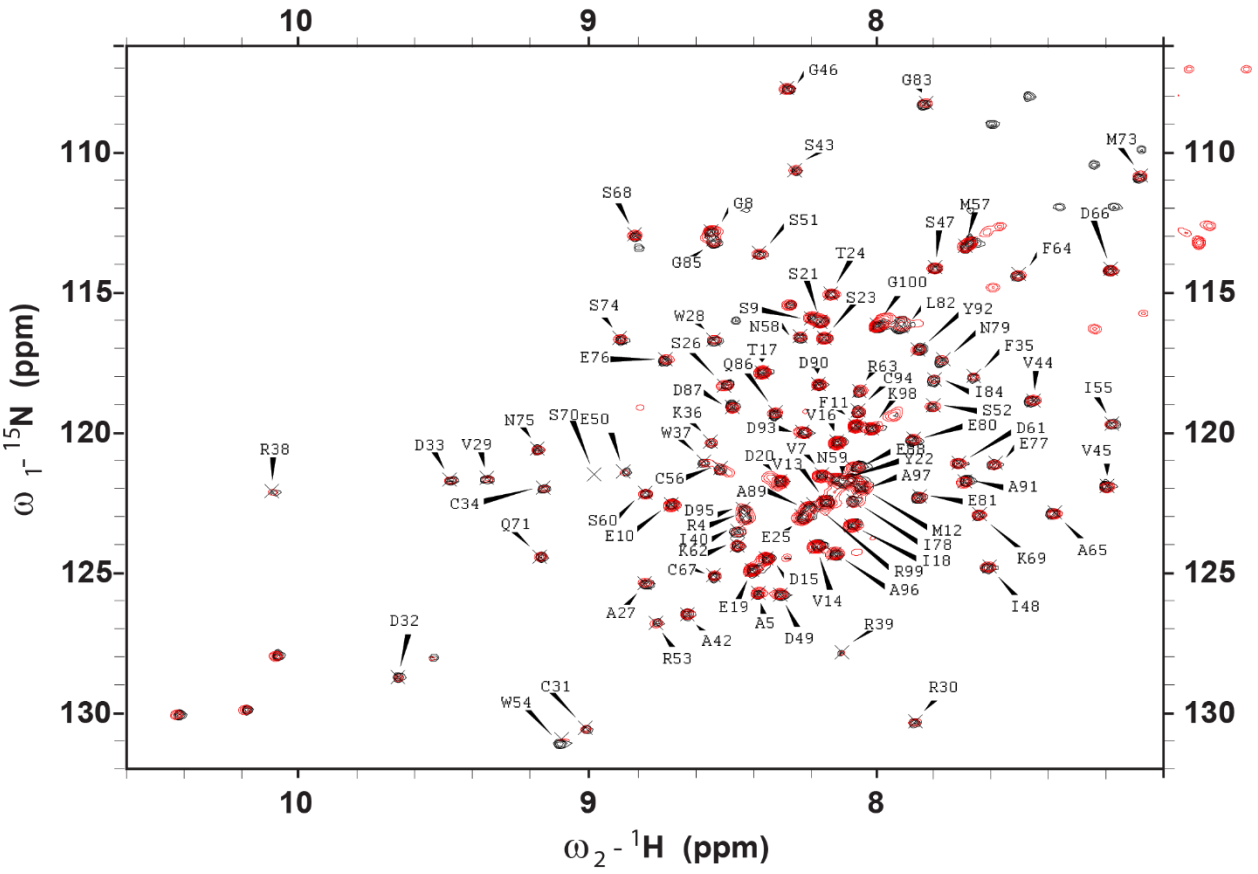


***Figure 4.11: Overlay of $^{15}$N HSQC of CWs with H3K4me1 mimicking peptide present.*** *Data from the Hoppmann 2011 paper colored black, and compared to NMR data generated in this study that are colored red. It should be noted that a uniform, manual adjustment of the solution structure $^{15}$N HSQC image size has been performed. The ppm range shown inside the frame, are from the solution structure. There are more peaks present in our dataset located downfield on the $^{1}$H axis (below 7 $^{1}$H ppm). Half of the W891 (W54 in CWs) peak shown at 130 $^{15}$N ppm, and 9.1 $^{1}$H ppm, have been spectrally folded in our data. The whole crosspeak is thus not shown in this spectral range.*

histone peptide. Here, the chemical shift values from the Hoppmann paper were not available. However, as shown in Figure 4.11 the images from the publication of the chemical shift data show a clear resemblance to the HSQC spectrums of CWs in this study. Our spectrum coloured red in Figure 4.11 have several more peaks, indicating that the sample situations are not the same. All the peaks that were annotated as backbone amides in the published data are present at nearly the same ppm values in our data. Thus, this assignment was transferred. However, since the raw data was not available, the peaks was re-picked in our data and labelled according to the assignment in Hoppmann et al. 2011.

### 4.5.3 NMR relaxation studies of CWs

One of the central aims of this thesis is to test the lid hypothesis for the interaction of ASHH2 to H3K4me1 by investigating the mobility of the CW domain on picosecond to nanosecond timescales. NMR relaxation studies were performed on the CWs protein both in the presence and absence of ligand. Three types of data, all containing information about molecular movement on the on this timescale were acquired. $T_1$ and $T_2$ relaxation data contains information about overall tumbling as well as the local flexibility of the polypeptide chain and heteronuclear NOEs (hNOE) provides information about the motion of NH-bond vectors in space (Doucleff et al. 2011).

*Table 4.2: Global relaxation parameters*

| Constant Name | Without ligand | With ligand |
|---|---|---|
| **# accepted residues** | 23 | 37 |
| **Global correlation time ($\tau_c$)** | 8.1 ns $\pm$ 0.39 ns | 8.8 ns $\pm$ 0.34 ns |
| **$D_{\parallel}/D_{\perp}$** | 1.24* | 1.20* |
| **Anisotropy** | 0.996 | 1.15 |
| **Rhombicity** | 82.1 | -1.23 |

**Note that both in the presence and absence of H3K4me1, CWs has a $1.2 \leq D_{\parallel}/D_{\perp} < 1.5$.*

For $T_1$, and $T_2$ the decay upon increase in mixing time were successfully fitted to a model for exponential decay for all amino acids except the first three residues that were not assigned and proline that lack the backbone amide group. The first residues are likely to have a high degree of water exchange. This claim is supported by predictions of a large exchange of protons with water with half-lives less than 5.62 x$10^{-04}$ $s^{-1}$ in the SPHERE webserver for these residues (Zhang 1995). $T_1$, and $T_2$ peak intensity upon increasing mixing times are shown for the two tryptophan residues in Figure 4.12. The decay constants $T_1$ and $T_2$ value are reported for each residue, and used in the model free analysis. This analysis utilises global parameters of CWs with and without ligand that is tabulated in Table 4.2 are used uniformly for all residues. From the "accepted residues" the global isotropic correlation time $\tau_c$ was calculated. In addition, the ratio between the parallel and the perpendicular NH bond vectors, anisotropy and rhombisity of the molecule

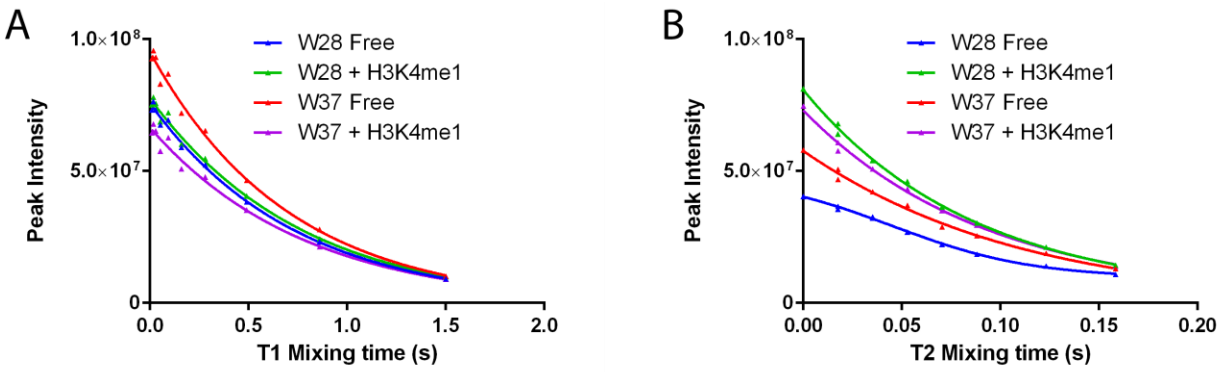


***Figure 4.12: $T_1$ and $T_2$ peak intensity of W28 and W37 in CWs with and without H3K4me1 upon increasing mixing times.*** *(A) $T_1$ relaxation data with and without ligand. (B) $T_2$ relaxation data with and without ligand. Both $T_1$ and $T_2$ data fit to an exponential decay model which was also true for all residues except G1, S2, R3, and P41 of CWs.*

were predicted. The obtained values are tabulated in Table 4.2. Note that these calculations were based on the selection criteria described in the methods part, and that a higher number of residues were accepted for CWs with ligand bound. This is in agreement with a more stable protein, since the criteria mainly reject highly mobile residues (e.g. rejection of hNOEs < 0.7). The global correlation time became 0.7 ± 0.52 ns longer when ligand was added which is consistent with an increase in the size of the complex, corresponding to the addition of a 1.51 kDa unit (Doucleff et al. 2011). The choice whether an isotropic or anisotropic model should be followed was unclear since the global correlation time was in the intermediate region ($1.2 < D_{\parallel}/D_{\perp} < 1.5$) between typical isotropic an anisotropic proteins (Mannhold et al. 2006). Calculations where continued under the isotropic assumption since the values are closer to 1.2 than 1.5.

