The FEBS Journal

FEBS PRESS science publishing by scientists

The *Arabidopsis* (ASHH2) CW domain binds monomethylated K4 of the histone H3 tail through conformational selection

Olena Dobrovolska[1], Maxim Brilkov[1], Noelly Madeleine[1,2], Øyvind Ødegård-Fougner[3], Øyvind Strømland[2], Stephen R. Martin[4], Valeria De Marco[5], Evangelos Christodoulou[4], Knut Teigen[2], Johan Isaksson[6], Jarl Underhaug[7], Nathalie Reuter[7], Reidunn B. Aalen[8], Rein Aasland[8] and Øyvind Halskau[1]

1 Department of Biological Sciences, University of Bergen, Norway
2 Department of Biomedicine, University of Bergen, Norway
3 Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany
4 Structural Biology Science Technology Platform, Francis Crick Institute, London, UK
5 King's College London, UK
6 Department of Chemistry, The Arctic University of Tromsø, Norway
7 Department of Chemistry, University of Bergen, Norway
8 Department of Biosciences, University of Oslo, Norway

Keywords
conformational selection; CW; dynamics; histone-tail binding; induced fit; methylation; structure

Correspondence
R. Aasland, Department of Biosciences, University of Oslo, Post Box 1066 Blindern, 0316 Oslo, Norway
Tel: +47 22 84 45 63
E-mail: rein.aasland@ibv.uio.no
and
Ø. Halskau, Department of Biological Sciences, University of Bergen, N-5020 Bergen, Norway
Tel: +47 55 58 45 63
E-mail: oyvind.halskau@uib.no

(Received 9 July 2019, revised 17 December 2019, accepted 20 February 2020)

doi:10.1111/febs.15256

Chromatin post-translational modifications are thought to be important for epigenetic effects on gene expression. Methylation of histone N-terminal tail lysine residues constitutes one of many such modifications, executed by families of histone lysine methyltransferase (HKMTase). One such protein is ASHH2 from the flowering plant *Arabidopsis thaliana*, equipped with the interaction domain, CW, and the HKMTase domain, SET. The CW domain of ASHH2 is a selective binder of monomethylation at lysine 4 on histone H3 (H3K4me1) and likely helps the enzyme dock correctly onto chromatin sites. The study of CW and related interaction domains has so far been emphasizing lock–key models, missing important aspects of histone-tail CW interactions. We here present an analysis of the ASHH2 CW-H3K4me1 complex using NMR and molecular dynamics, as well as mutation and affinity studies of flexible coils. β-augmentation and rearrangement of coils coincide with changes in the flexibility of the complex, in particular the η1, η3 and C-terminal coils, but also in the β1 and β2 strands and the C-terminal part of the ligand. Furthermore, we show that mutating residues with outlier dynamic behaviour affect the complex binding affinity despite these not being in direct contact with the ligand. Overall, the binding process is consistent with conformational selection. We propose that this binding mechanism presents an advantage when searching for the correct post-translational modification state among the highly modified and flexible histone tails, and also that the binding shifts the catalytic SET domain towards the nucleosome.

Abbreviations
H3, Histone H3; H3K4meX, histone-tail peptide methylated 0-3 times at position K4; MALS, multi-angle light scattering; MD, molecular dynamics; PTM, post-translational modification; RMSF, root mean square fluctuation; SEC, size-exclusion chromatography; WT, wild-type.

Databases

Structural data are available in the PDB database under the accession code 6QXZ. Resonance assignments for CW42 in its *apo*- and *holo*-forms are available in the BMRB database under the accession code 27251.

Introduction

Chromatin structure, and thereby gene expression, is dynamically regulated by post-translational modifications (PTMs) on the N-terminal histone tails protruding from nucleosomes. These PTMs include methylation, acetylation, phosphorylation and many other modifications. They are thought to constitute a histone code, where unique combinations of PTMs are associated with specific effects on gene expression [1]. The PTMs are established and altered by 'writer' and 'eraser' enzymes that add and remove modifications, respectively, and the ensuing pattern of PTMs on the histone tails is interpreted by 'reader' protein domains [2]. Methylation of histone N-terminal tail lysine residues is carried out by methyltransferases that harbour a catalytic SET domain, and target lysine residues can either be mono-, di- or trimethylated on the ε-nitrogen [3]. The modified lysine residue can be demethylated by one of two classes of lysine demethylases, either a flavin adenine dinucleotide-dependent oxidase or a Fe (II) and α-ketoglutarate-dependent hydroxylase [4]. Methylated lysine residues can be recognized by members of the 'royal family' of protein domains, which are the chromo, MBT, chromo barrel, Tudor and PWWP domains [5]. It is also known that some WD40 domains and PHD fingers can recognize unmodified or methylated lysine residues [6]. The CW domain family has also been identified as another family of proteins that can recognize methylated lysine residues both in animals [7] and in plants [8].

The CW domain family is named after and identified by conserved cysteine and tryptophan residues found in its primary structure. Proteins containing the domain have been found in higher-order plants, vertebrates and vertebrate-infecting parasites [9,10]. The CW domain is found in proteins in combination with other domains such as PWWP and SET, and it has also been identified in chromatin remodellers and demethylases [7–13]. The role of the CW domain in most proteins is to recognize and bind to methylated histone H3 (H3) N-terminal tails at the K4 position (H3K4meX, where X is the number of methyl groups). Depending on the protein, the CW domain displays a different specificity for the degree of methylation [7,8,14]. The other mammalian CW domain-containing proteins ZCWPW1, ZCWPW2, MORC3 and MORC4 display specificity for H3K4me2/me3 [7,11]. CW containing multidomain proteins found in animals and plants are not orthologues, and their overall domain organizations are different [9].

The small, flowering plant *Arabidopsis thaliana* codes for an enzyme named ASHH2 which methylates position K36 on H3. This 1759-amino acid-long enzyme contains a CW domain that binds specifically to monomethylated H3K4 followed by an AWS domain and then a SET domain where the methyltransferase activity resides [8]. ASHH2 is a major regulator of growth and development in *Arabidopsis,* as mutations in ASHH2 result in dwarf plants with alterations in flowering time, fertility, branching, organ identity, programmed cell death and pathogen defence [15,16]. ASHH2 di- and trimethylates H3K36 from their monomethylated state, and in loss-of-function mutant plants, a global reduction in H3K36me3/me2 and a corresponding increase in H3K36me1 are linked to an early flowering phenotype [17,18]. Further pleiotropic effects include reduced fertility as well as homeotic changes in floral organs in plants where the ASHH2 gene is mutated [19,20]. *Arabidopsis* also contains another H3K36 methyltransferase ASHH1, but in contrast to ASHH2, this protein lacks a CW domain and based on the severe pleiotropic effects of the ASHH2 mutant the two proteins are not redundant [18].

In recent years, several structures of CW domains in their *apo* and *holo*, that is their unbound and bound states, have been solved [7,8,14,21]. A shared feature of interaction is the conserved tryptophans scaffolded by a β-sheet which provides part of the pocket that accepts the methylated lysine. Another feature of CW domains not highlighted by earlier investigations is the fact that their tertiary structure comprises just a few short secondary structure elements, while flexible coils dominate the rest of the fold. Disorder and flexibility are prevalent both in histone tails and in proteins involved in chromatin remodelling, and recent bioinformatic studies have highlighted the need for investigations focusing on functional flexibility [22,23]. As far as we can determine, the structural biology of CW domains has not been investigated systematically with functional flexibility in mind. We are therefore interested in whether CW binding is coupled to changes in structure, stability and mobility at the level of individual amino acids, and secondary and tertiary levels of organization. To explore this question, a comprehensive structural and dynamic analysis of the ASHH2 CW-H3K4me1 complex using NMR, molecular dynamics (MD) and lower-resolution techniques were performed, followed by mutagenesis of residues implicated in functional flexibility to assess their effect on affinity. From our analysis, ASHH2 CW emerges as a dynamic domain that undergoes a global reorganization to become more compact but still remains relatively flexible. We found that the mechanism of binding relies on protein flexibility and is best described by a conformational search for the correct histone modification. CW domains have not yet been reported to act through such mechanisms, and it is possible that this mechanism confers an advantage in the highly complex and dynamic chromatin environment.