$T_1$ data depicted in Figure 4.113 A, show that the values of the fittings are generally higher, especially in the central regions of the CWs construct in the presence of the H3K4me1 peptides. The N-terminal termini also acquire significantly higher $T_1$ values when the ligand is present. In contrast, a more ambiguous change occurs at the C-terminal end. The tendency is that CWs obtain lower $T_1$ values when ligand is added, but A89 shown a much higher $T_1$ and C94 remains unchanged.

In Figure 4.13 B, the CWs $T_2$ values in the core region are similar for the ligand bound and unbound situation. However, the $T_2$ values are higher in the N-terminal and lower in the C-terminal for H3K4me1 bound CWs. Several of the central amino acids obtain lower $T_2$ values when the ligand is present. This includes residue 47-48, and 51- 53. However, $T_2$ measurements have large uncertainties in these regions.

hNOE experiments were conducted with and without the hNOE effect, in setup where the two data types where alternating and intercalated in the same dataset. The dataset was split into two datasets using the "split" command in TopSpin, and the integral ratio between the peaks for each amino acids was calculated and used in model free analysis. hNOEs are consistent with $T_1$, and $T_2$ data, and are shown in Figure 4.13 C. In the N-terminal region, hNOE values are

reduced when ligand is added, whereas higher hNOEs are found in the C-terminal in the presence of ligand. Higher hNOE values are found in the central region from residue 43 to 53. The C-terminal helix has a significant increase in hNOE values. This stability shift transverse the defined stability line at 0.7 hNOE, and indicates induction of a less mobile helix upon ligand binding.

### 4.5.4 Mobility of CWs with and without ligand

In order to extract dynamic information about the protein in solution, $T_1$, $T_2$ and steady-state hNOE data where used to conduct a model free analysis. This analysis assumes that the global correlation time $\tau_c$ is the same for all atoms and uses a least-squares data fitting algorithm to fit 5 equations (M1-M5 in the Dynamics Centre user manual ch. 8), which contain parameters describing different types of protein internal motions - $S^2$, $\tau_e$, $R_{ex}$, and $S_f^2$. Depending on the particular model used to account for experimental data, various combinations of the $S^2$ order parameter, the local correlation time $\tau_e$, the rate of exchange $R_{ex}$, and the order parameter for fast movements $S_f^2$ can be used to gain insight into the local mobility of the protein.

The computed $S^2$ describes how much of a residues motion that comes from local motion and is shown in Figure 4.13 D. The $S^2$ parameter range from 0-1 where 1 correlates to no local motion and 0 to a fully flexible protein (Lipari & Szabo 1982). In this work, residues with $S^2$ values that are higher than 0.8 are considered to be immobile relative to the overall motion of the tumbling protein (Suggested in the DC manual and in litterature e.g. Doucleff et al. 2011). Interestingly, as shown in white spheres in the Figure, the unbound CWs has several regions of high mobility in addition to the flexible termini. Most prominently, several amino acids in the central unstructured region (aa 46 - 55) have $S^2$ values of about 0.6. Furthermore, the C-terminal helix has $S^2$ values between 0.7 and 0.8 in the unbound state. This includes isoleucine 78 that is occupying the binding pocket (I915 in Uniprot), but not leucine 82, which is also docked in the binding site (L919 in Uniprot). This show that there is a significant movement of the lid and that

in this region about 25 % of the motion come from local motion (the rest comes from global tumbling $\tau_c$)), in the unbound state.

Upon binding, most of the amino acids in the globular part of the protein (from the first beta-sheet to the C-terminal helix) obtain a less mobile configuration compared to the unbound protein, and most of the amino acids have a higher $S^2$ value than 0.8. This shows that the domain has less local motion upon binding.

Several residues have motions that can be explained by a transition between two distinct states. The rate of this transition is described by the $R_{ex}$ parameter and is shown in Figure 4.14 A. For CWs without ligand no clear trend in $R_{ex}$ has been found. It appears that amino acids, mostly within the globular region but not in the secondary structure elements of the protein, have been fitted to models with a $R_{ex}$ component. When ligand is added, several more residues in the globular part of the protein fit to a model with a $R_{ex}$ component in it. This also includes several amino acids in the C-terminal helix and in the residues right before it (from S70 to E81). The frequency of the exchange between the states is approximately 0.7 $s^{-1}$. This is in the typical timescale of relatively fast intermolecular interactions (Timescale: ms-s), or alternatively very slow intramolecular loop movements.

Not all residues have a clear transition between two distinct states, but have several states or very unrestricted movement patterns. These residues are typically fitted to equations that contain a variable with local correlation time ($\tau_e$) (or $\tau_s$ for very fast movements). $\tau_e$ and $\tau_s$ describe the time it typically takes for a given residue to repeat its complex motion. $\tau_s$ values are shown in Figure 4.14 B, and show that this is a typical motion that is common in the unstructured termini. These amino acids repeat their motion in about 5 – 8 ns. The core region has a larger variation, but no obvious trend can be found.

Residues fitted with the $\tau_e$ parameter have much faster movements, which repeat themselves on the picosecond - nanosecond timescale as shown in Figure 4.14 C. The $\tau_e$ values are prominent in the central regions of the CWs domain, but there are many more that have this fast movement when the ligand is bound. There are, however, a few amino acids (T17, E19, D49,

S70, L82, and Y92) that have much higher values (>2 ns) when the ligand is not present. Only Y92 has this high $\tau_e$ after binding. These fast motions are also described by the $S_f^2$ parameter shown in Figure 4.14 D, but no general trends have been found in either the bound or the unbound state.

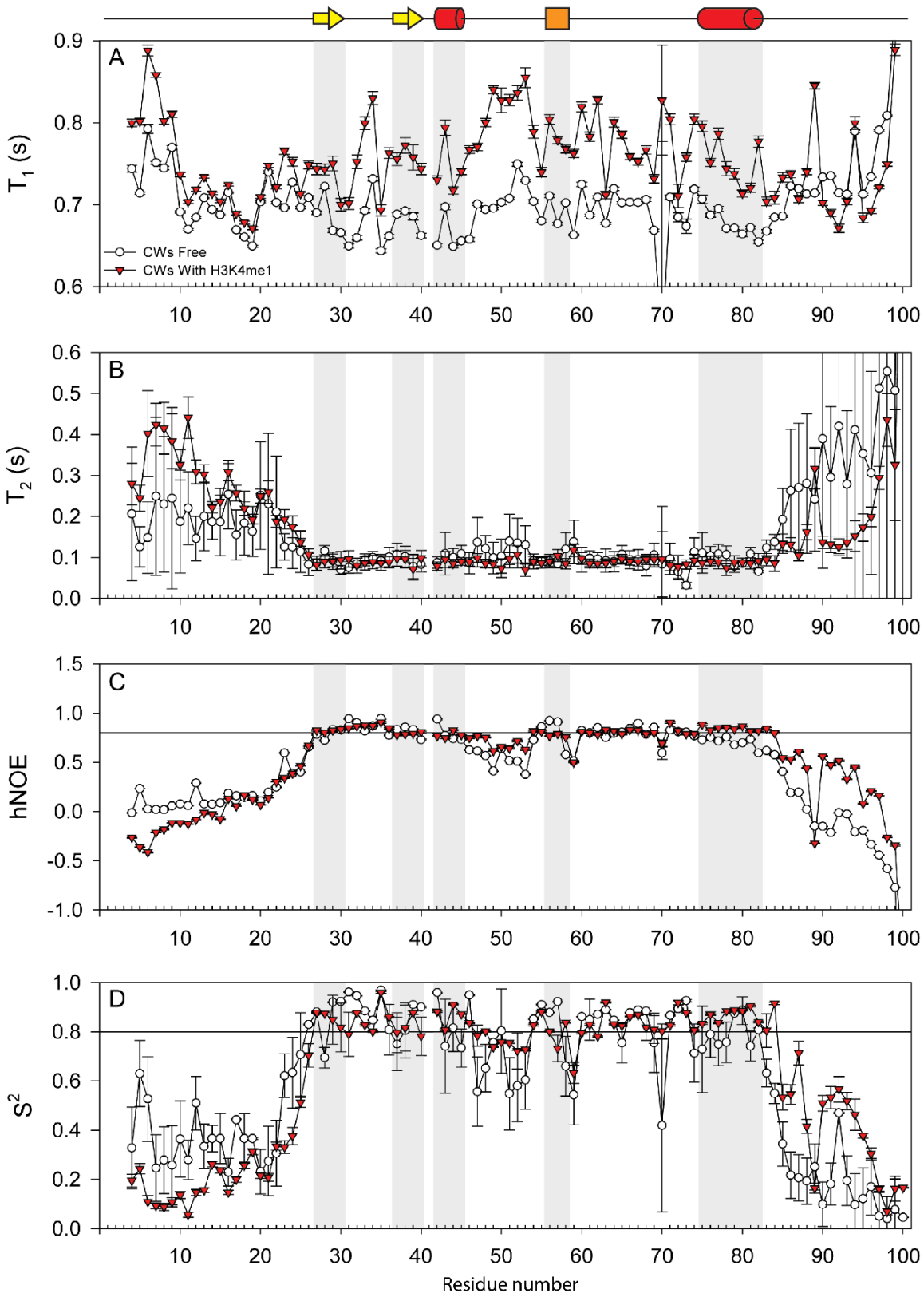


***Figure 4.13: Relaxation data of CWs with (red triangle) and without (white circle) ligand.*** *5 times molar excess H3K4me1 peptide was used as ligand. (A) Spin-lattice relaxation ($T_1$). (B) Spin-spin relaxation ($T_2$). (C) Heteronuclear NOEs. (D) Individually selected $S^2$ values from models (M1-M5), following selection criteria described in Methods 3.7.3. A horizontal line at 0.7 in hNOE, and at 0.8 in $S^2$ indicate what parts of the protein can be considered to have stable local conformation. G100 is scaled out in panels A-C. Its values are $T_1$ = 1.4s ± 0.001s, $T_2$ = 1.3s ± 2.5 s and hNOE = -2.5 ± 0.005 for free CWs and $T_1$ = 1.3s ± 0.002 s, $T_2$ = 1.1 s ± 0.2 and hNOE = -1.7 ± 0.0054 with ligand.*

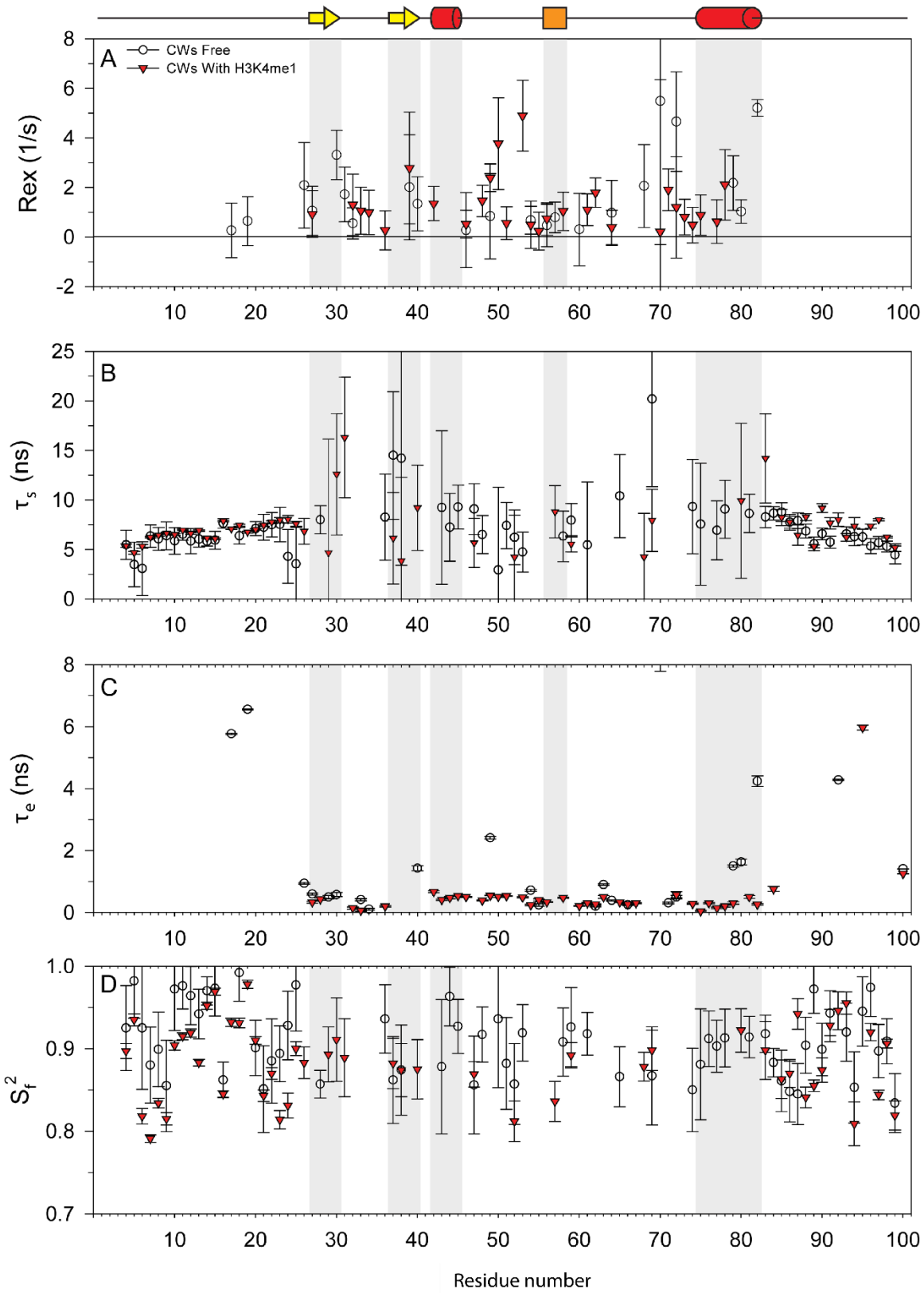


***Figure 4.14: Local motion of the CWs backbone.*** *CWs without ligand shown as white circles and CWs with 5 times molar excess H3K4me1 peptide shown as red triangles plotted against amino acid sequence. Secondary structure is shown above the panel, and structural elements are marked as shadows on the graphs. (A) Exchange rates ($R_{ex}$) between different conformational states of the backbone amides. (B) The correlation time $\tau_s$. (C) Fast correlation time $\tau_e$. (D) Order parameter for fast movement $S_f^2$.*

# 4.6 Backbone assignment of CW42

## 4.6.1 Backbone assignment of CW42 without ligand

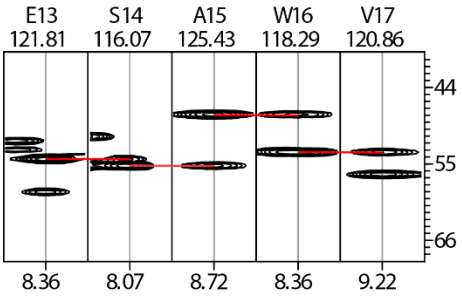


***Figure 4.15:*** *Example of a HNCA strip that show backbone assignment of CW42 without ligand. The path between $C_\alpha$, and $C_\alpha^{-1}$ in the next residue is marked with a red line.*

The main work done in this thesis used the CWs construct. In addition, work was also done on the shorter and more stable construct CW42, with the aim of solving its ligand-bound structure and observe the contribution of the unstructured termini. The backbone assignment of CW42 without peptide was also based on the $^{1}H\,^{15}N$ HSQC (HSQC) assignment in the Hoppmann 2011 paper. However, since this is a truncated version of the CW domain, especially the terminal residues are expected to undergo changes in their chemical

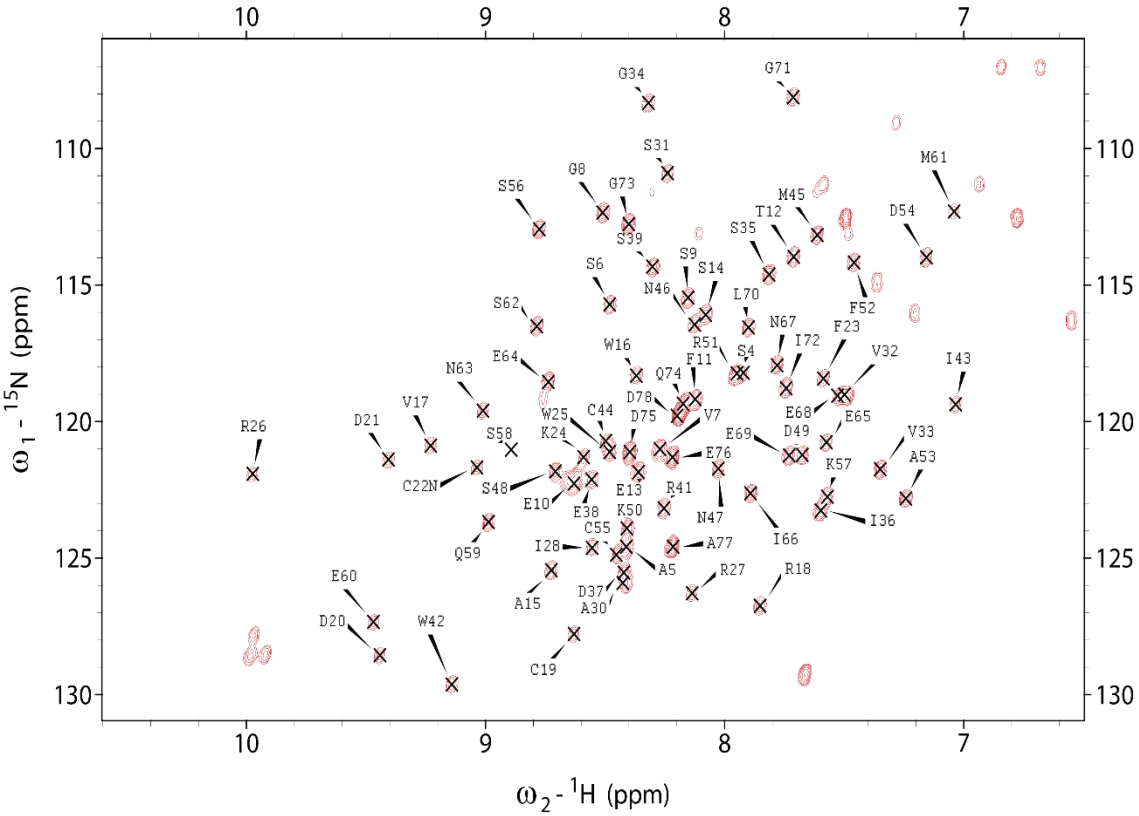


***Figure 4.16: Backbone assignment of a 1.5 mM $^{15}N, ^{13}C$ solution of CW42 without ligand.*** *Peak assignment indicated by crosshear on each peak in the HSQC spectrum.*

environments. In order to complete and validate assignment of CW42, a HNCA spectrum was used. In HNCA spectrums, a crosspeak from the preceding $C_\alpha$ in the sequence ($C_\alpha^{-1}$) can be seen in addition to the $C_\alpha$ itself. The $C_\alpha^{-1}$ is usually identifiable from its much weaker crosspeak intensity. A selected strip where $C_\alpha$ resonances at the same ppm value in the $^{13}C$ dimension is depicted in Figure 4.15, where the red line indicates matching ppm values between $C_\alpha^{-1}$ and $C_\alpha$ between successive residues. The full backbone amide assignment is shown in the HSQC depicted in Figure 4.16. Here it can be observed that only a few peaks have overlaps with others, and that also the CW42 construct has a separation of the peaks which are consistent with a folded protein. Most of the peaks, with the exceptions of G1, S2, S58, have a similar width and height. These peaks cannot be observed in the HSQC. The unassigned peak situated at $^{15}N$ 129.5 ppm and $^{1}H$ 7.7 ppm has a crosspeak at $^{13}C$ 53.7 ppm in the HNCA is suspected to be A79 since a suitable $C_\alpha^{-1}$ crosspeak at $^{13}C$ 54.1 ppm is found in D78. The chemical environment of this residue is drastically changed compared to the corresponding amino acid for CWs since it is the last residue in the sequence for CW42. Thus a large shift was expected for A79 in CW42 compared to CWs, making the assignment of CWs not transferrable to CW42 for this residue. This peak was not assigned, since only information on the $C_\alpha$ was available, and $C_\alpha$ shifts have a high degree of overlap between the residues.