Results

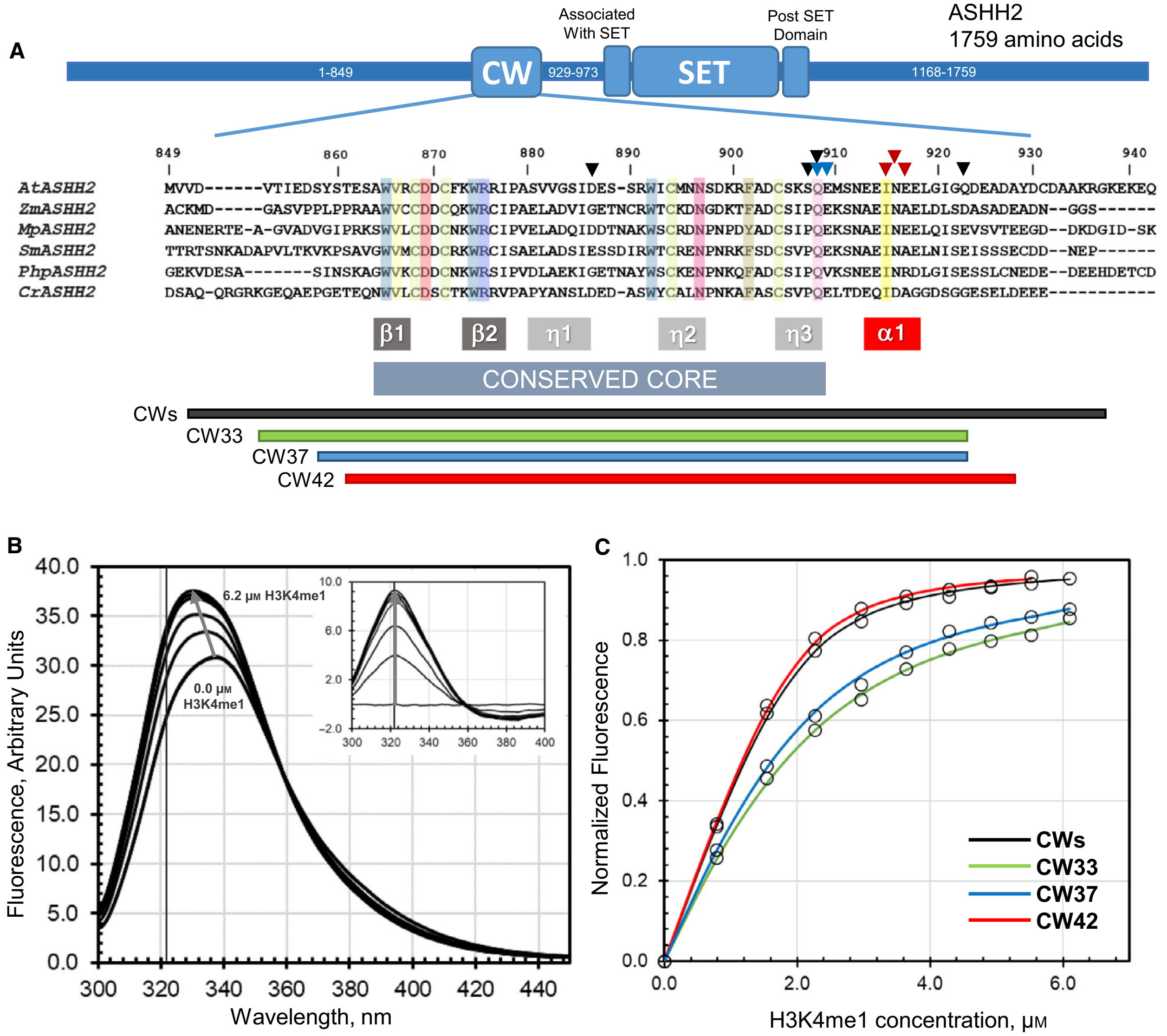
The CW domain reorganizes to a more compact form upon binding to H3K4me1

The functional domains of ASHH2 lie within long stretches of amino acids that are predicted to be disordered or to contain orphan secondary structure not associated with any known fold. Among the folded domains, CW is N-terminally situated and is also flanked by disordered segments (Fig. 1A). The NMR structure of the unbound ASHH2 CW domain was determined by Hoppmann *et al.* [8] using a construct denoted CWs (PDB code: 2L7P, Fig. 1A). For structure determination of the complex, a screen of 20 additional constructs was subsequently performed, initially aimed at finding expressible and high-affinity binders amenable for co-crystallization with the bound ligand. Of these, the constructs denoted CW33, CW37 and CW42, all covering the evolutionary conserved residues of the CW domain (Fig. 1A), where high-affinity binders of the histone tail mimick H3K4me1 (ARTKme1QTARY, with one methyl substitution at the ζ-position of K4), as determined by an intrinsic fluorescence-based binding assay (K_ds in the range of 0.21–0.85 μM Fig. 1B-D). Their affinities for H3K4me2 and H3K4me3 were also determined (sequence as for H3K4me1, but with two and three methyl substitutions at K4, respectively), and the K_d values ranged from 0.7 to 7.3 μM (Fig. 1D, examples of binding curves in Fig. S1). Several crystallization attempts were unsuccessful, and consequently, the decision was made to characterize the complex using NMR. For subsequent work, CW42 construct was selected as it expressed well and had an affinity indistinguishable from that of the longer CWs [24].