## 4.6.2 Backbone assignment of CW42 with ligand

For CW42 with peptide there were more differences between the HSQC fingerprint in our data, compared to the CWs fingerprint in the Hoppmann 2011 paper. In order to solve and verify the backbone assignment for this, a CCCONH, and a CBCAHN was acquired in addition to a HNCA. In the CBCANH, intense $C_\alpha$, and $C_\beta$ crosspeaks can be found, in addition to $C_\alpha^{-1}$, and $C_\beta^{-1}$ of lesser intensity.

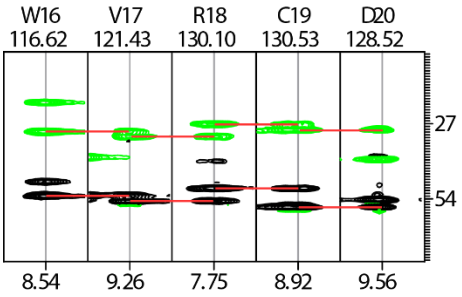


***Figure 4.17: CBCANH strip of CW42 in the presence of H3K4me1.*** *$C_\alpha$, and $C_\alpha^{-1}$ are colored black. $C_\beta$,* and $C_\beta^{-1}$ are colored green. *The path between the resonance from the amino acid, and from the n-1 amino acid is shown as red lines.*

In Figure 4.17, selected strips show how the residues were joined. In this assignment all amino acids except 1-4 were assigned. S907 (S58 in CW42) is visible in this spectra, although it appears weaker compared to the other peaks. Overall, there is not much overlap between adjacent peaks. The full assignment is shown on the HSQC depicted in Figure 18 where a larger

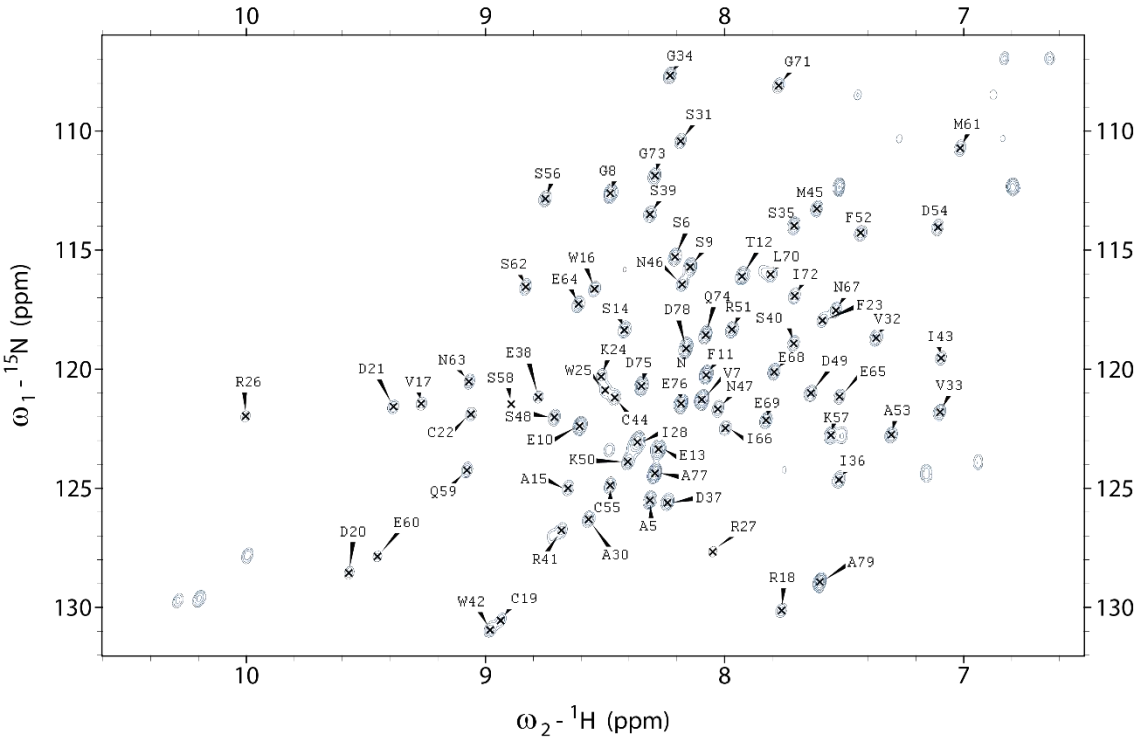


***Figure 4.18: Backbone assignment of 1.2 mM solution of $^{15}N$ and $^{13}C$ labelled CW42 with 5 molar excess of H3K4me1.*** *Peak selection is indicated by a crosshair on individual HSQC crosspeaks. S58 ($^{15}N$ 122 ppm, $^{1}H$ 8.7 ppm) is not visible at these contour level settings.*

separation of the peaks can be seen (compared to Figure 16), suggestive of more folded protein.

### 4.6.3 Change in chemical shifts of CW42 in presence and absence of ligand

Although performed also by Hoppmann et al. 2011 for CWs, we confirm the location of regions that are important for ligand binding by comparing the chemical shifts between CW42 in the presence and absence of ligand. The values from the δCSP formula shown in Equation 3.4 are shown in Figure 4.19, and they indicate which amino acids have changes in their chemical shifts of the backbone amide when the ligand is added. In general, there are relatively large changes. The largest changes are not in the secondary structure elements. Most of the amino acids that are higher than the average perturbation (dashed like Figure 4.19), are either positively or negatively charged amino acids. One of the cysteines involved in $Zn^{2+}$ coordination has a large shift. Interestingly, none of the tryptophan residues in the beta sheets have large perturbation upon binding. There is a general increase in changes near the C-terminal helix. In this region the changes are primarily in polar amino acids, including the negatively charged glutamine.

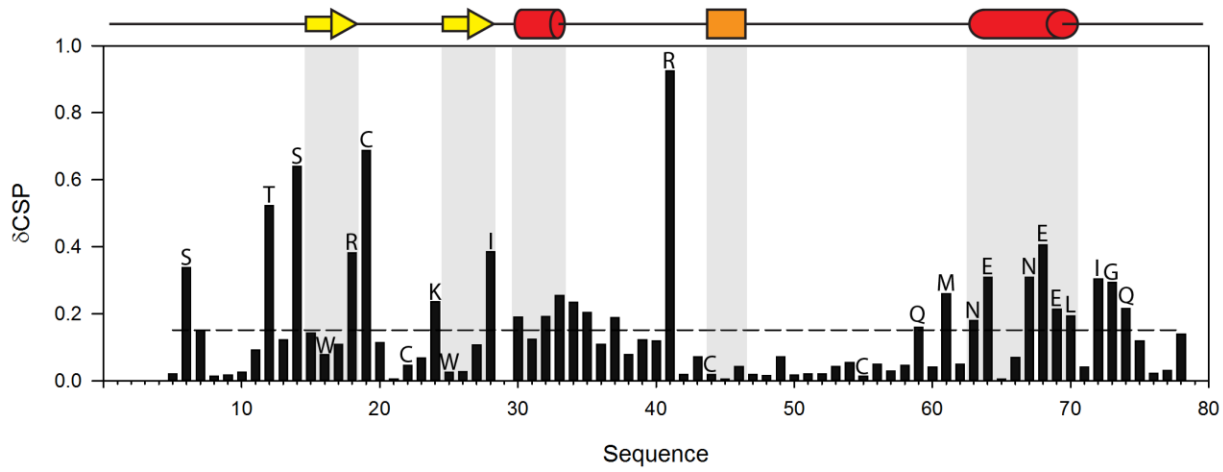


***Figure 4.19: Change in chemical shifts between CW42 in the presence and absence of ligand.*** *δCSP was calculated as shown in Equation 3.4. The chemical shifts of amino acid 1-4, and 79 where not assigned. The dotted line indicates the average height of the bars, and selected amino acid residues are shown above the bars. The secondary structure is shown above the graph and a shadow highlight structural elements in the graph.*

### 4.6.4 Stability of CW42 in solution over time

The NMR-data of CWs for mobility studies was accumulated on a sample that was kept and reused for 8 weeks. Furthermore, the CW42 sample that was prepared to make the solution structure has been stored and reused for 5 months and is still in use. To ascertain that the samples are stable, we regularly accumulate HN-HSQC experiments (before and after each major experiment). By automatic peak centring in SPARKY in comparison by calculating the δCSP (Equation 3.4) between the first, and the last HSQC experiment only small changes are observed during a 3 month time period (Figure 4.20). With an average δCSP of 0.0067 and the largest shifts at alanine 79 of 0.019 it can be concluded that the protein is stable over several months.

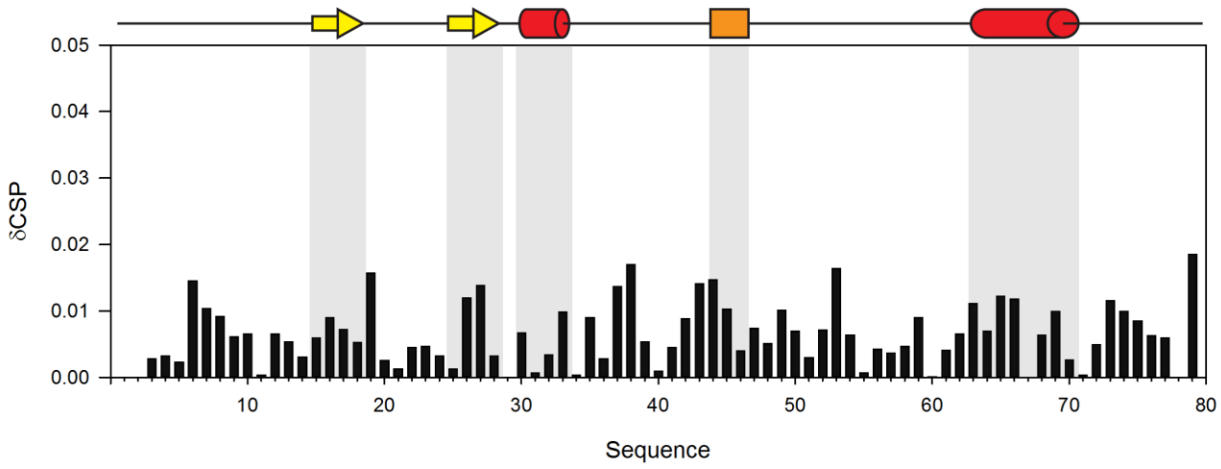


***Figure 4.20: Chemichal shift changes for CW42 after 3 month storage.*** *The sample was stored at 4 °C when not in use, and NMR data was acquired at 25°C. δCSP was calculated using Equation 3.4. Note the difference Y-axis scale difference between Figure 4.20 and 4.19.*

# 5 Discussion

The main purpose of this work has been to test the lid hypothesis as part of the molecular mechanism for histone tail binding by the ASHH2 CW domain. This was achieved by studying both global, and local motion of the CWs protein and by observing changes in the CWs, CW42 and CWΔLID protein's melting points as they interacted with H3K4me1 peptides using fluorescence. Moreover, it was also investigated how modifications on K9 affect binding strengths to the CWs domain. The main findings in this work are that the C-terminal helix is mobile in the unbound state and that this mobility is decreased upon binding. Binding also results in the protein adopting a more stable and compact conformation. When the C-terminal helix was removed, the stability of the protein was severely reduced. Furthermore, binding was not drastically affected by acetylation or dimethylation of K9 in addition to monomethylation of K4, but K9ac has a slightly negative effect on binding.

## 5.1 Binding parameters

The first 8 N-terminal residues of H3 are known to be important determinants for H3 binding. This study investigated the additional contribution of K9 in CWs. ITC data clearly shows that the combination of dimethylation at K9 and monomethylation at K4 on H3 (H3K4me1K9me2) has no significant effect on binding strength relative to the control, H3K4me1. For H3K4me1K9ac, a higher $K_D$ is observed showing that K9 acetylation has a small negative effect on binding. The acetylation of lysine residues neutralizes the positive charge of the ε nitrogen. In addition, the residue becomes a proton acceptor instead of a proton donor. This suggests that K9 also participates in the binding in a PTM-dependent manner, but that this is effect is both small and not essential for binding. The change in charge by acetylation may simply shift the histone tail slightly and introduce a small disfavourable rearrangement that may shifts e.g. cation-π interactions into a sub-optimal situation. Structural alignments of CWs with MORC3, ZCWPW1, and ZCWPW2 with their ligands bound, gives no clear answer as to where on CWs the K9 side chain would reside. However, it is most likely in the close vicinity of CWs, since acetylation of K9

has an effect on binding. When the solution structure is completed and available in full detail, a much better understanding of how the ligand resides in CWs can be achieved.

From the solution structures of CW domains, several residues besides K4 on the histone tail are important for the selective binding of K4 (e.g. A1, R2, T3, Q5, T6, and R9) (Figure 5.1). It appears that binding is negatively affected by acetylation without completely disrupting it. The effects of acetylation of K9 are relatively small, but in a cellular context, even small changes can have large effects. *In vitro* work with histone tails is usually performed on relatively short peptides compared to the full length histone tails (e.g. H3s histone tail is 27 aa long). Based on these finding further work should use peptides that also include K9 in order to be closer to the biological situation. It is tempting to suggest that by having this large amount of determinants that affect binding, the binding strengths is the sum of many small contributions, rather than being restricted to one or two interactions. This would explain how histone recognition modules are able to keep their specificity towards a certain residue in the myriad of modifications that alter their ligands. However, a much more systematic approach with many more combinations of PTMs should be tested in order to reach such a conclusion.

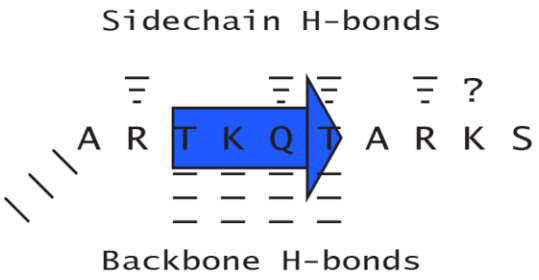


***Figure 5.1: Typical hydrogen bonding sites on H3 when it is recognized by CW domains*** *(and PHD fingers). The residues of H3 from A1-S10 are shown. (⋮) Hydrogen bonds between the side chains in the histone tail. (=) Hydrogen bonds between the backbone. These indicate that this histone is bound through beta augmentation. (⋰) Hydrogen bonds between the N terminal protons on A1 and CW.*

In this study, methylated histone-tail mimicking peptides were used and all the experiments were conducted *in vitro*. Whether this is directly applicable in a biological setting with whole histones in the crowded nucleus is uncertain. However, relating affinity to structural information is accepted as valid and relevant approach for a very wide range of biological phenomena. The focus here is to observe the changes that occur upon changing the methyl substitutions of the ligand, which are known to be related to epigenetic effects. All the peptides used in this study included K9 so that its relative contribution to binding is included in all our data.

In this study, the angle of the tryptophan residues on the beta sheet of three CW domains with affinity towards all three different methylation states, have been measured. The CW domains with specificity towards H3K4me1 (ASHH2), has a smaller angle that that with specificity towards H3K4me2 (MORC3), which again is smaller than the angles found in ZCWPW1 that has the highest affinity towards H3K4me3. Holmedal 2013 suggested that both the core, and the helix functions as determinants for binding. The angle of these tryptophan residues might be the contribution from the core.

## 5.2 The N- and C-terminals have a large effects on ASHH2 stability

The shorter ASHH2 protein, CW42, has a higher $T_m$ than the longer CWs, indicating that the lengths of the termini are important for stability. It is not clear why the longer CWs are less stable. One possibility is that this protein has a cysteine residue in the C- terminal which is absent in CW42. Cysteine residues are known to be thermolabile, since they tend to become oxidised at higher temperatures (Russell et al. 1997), thus being more susceptible towards the formation of disulphide bridges. However, it is unclear whether this would affect $T_m$ at all, or whether other effects contribute to a greater extent since this cysteine is in the unstructured region. Another possibility is that the C-terminal contributes unfavourably to the overall fold of the protein by means of transiently exposing hydrophobic parts, or competing with the core structure's bonds (such as hydrogen bonds and salt bridges). The protein also has a high degree of local motion in its core domain, so such exposure is possible. These suggestions are in agreement with statistical studies on the available structures in The Protein Data Bank that conclude that protein packing, amount of buried non-polar amino acids, and protein size are not key factors in giving a protein its $T_m$. Proteins with a high $T_m$ usually have a higher number of hydrogen bonds, salt bridges, more arginine and tyrosine residues and fewer cysteine and serine residues (Kumar et al. 2000).

## 5.3 The C-terminal helix is essential for the ASHH2 CW domain fold

While the shorter (CW42) of the two ASHH2 CW domain proteins are the most stable, complete removal of what is reported as a flexible region in the published structure results in no binding (Hoppmann et al. 2011). This could be attributed to both failure to bind and a failure to fold. We find that when the C-terminal helix is removed in CWΔLID, the core CW domain becomes unstructured, even at 4 °C. However, it does remain soluble at 4 °C for several hours. This shows that the helix is important for the stability of this CW domain, even though it is rather unstructured (residues has $S^2 < 0.8$). It should be noted that several steps in the purification are performed at room temperature, and since no re-naturation has been observed after thermal denaturation the actual $T_m$ might be as high as 25 °C. However, this does not alter the conclusion that the C-terminal helix is essential for the stability of the CW proteins. Other CW domains have also been reported to have low solubility and form aggregates when removed from their natural conditions (Zhang et al. 2013). This suggests that covering the binding site may be important. On the other hand, there are examples where such a cover is missing, such as in the case of ZCWPW1, which implies that this is not essential for all the subtypes of this domain. Therefore, it seems more likely that only certain CW domains require a covered binding site, and one suggestion is that CW domains with specificity towards H3K4me1 would be among those more likely to require such a "lid". These monomethylated histone-tail binding proteins require a mechanism for excluding the much larger dimethyl and trimethyl groups on lysine residues from the binding site to achieve selectivity. Since having the "lid" in the binding pocket makes the protein more stable, other parts of the protein may have acquired mutations that rendered the protein "lid dependent". This would be in agreement with our data and with the structural information available so far, but would require an extensive phylogenetic analysis in combination with experimental validations of the stability of many more CW domains in order to be certain. Our group has previously reported that this helix is important for binding, as truncated proteins where the C-terminal helix was removed did not bind to either of the methylation states of H3K4 (Hoppmann et al. 2011). In light of the results presented here, that

conclusion may not have been valid, since the protein was most likely not folded when the experiments were performed. It should be noted that intrinsic tryptophan fluorescence is not a direct measure of protein folding, but a measure of how solvent-accessible the tryptophan residues are. In light of this, a proton spectrum of CWΔLID and a natural abundance HSQC were acquired on freshly purified CWΔLID. The proton spectrum revealed a spectrum characteristic of an unfolded protein and no peaks were found for the HSQC. This was most likely because of very low sensitivity in combination with the observed aggregation during the ~23 hour long experiment at 25 °C.