One noticeable property of the fluorescence binding studies was that the ligand caused a λ_{max} shift towards shorter wavelengths of the spectrum (Fig. 1B). Such behaviour is characteristic of tryptophans entering a more solvent-protected environment within a fold [25]. This makes sense if the ligand covers the two tryptophans of the binding site upon binding, if there is a consolidation of the overall fold of the domain upon binding or both. We, therefore, compared the temperature stability of the *apo*- and *holo*-forms using intrinsic tryptophan fluorescence, as well as estimating their hydrodynamic sizes. We found that the T_m of the CW42-H3K4me1 complex was about 6 °C higher than that for uncomplexed CW42 (T_m of 58.0 ± 1.4 °C vs 64.4 ± 1.0 °C, Fig. 2A,B). For size estimations, size-exclusion chromatography with multi-angle light scattering (SEC-MALS) was used as well as diffusion constant measurements using pulsed-field NMR. MALS data showed lower effective hydrodynamic radius, that is the elution time on an SEC column increases, even as the molecular mass of the complex goes up (Fig. 2C). The *holo*-state also shows a shoulder towards the unbound state. Curiously, increasing the ligand concentration beyond further beyond twofold excess did not remove this feature. The NMR diffusion rate measurements collected for the protein and the complex support this observation. The observed diffusion rates correspond to roughly $2.1 \cdot 10^{-10}$ $m^2 \cdot s^{-1}$ and $2.6 \cdot 10^{-10}$ $m^2 \cdot s^{-1}$ for the *apo*- and *holo*-forms (Fig. 2D). Using the Stokes–Einstein relationship [26], these diffusion rates correspond to approximate hydrodynamic diameters of 1.7 and 1.4 nm, respectively. The T_m, MALS and diffusion data support the view that the domain undergoes compaction and stabilization of its structure upon ligand binding. In the following, we elucidated how this was reflected in the structure and dynamics of CW42 at a more detailed level.

Apo- vs *holo*-structural comparison shows C-terminal α1-helix differences and posthelical coil involvement in binding

The most suitable approach for exploring the detailed in-solution molecular changes associated with the binding is comparing the NMR structures of the *apo*- and *holo*-forms of CW. Previously, we published the structure of the free ASHH2 CW domain [8], and now, we present the solution structure of the CW42-H3K4me1 complex. The structure was submitted to the Protein Databank (PDB code: 6QXZ), and a summary of NMR structural statistics and an ensemble representation of the 20 energy-minimized conformers can be viewed in Fig. 3A,B. In Fig. 4A, the *apo*-structure (PDB code: 2L7P, Hoppmann *et al.* [8]) and the *holo*-structure are superimposed. The chemical shifts



D

	K4me1	K4me2	K4me3	K4me0
CWs	**0.21** ± 0.013	**0.91** ± 0.057	**4.3** ± 0.83	n.a.
CW42	**0.22** ± 0.08	**0.72** ± 0.12	**3.7** ± 0.7	n.d.
CW37	**0.65** ± 0.09	**1.22** ± 0.15	**5.74** ± 0.46	n.d.
CW33	**0.85** ± 0.11	**1.58** ± 0.34	**7.30** ± 1.2	n.d.

All K_d values in μM.

Fig. 1. The CW domain of ASHH2 binds H3K4me1 with high affinity. (A) ASHH2 domain organization, with multiple sequence alignment of the evolutionarily conserved CW domains from dicotyledonous flowering plants (*Arabidopsis thaliana,* Q2LAE1), monocotyledonous (maize, Zea mays, A0A1D6HAE7), liverworts (*Marchantia polymorpha,* A0A2R6W143), spikemosses (*Selaginella moellendorffii,* D8SGM1), mosses (*Physcomitrella patens,* A0A2K1L195) and green algae (*Chlamydomonas reinhardtii,* A0A2K3DEA3). The codes in parenthesis identify the UniProt entries used in the alignment, which was generated using ClustalW. Mutations performed in this study (▼), the Hoppmann *et al* study (▼) and the Liu *et al.* study (▼) are indicated on the sequence. The CW core that is conserved across species, as well as secondary structure elements, is shown below the multiple sequence alignment, as is a subselection of 3 of the 20 constructs initially prepared as possible crystallization candidates and how they relate to the main sequence and the CW construct used in the Hoppmann *et al.* paper. (B) Representative intrinsic tryptophan fluorescence spectra (excitation wavelength 290 nm) used for binding assays of CW42 to H3K4me1. The vertical line indicates the wavelength at which emission intensities were used, as determined in the inset. Inset: ΔFluorescence intensity where spectra of CW42 in the absence of any ligand are subtracted from spectra of increasing amounts of H3K4me1. Units are otherwise the same as in the main panel. The wavelength at which the ΔFluorescence intensity was maximal was 322 nm for CW42, and in the range of 319–322 nm for the other constructs. (C) K_d determinations of CWs (—), CW42 (—), CW37 (—) and CW33 (—). Normalized ΔFluorescence intensity values at the wavelength as determined in the inset of Panel B were plotted against ligand concentrations (0.0–7.2 µM H3K4me1). For CWs and CW42, the wavelengths used were 321 and 322 nm, respectively. Protein concentrations were constant throughout anyone titration, but could vary somewhat from construct to construct (always within 2.0–2.4 µM). The data were fitted using nonlinear least-square methods to Eq. 1, yielding three K_d values in each instance. (D) Tabulated affinities (in µM) of CW33, CW37, CW42 and CWs binding to H3K4me0/1/2/3. N.a., not applicable, n.d., not determined. Values are given as means of each individual K_d determined within sets of matching parallels. Error bars are one standard deviation, $n = 3$.