## 5.4 The ASHH2 CW domain adopts a more stable conformation upon ligand binding

Upon binding, there is an increase in $T_m$ for both CWs and CW42, suggesting that the protein adopts a more stable conformation when the ligand is present. The differences between bound and unbound form are about the same for CWs and CW42 ($\Delta T_m$ of about 5 - 10 °C), which suggest that the truncation from CWs to CW42 has not affected the key aspects of ligand binding.

It should at this point be noted that, in contrast to CWΔLID, both CWs and CW42 have a stable fold at the NMR experimental temperature both with and without ligand. Exposure to 25 °C over prolonged periods of time had no effect: comparison of HSQCs taken both when the protein was fresh and several months later were nearly identical Figure 4.20.

## 5.5 Unfolding by $Zn^{2+}$ depletion is partially reversible

The post-transition states of CW domains that were depleted of their $Zn^{2+}$ ions were very similar to those obtained by thermal denaturation. This shows that $Zn^{2+}$ is a requirement for a stable fold. When excess $Zn^{2+}$ was added, I335/I355 ratios indicative of a native state were partially recovered, which suggests that $Zn^{2+}$ removal is at least partially reversible. Interestingly, CW42 has an almost complete recovery, which may be explained by the fact that this protein has a

more stable fold than CWs (measured with thermal denaturation). Another aspect may be that the extra-long termini in CWs may complicate refolding significantly compared to the shorter CW42.

## 5.6 Local mobility of CWs support the lid hypothesis for the ASHH2 CW domain

The lid hypothesis suggests that there must be a movement in the C-terminal helix in order for the CW domain to bind to histone tails. In general, the relaxation data suggest that the unstructured N-terminal can achieve more motion when the peptide is bound and that the C-terminal, including the C-terminal helix, becomes more stable after binding (Figure 5.2). This trend can be observed directly in $T_1$, $T_2$, and hNOE measurements (Figure 4.13 A-C) and the type of movement can be assigned through the model free parameters ($S^2$, $S_f^2$, $\tau_c$, $\tau_e$, and $\tau_s$) (Figure 4.13 D, and Figure 4.14 A-D).

The global parameters of CWs were calculated on the basis of residues that are within typical $T_1$, $T_2$, and hNOE values of amino acids in a fold. Only 23 amino acids fulfilled these criteria for the unbound protein, while in the bound configuration this number increased to 37 residues. This suggests that more amino acids contribute to the fold. This is consistent with the thermal denaturation data where a higher $T_m$ was observed in the presence of ligand. The global correlation time was increased by 0.7 ns ± 0.52 ns, which is consistent with a small increase in mass. Interestingly, a combination of gel filtration and multi-angle light scattering experiments have shown that the hydrodynamic radius decreased upon binding (unpublished data by Rein Aasland, University of Bergen and Valeria De Marco, The National Institute for Medical Research, London). This is in agreement with our data when the relative contributions of the termini are included.

A fundamental assumption in the model-free analysis is that the protein has either a global isotropic or anisotropic tumbling. This is determined based on the ratio between parallel, antiparallel and perpendicular NH vectors (Diffusion anisotropy = $D_{\parallel}/D_{\perp}$). Proteins with $D_{\parallel}/D_{\perp}<$ 1.2 are defined as isotropic, whereas proteins with $D_{\parallel}/D_{\perp}>$ 1.5 are considered to be anisotropic. For CWs $D_{\parallel}/D_{\perp}$= 1.24 in the unbound state and 1.20 in the presence of ligand which falls between these defined values. In the data presented in this thesis an isotropic model was therefore chosen, since no intermediate model exists and since both values are closer to the isotropic values. Although the absolute values of the parameters changed slightly when a anisotropic analysis was conducted, the same conclusions could be drawn also in this case.

The local parameters of motions described by $S^2$, $S_f^2$, $\tau_e$, and $\tau_s$ are in agreement with a mobile C-terminal helix that becomes more stable upon binding. By crossing the defined stability line ($S^2 > 0.8$), CWs has $S^2$ values in the C-terminal helix below what is normally considered to be a stable structure. In the presence of ligand, the $S^2$ parameter becomes higher than this threshold. This suggest a more stable configuration in the ligand-bound state and harmonizes with the fluorescence stability measurements in the presence and absence of ligand. As summarized in Figure 5.4 the changes in $S^2$ suggest a decrease in mobility for the C-terminal region, and an increase in mobility for the N-terminus.

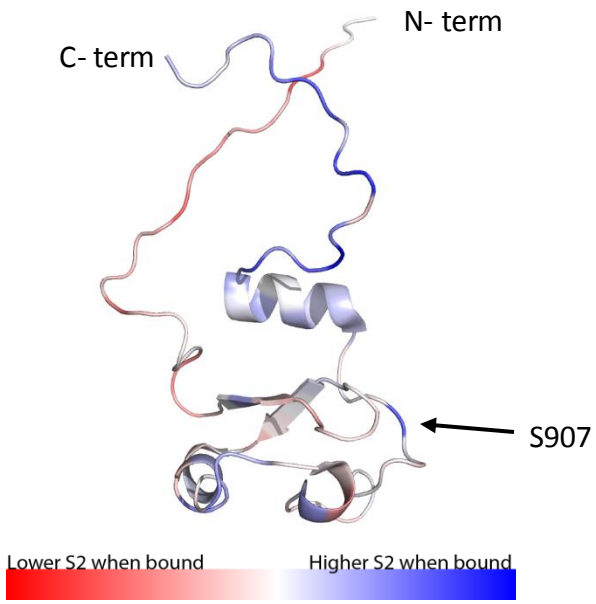


***Figure 5.2: Visual summary of changes in $S^2$ values plotted onto the ASHH2 CW domain (PDBID: 2L7p).*** *(Blue) decrease in local mobility (Red) increase in local mobility. S907 (S58 in CW42, S70 in CWs) in ASHH2 is marked with an arrow, and the N- and C- terminal are labelled.*

Furthermore the transition rates of ($R_{ex}$ of about 0.7 $s^{-1}$) for these residues are very roughly on the same timescale as the disassociation rates ($k_d$ = 0.015 $s^{-1}$) that have been found by surface

plasmon resonance (Unpublished data), in loose agreement with transitions of the same type. The region from residue number 43 to 53 also undergoes changes upon binding to the ligand and data suggests a transfer to a more stable configuration also for these residues. Regardless, this part of the domain is not believed to take direct part in the binding.

Models were manually selected to yield $S^2$-values for each amino acid, as shown in Figure 4.13 D. The simplest of the models (M1), only allows the $S^2$ parameter to change and is in most cases too simple to explain the protein dynamics of this backbone. The increasingly complex models introduce additional levels of movement, which gives rise to lower $S^2$ values, but also introduces larger sources of errors. The general trends and differences between the bound and unbound situations reported here are, however, consistent between all the models, with the possible exception of M1. This indicates that while it is important to be critical when choosing which model to use in a given instance, the overall results are robust.

In summary, the relaxation data are compatible with our model for binding of the CW domain to histone tails. Both termini are locally mobile as shown by $T_1/T_2$, hNOE and model free analysis. The C-terminal is the more mobile of the two when the peptide is not present. Addition of the peptide increases the local motion of the N-terminus, but reduces the local motion in the C-terminus, consistent with consolidation of the proposed lid-section across the ligand as it interacts with the binding pocket. The mobile mid-section of the protein is stabilized as the peptide binds, suggesting that this part may also be involved in the interaction, either directly or by allosteric means. The change in overall shape from almost spherical to slightly ellipsoid may be due to a reorientation of the termini; the N-terminus is expelled from the body of the protein, while the C-terminus is packed tighter to the body of the protein.

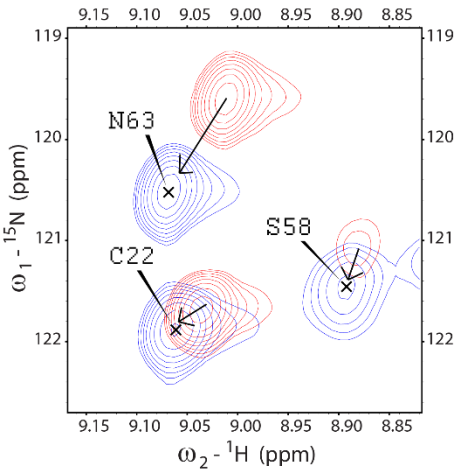


***Figure 5.3: Changes in S58 intensities upon ligand binding.*** *The figure displays an HSQC section of CW in presence (Blue) and absence (Red) of H3k4me1. Note the intensity difference of S58 as the ligand binds. Arrows indicate changes in chemical shift values upon binding.*

There are some residues that display interesting features that have not been addressed in detail in previous data. Of particular interest is serine 907 in full length ASHH2 (S58 in CW42 and S70 in CWs) which has a very weak signal when the ligand is not present (Figure 5.3). There are several factors in a sample that can contribute to a weaker signal, including a high degree of amide proton exchange upon binding or a transition between different conformational states. Using the SPHERE webserver no large proton exchange is predicted for S907, thus suggesting that this cannot account for the large effect. It should be noted that this server computes water exchange rates based on the nearest residue only and does not take 3D structure into account. More interestingly, the signal increased in intensity upon binding (Figure 5.5), implying that binding has a large effect on this residue despite the low δCSP value of 0.06. This suggests that this residue has more than one conformational state in the unbound conformation, but that the residue conforms towards one state upon binding. However, the chemical environment is similar in both the bound and the unbound state. Since the localization of this residue is not in the direct vicinity of the binding site (arrow in Figure 5.2) it may simply act as a "hinge" that allow the C-terminal helix to move relative to the core domain.

## 5.7 NMR solution structure

The main goal of this thesis has been to look at the changes in local motion of the CW domain as it is bound by histone-mimicking peptides. Many of the conclusions that have been reached in this thesis would benefit tremendously by having a solution structure with an incorporated ligand, as this would show details of CW-ligand interaction and the proposed reorganisation of the protein domain structure, including the lid and central regions. Such a structural characterization was initially not part of this master project. Despite this, the work to produce a structural model in the presence of the ligand was started during the final stages of this thesis. While backbone and side-chain assignment for this structure are available, the work needs to implement filtered-type NOESY experiments that separate NOEs from protons coupled to $^{15}N$ and $^{13}C$ nuclei from those that are not. The final structure is thus not completed since the NOE restraints are not yet available.

# 6 Conclusion

The work in this thesis has been aimed towards understanding the determinants for binding of the CW domain in ASHH2 and the data are in agreement with the lid hypothesis. The results presented here show that the C-terminal helix has a local motion that can give space for the ligand in the binding pocket. Furthermore this work implicates K9 as an additional contributor towards this binding, since a small but significant decrease in binding was achieved by its acetylation.

Based on our findings, we suggest that the ASHH2 CW C-terminal region functions as a lid that provides specificity towards H3K4me1, and provide protection for the binding pocket from the solvent in the absence of binding partner. The helix has to move away to make room for the ligand upon binding, since there is not enough space for the monomethyl lysine between the helix and W865 and W874. This energetically unfavourable transition may be achieved through a general stabilization of the whole CW domain as the correct ligand binds, which results in an energetically favourable reaction overall, indicated by an increase in $T_m$. This may be a key element in the peptide binding, since the energetically favourable helix docking in the pocket has to be disrupted upon binding. However, this helix is not only important for specificity (Our group has previously shown that when the C-terminal regions of the ASHH2 and the ZCWPW1 domain were swapped, this resulted in semi-functioning protein (Holmedal 2013)). When it is removed, the protein becomes prone to aggregation. This suggests that the ASHH2 CW domain requires a C-Terminal stabilizing region in order to stay folded.

Given the structural changes in the ASHH2 CW domain that occur during ligand binding, it is tempting to speculate that this could translate to an effect on the adjacent SET domain (methyltransferase) of ASHH2, thus participating in the regulation of the activity of the ASHH2 enzyme.

# 7 Future work

This study demonstrates the importance and contributions of the CW nonconserved C-terminal region upon ligand interaction. Specifically we demonstrate that the C-terminal helix in ASHH2 CW has mobility enabling interaction with monomethylated H3K4.This can partially explain several aspects of our models for CW domain binding. However, there are still many unanswered questions.

Determinants restricting binding of monomehylated lysine is then of particular interest. In this study, an interesting observation was made on the angle of the tryptophan residues. There appears to be a 5 degree larger angle between the tryptophan residues in MORC3 than ASHH2 and a 5 degree larger angle between the tryptophan residues in ZCWPW1 than in MORC3. This could thus provide an additional selection criterion in our model for binding. However, this analysis has only been conducted on these three proteins, therefore many more should be included in order to draw such conclusions. Furthermore, the resolution of these structures gives the possibility for misinterpretations, something that could have been eliminated in ultra-high resolution crystal structures, or by evaluation of the side chain dynamics of these residues by NMR.

A major motivation towards gaining insight into the binding mechanisms of HRMs is to use them as molecular probes that can, with the correct fusion proteins, be used in epigenetic engineering tools. These could be used as fusion proteins with enzymes of interest and could potentially be used to target e.g. methyl transferases to certain regions like enhancers or promoters, and observe the effects of it. However, CW domain mutants and fusion proteins have been shown to give unexpected results, including aggregates (Zhang et al. 2013), or unexpected effects between different cell types (Personal communication with Prof. Reidun Aalen, University of Oslo). By using the biophysical approach to understand more of the molecular determinants for binding for all CW domain types, more reliable tools can be engineered.

The C-terminal regions of CW domains have been shown to be important for binding, both when this region is unstructured, like it is for ZCWPW1 and MORC, and when this region has secondary structure, such as in the case of ASHH2. A phylogenetic study of the development of these regions could provide insight into how these domains acquired their high degree of affinity and specificity.

The effects of combinatorial PTMs still remains hard to predict. By performing e.g. ITC or NMR titrations of many combinations of differently modified peptides in to solutions of CWs, a better understanding of how these PTMs affect binding can be obtained. NMR could be conducted on peptides uniformly labelled with $^{15}N$, in order to observe the effects that binding has on the histone tail backbone. This can also give information about how CW and PHD domains achieve their specificity towards the "correct" lysine.

The approaches that have been used in this study rely on knowing which ligand is preferred for the domain in question. When a new domain is identified the preferred ligand has to be identified, and the binding constant must be determined. Today this is a cumbersome process. By developing a quick and convenient assay involving tools such as Surface plasmon resonance (SPR), one could screen many CW domains fast and easy by running them over immobilised peptide arrays. Furthermore, a novel technique that uses libraries of "DNA-barcoded" nucleosomes with distinct combinations of PTMs has recently been shown to be able to investigate combinatorial PTMs very rapidly (Nguyen et al. 2014).

# 8 References


Aasland, R., Gibson, T.J. & Stewart, A.F., 1995. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends in biochemical sciences*, 20(2), pp.56–9. Available at: http://www.sciencedirect.com/science/article/pii/S0968000400889574 [Accessed September 7, 2014].

Allfrey, G., Faulkner, R. & Mirsky, A.E., 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *PNAS*, 315(1938), pp.786–794. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=300163&tool=pmcentrez&rendertype=abstract [Accessed August 24, 2014].

Armstrong, L., 2014. *Epigenetics*, Taylor & Francis Group. Available at: http://books.google.com/books?id=sMnwLQAACAAJ&pgis=1 [Accessed October 24, 2014].

Bannister, A.J. et al., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824), pp.120–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11242054 [Accessed November 18, 2014].

De Beer, T.A.P. et al., 2014. PDBsum additions. *Nucleic acids research*, 42(Database issue), pp.D292–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965036&tool=pmcentrez&rendertype=abstract [Accessed October 3, 2014].

Berger, S.L. et al., 2009. An operational definition of epigenetics. *Genes & development*, 23(7), pp.781–3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3959995&tool=pmcentrez&rendertype=abstract [Accessed November 6, 2014].

Brogaard, K. et al., 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 486(7404), pp.496–501. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3786739&tool=pmcentrez&rendertype=abstract [Accessed November 16, 2014].

Dhalluin, C. et al., 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, 399(6735), pp.491–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10365964 [Accessed September 7, 2014].

Dinkel, H. et al., 2014. The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic acids research*, 42(Database issue), pp.D259–66. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3964949&tool=pmcentrez&rendertype=abstract [Accessed October 17, 2014].

Doucleff, M., Hatcher-Skeers, M. & Crane, N.J., 2011. *Pocket Guide to Biomolecular NMR*, Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: http://link.springer.com/10.1007/978-3-642-16251-0 [Accessed October 25, 2014].

Eltsov, M. et al., 2008. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences of the United States of America*, 105(50), pp.19732–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2604964&tool=pmcentrez&rendertype=abstract [Accessed November 15, 2014].

Everid, A.C., Small, J. V & Davies, H.G., 1970. Electron-microscope observations on the structure of condensed chromatin: evidence for orderly arrays of unit threads on the surface of chicken erythrocyte nuclei. *Journal of cell science*, 7(1), pp.35–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5476861 [Accessed November 6, 2014].

Flemming, W., 1882. *Zellsubstanz, kern und zelltheilung*, Leipzig: Leipzig, F. C. W. Vogel. Available at: http://books.google.com/books?id=ndYcngEACAAJ&pgis=1 [Accessed October 22, 2014].

Gasteiger, E. et al., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker, ed. *The Proteomics Protocols Handbook*. Totowa: Humana Press Inc, pp. 571–607. Available at: http://web.expasy.org/protparam/protpar-ref.html [Accessed October 30, 2014].

Goddard, T.D. & Kneller, D.G., 2007. SPARKY 3.

Haynes, S.R. et al., 1992. The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic acids research*, 20(10), p.2603. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=312404&tool=pmcentrez&rendertype=abstract [Accessed September 7, 2014].

He, F. et al., 2010. Structural insight into the zinc finger CW domain as a histone modification reader. *Structure (London, England : 1993)*, 18(9), pp.1127–39. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20826339 [Accessed August 26, 2014].

Higman, V.A., 2012. Protein NMR: A Practical Guide. Available at: http://www.protein-nmr.org.uk/ [Accessed October 30, 2014].

Holmedal, M.K., 2013. *Master Thesis - The Molecular Basis for Ligand Binding Specificity of the CW Domain*. University of Bergen.

Hoppmann, V. et al., 2011. The CW domain, a new histone recognition module in chromatin proteins. *The EMBO journal*, 30, pp.1939–1952.

Kossel, A., 1911. *Ueber die chemische Beschaffenheit des Zellkerns....*, Available at: http://books.google.com/books?id=bunImAEACAAJ&pgis=1 [Accessed October 22, 2014].

Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17320507 [Accessed October 20, 2014].

Kumar, S., Tsai, C.J. & Nussinov, R., 2000. Factors enhancing protein thermostability. *Protein engineering*, 13(3), pp.179–91. Available at: http://peds.oxfordjournals.org/cgi/doi/10.1093/protein/13.3.179 [Accessed November 9, 2014].

Lachner, M. et al., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410(6824), pp.116–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11242053 [Accessed October 18, 2014].

Li, H. et al., 2006. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, 442(7098), pp.91–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16728978 [Accessed October 30, 2014].