and the position of the side chain of K4me1 are suggestive of cation–π interactions with the indole group of W874 [27] (Fig. 4A). Four residues comprise the top of the hydrophobic pocket hosting the monomethylated lysine: I915, L919, I921 and Q923 (Fig. 4B,D). All of these residues display NOE connectivities with the ligand, indicating stable contacts in solution (see examples in Fig. 3C). A comparison between the backbone of the *apo*- and the *holo*-NMR structures does not indicate a large reorganization of the protein domain upon binding. The most prominent difference is in the η1-loop that changes position to interact with the N-terminal of the ligand (backbone displaced by up to 8 Å). There is also a minor reposition of the C-terminal α1-helix to accommodate the ligand (backbone displaced by ~ 2 Å), as predicted in the Hoppmann *et al.* paper [8].

A crystal structure of the ASHH2 CW domain in complex with H3K4me1 has also recently been published by Liu *et al.* (PDB code: 5YVX) [14]. The η1 region, as well as the I915 and L919, was identified as crucial for ligand binding, and the interactions were discussed in terms of lock–key arguments for the N-terminal part of the ligand. To exploit all existing structural data, we also include this crystal structure in our analysis. Comparing the two known *holo*-structures shows that their backbones match closely except at the α1-helix (Fig. 4C, RMSD between residues S863-Q908 of the *holo*-forms is 0.913 Å), a part of the domain that is crucial for correct binding [8,14]. In the Liu *et al.* structure, this helix is longer than the *holo*-NMR structure, and the structure terminates immediately after the helix. Moreover, the C_αs is displaced by roughly 4 Å towards the C-terminal end of the helix (Fig. 4C) relative to our NMR structure. In order to make the protein domain crystallize, Liu *et al.* introduced an E917A mutation into the α1-helix at a site that is partially conserved (Fig. 1A). While this mutation still allowed the ligand to bind with a somewhat reduced affinity [1.3 ± 0.2 mM (WT) vs 2.79 ± 0.36 mM (E917A)], it may together with the lack of the C-terminal coil have caused the α1-helix to become displaced relative to the NMR structure of the wild-type version of the domain. Liu *et al.* [14] report that N916, located within the α1-helix and positioned next to the E917A mutation, is crucial for both binding of the methylated ligand and the methylation-dependent binding profile of the CW domain. In our structure, we find no evidence for stable contacts between it and the ligand, and at the same time, the ligand is in our case surrounded by the key residues (Fig. 4B,D). We also note that between L919, I921 and Q923, there are two glycines (Fig. 4D). These make no contacts with the ligand, but allow the key residues space and flexibility they need to pack tightly around the ligand. This is markedly different than the configuration found in the crystal structure (Fig. 4E).

The CW42-H3K4me1 complex is stabilized by intermolecular β-augmentation, the α1-helix and the C-terminal coil

Several related structures, including that of MORC and zinc finger CW, report β-augmentation as being part of the binding mechanism, that is that an intermolecular β-sheet is formed upon binding [7,21,28]. To examine whether this is a stable feature of the CW42-H3K4me1 complex, three replicates of 50-ns

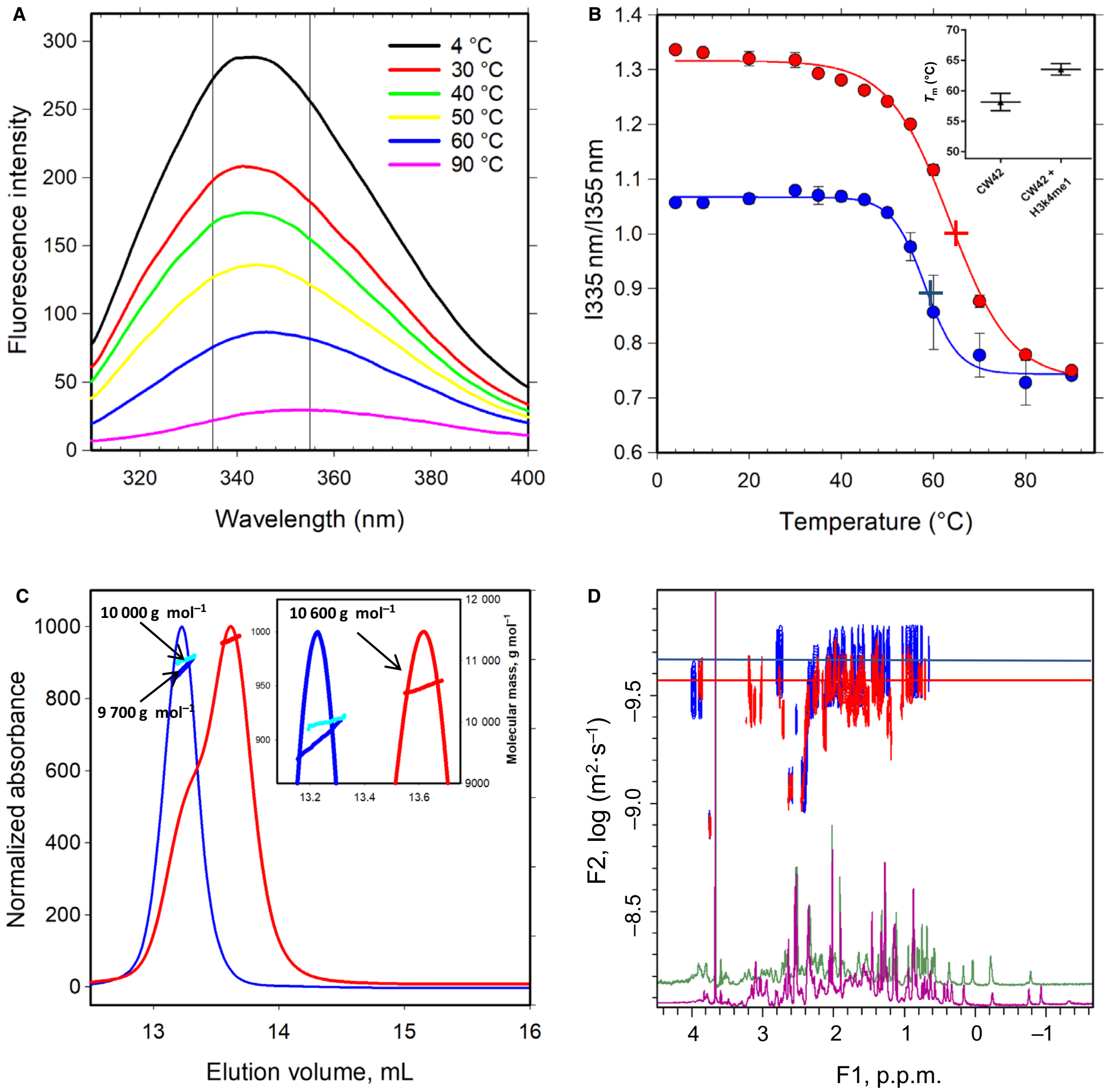


Fig. 2. CW42 becomes more stable upon binding to H3K4me1. (A) Representative intrinsic tryptophan fluorescence spectra of CW42 in the absence of H3K4m1 at 4–90 °C. Tryptophans were excited at 295 nm, and the emission scanned from 310 to 450 nm. Vertical lines at 335 and 355 nm indicate the wavelengths at which intensity values were the intrinsic tryptophan signal dominated by folded and unfolded protein states, respectively. (B) Thermal denaturation profile of bound and unbound CW42. The I335 nm/I355 nm ratios derived from fluorescence data in the presence (●) and absence (●) of H3K4me1 ligand were plotted vs temperature. Each data point represents the mean of three parallels, and error bars are shown as one standard deviation where these exceed the size of the symbols. The data series for the bound (—) and unbound (—) situation were then fitted (nonlinear least squares) to a 4-parameter sigmoidal expression, yielding the midpoint of the denaturation curve, T_m, as an output in the presence (+) and absence (+) of H3K4me1. Inset: summary of T_m for CW42 with and without ligand bound. Error bars show 95% confidence interval of the fits in the main panel. (C) SEC-MALS elution profiles of CW42 in the presence (—) and absence (—) of H3K4me1, where each profile is shown as molecular mass (kDa) vs elution time (min). The molecular mass (g·mol^{-1}) for each elution as determined by static light scattering is shown as red and blue dots for the ligand present and absent situation, respectively. The average molecular masses for each peak are indicated (→). (D) Diffusion measurements of CW42 in the presence (red contours) and absence (blue contours) of H3K4me1. Horizontal