Lipari, G. & Szabo, A., 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society*, 104(17), pp.4546–4559. Available at: http://pubs.acs.org/doi/abs/10.1021/ja00381a009 [Accessed July 18, 2014].

Liu, Y.. T.W.. D.A.. B.C.. A.C.H.. E.A.M.. M.J.. S.G.C. (SGC), CW-type zinc finger of MORC3 in complex with the amino terminus of histone H3. *To be Published*. Available at: http://www.rcsb.org/pdb/explore/explore.do?structureId=4QQ4 [Accessed November 17, 2014].

Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648), pp.251–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9305837 [Accessed July 17, 2014].

Maeshima, K., Hihara, S. & Eltsov, M., 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Current opinion in cell biology*, 22(3), pp.291–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20346642 [Accessed November 15, 2014].

Mannhold, R. et al., 2006. *BioNMR in Drug Research*, John Wiley & Sons. Available at: http://www.google.no/books?hl=en&lr=&id=s7cIuiDnik8C&pgis=1 [Accessed September 20, 2014].

Marshall, M.S. et al., 2009. Potential energy curves for cation-pi interactions: off-axis configurations are also attractive. *The journal of physical chemistry. A*, 113(48), pp.13628–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19886621 [Accessed November 11, 2014].

Martin, D.G.E. et al., 2006. The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Molecular and cellular biology*, 26(21), pp.7871–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1636756&tool=pmcentrez&rendertype=abstract [Accessed November 20, 2014].

Mellor, J., 2006. It takes a PHD to read the histone code. *Cell*, 126(1), pp.22–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16839870 [Accessed November 19, 2014].

Munks, R.J. et al., 1991. Histone H4 acetylation in Drosophila. Frequency of acetylation at different sites defined by immunolabelling with site-specific antibodies. *FEBS letters*, 284(2), pp.245–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2060643 [Accessed September 11, 2014].

Murray, K., 1964. THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES. *Biochemistry*, 3, pp.10–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14114491 [Accessed September 5, 2014].

Nelson, D.L., Lehninger, A. & Cox, M.M., 2008. *Lehninger principles of biochemistry* 5th ed., W. H. Freeman. Available at: http://books.google.no/books?id=YbcVAAAAQBAJ [Accessed November 6, 2014].

Nguyen, U.T.T. et al., 2014. Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. *Nature methods*, 11(8), pp.834–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24997861 [Accessed September 26, 2014].

Nyborg, J.K. & Peersen, O.B., 2004. That zincing feeling: the effects of EDTA on the behaviour of zinc-binding transcriptional regulators. *The Biochemical journal*, 381(Pt 3), pp.e3–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1133908&tool=pmcentrez&rendertype=abstract [Accessed July 15, 2014].

Olins, D.E. & Olins, A.L., 2003. Chromatin history: our view from the bridge. *Nature reviews. Molecular cell biology*, 4(10), pp.809–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14570061 [Accessed October 22, 2014].

Oudet, P., Gross-Bellard, M. & Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4), pp.281–300. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1122558 [Accessed September 12, 2014].

Pace, C.N. & Shaw, K.L., 2000. Linear extrapolation method of analyzing solvent denaturation curves. *Proteins*, Suppl 4(February), pp.1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11013396 [Accessed October 1, 2014].

Palmer III, A.G. et al., 2005. *Protein NMR Spectroscopy: Principles and Practice* 2nd ed., Academic Press. Available at: http://www.amazon.com/Protein-NMR-Spectroscopy-Second-Edition/dp/012164491X [Accessed November 4, 2014].

Paro, R. & Hogness, D.S., 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 88(1), pp.263–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=50790&tool=pmcentrez&rendertype=abstract [Accessed November 16, 2014].

Peña, P. V et al., 2006. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature*, 442(7098), pp.100–3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3190580&tool=pmcentrez&rendertype=abstract [Accessed September 7, 2014].

Perry, J. & Zhao, Y., 2003. The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. *Trends in biochemical sciences*, 28(11), pp.576–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14607086 [Accessed October 7, 2014].

Phillips, D.M., 1963. The presence of acetyl groups of histones. *The Biochemical journal*, 87(1958), pp.258–63. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1201885&tool=pmcentrez&rendertype=abstract [Accessed October 26, 2014].

Prunell, A. & Kornberg, R.D., 1978. Relation of nucleosomes to DNA sequences. *Cold Spring Harbor symposia on quantitative biology*, 42 Pt 1, pp.103–8. Available at: http://symposium.cshlp.org/content/42/103.short [Accessed September 5, 2014].

Ragvin, A. et al., 2004. Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *Journal of molecular biology*, 337(4), pp.773–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15033350 [Accessed October 25, 2014].

Remaut, H. & Waksman, G., 2006. Protein-protein interaction through beta-strand addition. *Trends in biochemical sciences*, 31(8), pp.436–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16828554 [Accessed September 24, 2014].

Rochus, L.J.K., 2004. *The Computer Aided Resonance Assignment Tutorial*, Zürich: CANTINA Verlag.

Russell, R.J. et al., 1997. The crystal structure of citrate synthase from the hyperthermophilic archaeon pyrococcus furiosus at 1.9 A resolution,. *Biochemistry*, 36(33), pp.9983–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9254593 [Accessed November 9, 2014].

Ruthenburg, A.J. et al., 2007. Multivalent engagement of chromatin modifications by linked binding modules. *Nature reviews. Molecular cell biology*, 8(12), pp.983–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18037899 [Accessed October 30, 2014].

Santiveri, C.M. et al., 2008. The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *Journal of molecular biology*, 382(5), pp.1107–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18706910 [Accessed November 3, 2014].

Schindler, U., Beckmann, H. & Cashmore, A.R., 1993. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *The Plant journal : for cell and molecular biology*, 4(1), pp.137–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8106082 [Accessed September 7, 2014].

Schomburg, I. et al., 2004. BRENDA, the enzyme database: updates and major new developments. *Nucleic acids research*, 32(Database issue), pp.D431–3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=308815&tool=pmcentrez&rendertype=abstract [Accessed October 9, 2014].

Segal, E. et al., 2006. A genomic code for nucleosome positioning. *Nature*, 442(7104), pp.772–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2623244&tool=pmcentrez&rendertype=abstract [Accessed October 23, 2014].

Shi, X. et al., 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature*, 442(7098), pp.96–9. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3089773&tool=pmcentrez&rendertype=abstract [Accessed November 20, 2014].

Song, F. et al., 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science (New York, N.Y.)*, 344(6182), pp.376–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24763583 [Accessed November 16, 2014].

Strømland, Ø., 2012. *Master Thesis - Specificity of CW domains for modified histone peptides*. University of Bergen.

Tan, M. et al., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, 146(6), pp.1016–28. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3176443&tool=pmcentrez&rendertype=abstract [Accessed September 8, 2014].

Taverna, S.D. et al., 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology*, 14(11), pp.1025–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17984965 [Accessed July 14, 2014].

Turner, B.M., 1993. Decoding the nucleosome. *Cell*, 75(1), pp.5–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8402900 [Accessed September 11, 2014].

Turner, B.M., 2007. Defining an epigenetic code. *Nature cell biology*, 9(1), pp.2–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17199124 [Accessed October 6, 2014].

Ulrich, E.L. et al., 2008. BioMagResBank. *Nucleic acids research*, 36(Database issue), pp.D402–8. Available at: http://nar.oxfordjournals.org/content/36/suppl_1/D402.full?keytype=ref&ijkey=sOYRmrTXmKp5lxr [Accessed August 4, 2014].

Waterhouse, A.M. et al., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, 25(9), pp.1189–91. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2672624&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Williamson, R. a et al., 1997. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry*, 36(45), pp.13882–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9374866.

Wong, H., Victor, J.-M. & Mozziconacci, J., 2007. An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by the nucleosomal repeat

length. *PloS one*, 2(9), p.e877. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1963316&tool=pmcentrez&rendertype=abstract [Accessed November 7, 2014].

Woodcock, C.L., Safer, J.P. & Stanchfield, J.E., 1976. Structural repeating units in chromatin. I. Evidence for their general occurrence. *Experimental cell research*, 97(1), pp.101–10. Available at: http://linkinghub.elsevier.com/retrieve/pii/0014482776906595 [Accessed October 23, 2014].

Wu, C., Bassett, A. & Travers, A., 2007. A variable topology for the 30-nm chromatin fibre. *EMBO reports*, 8(12), pp.1129–34. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1963316&tool=pmcentrez&rendertype=abstract [Accessed November 6, 2014].

Wysocka, J. et al., 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442(7098), pp.86–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16728976 [Accessed November 18, 2014].

Yap, K.L. & Zhou, M.-M., 2010. Keeping it in the family: diverse histone recognition by conserved structural folds. *Critical reviews in biochemistry and molecular biology*, 45(6), pp.488–505. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2988946&tool=pmcentrez&rendertype=abstract [Accessed November 1, 2014].

Ye, Y. & Godzik, A., 2004. FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic acids research*, 32(Web Server issue), pp.W582–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=441568&tool=pmcentrez&rendertype=abstract [Accessed October 30, 2014].

Ye, Y. & Godzik, A., 2005. Multiple flexible structure alignment using partial order graphs. *Bioinformatics (Oxford, England)*, 21(10), pp.2362–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15746292 [Accessed November 11, 2014].

Zhang, Q. et al., 2013. Structure-function analysis reveals a novel mechanism for regulation of histone demethylase LSD2/AOF1/KDM1b. *Cell research*, 23(2), pp.225–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3567814&tool=pmcentrez&rendertype=abstract [Accessed August 20, 2014].

Zhang, Y.-Z., 1995. *Protein and Peptide Structure and Interactions Studied by Hydrogen Exchange and NMR*. University of Pennsylvania. Available at: http://books.google.no/books/about/Protein_and_Peptide_Structure_and_Intera.html?id=p-KcNwAACAAJ&pgis=1 [Accessed September 22, 2014].