axes represent the projection of ^{1}H experiments using bipolar gradient sets separated by diffusion delays and 3-9-19 water suppression. The vertical axis is the logarithm of the diffusion coefficient (D, m^2·s^{-1}). Cross-peaks represent fits of peaks extracted by fitting the 64 ^{1}H experiments to the decay function given by Eq. 3 in the Supplementary Information. Only selected peaks from the nonexchanging, upfield region were used to estimate the mean D, as either buffer components or the ligand do not influence this spectral region. The log D value for CW42 in the presence and absence of ligand is indicated by horizontal lines, (—) and (—), respectively. Representative 1D ^{1}H spectra acquired in the presence (magenta) and absence (green) of the ligand are shown at the bottom of the panel.

A

Restraints used in structure calculation	Number
Total number of NOE distance restraints	1056
Intra-residual NOEs	285
Short-range, $\|i\text{-}j\| = 1$	330
Medium-range, $1 < \|i\text{-}j\| < 5$	204
Long-range, $\|i\text{-}j\| \geq 5$	147
Intermolecular NOEs	74
Number of upper distance limits for Zn^{2+}	8
Number of lower distance limits for Zn^{2+}	8
TALOS N ϕ/ψ dihedral angle restraints	94
Structure statistics, 20 conformers	
CYANA target function value (Å^2)	2.78 ± 0.22
Maximal distance constraint violation (Å^2)	0.36 ± 0.04
Maximal torsion angle constraint violation (Å^2)	0.54 ± 0.49
AMBER energies in implicit solvent (kcal/mol)	–3867.0074
OneDep – Ramachandran statistics	
Residues in favorable regions (%)	89
Residues in allowed regions (%)	9
Residues in outlier regions (%)	2
Root mean square deviation to average coordinates (Å)	
N, C^{α}, C' (860–910)	0.36 ± 0.09
Heavy atoms (860–910)	0.97 ± 0.14

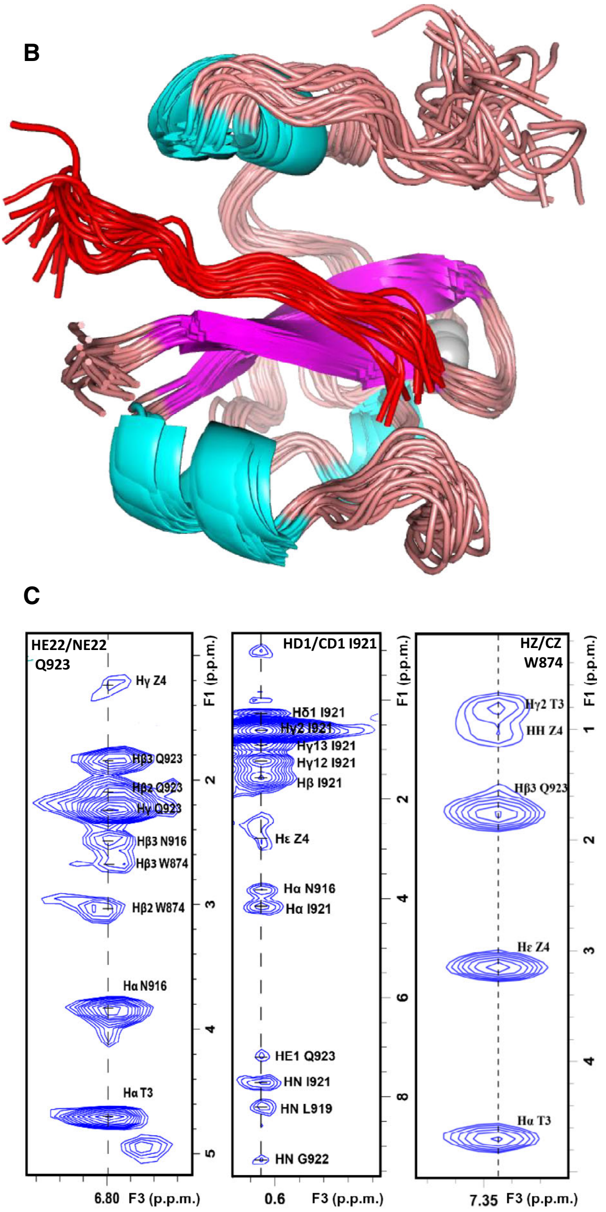


Fig. 3. NMR structure of the CW42-H3K4me1 complex. (A) NMR restraints and structural statistics for CW42-H3K4me1 complex. (B) The structural ensemble of the 20 minimized NMR-derived structures, backbone Cα atoms aligned to the medoid structure, conformer 15. (C) Strip plots for residues Q923 (HE22/NE22), I921 (HD1/CD1) and W874 (HZ/CZ) derived from the filtered-edited 3D NOESY experiments showing intra- and intermolecular NOE connectivities to both the CW domain and the bound ligand. Graphical representations of structures were prepared in PYMOL 1.5 (Schrödinger, New York, NY, USA).

MD simulations were performed using a representative conformation from our NMR structure, and Liu *et al.*'s crystal structure. The replicates were identical except for different initial velocities, and we find that both structures of the complex are stable and able to hold the ligand within its binding pocket as evidenced by their RMSD values throughout the simulation (Fig. 5A, Table S1). The high variations observed from 12 to 20 ns in the RMSD values of the NMR structure simulations are due to a displacement of the C-terminal coil in one replicate (the C-terminal coil is receding from the ligand). The C-terminal coil comes back to interact with the ligand from 20 ns until the end of the simulation. Hydrogen bond analysis of the complex combined with secondary structure analysis of the ligand along the MD simulation trajectory of

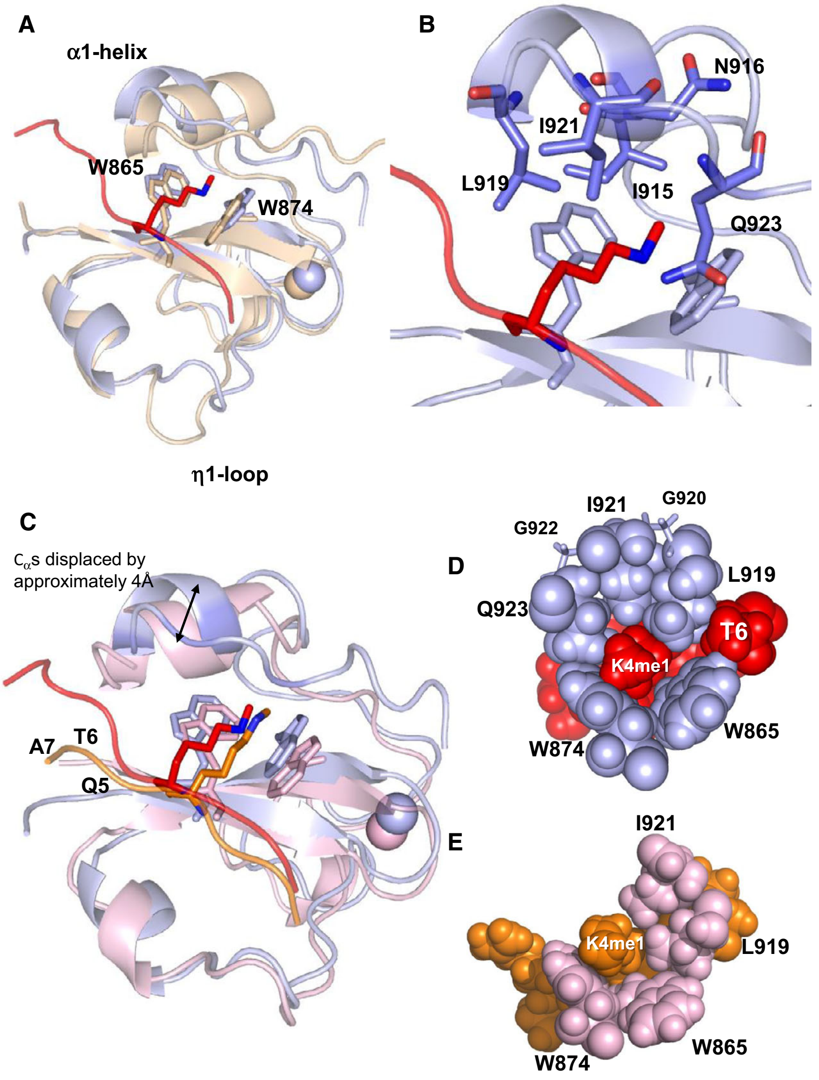
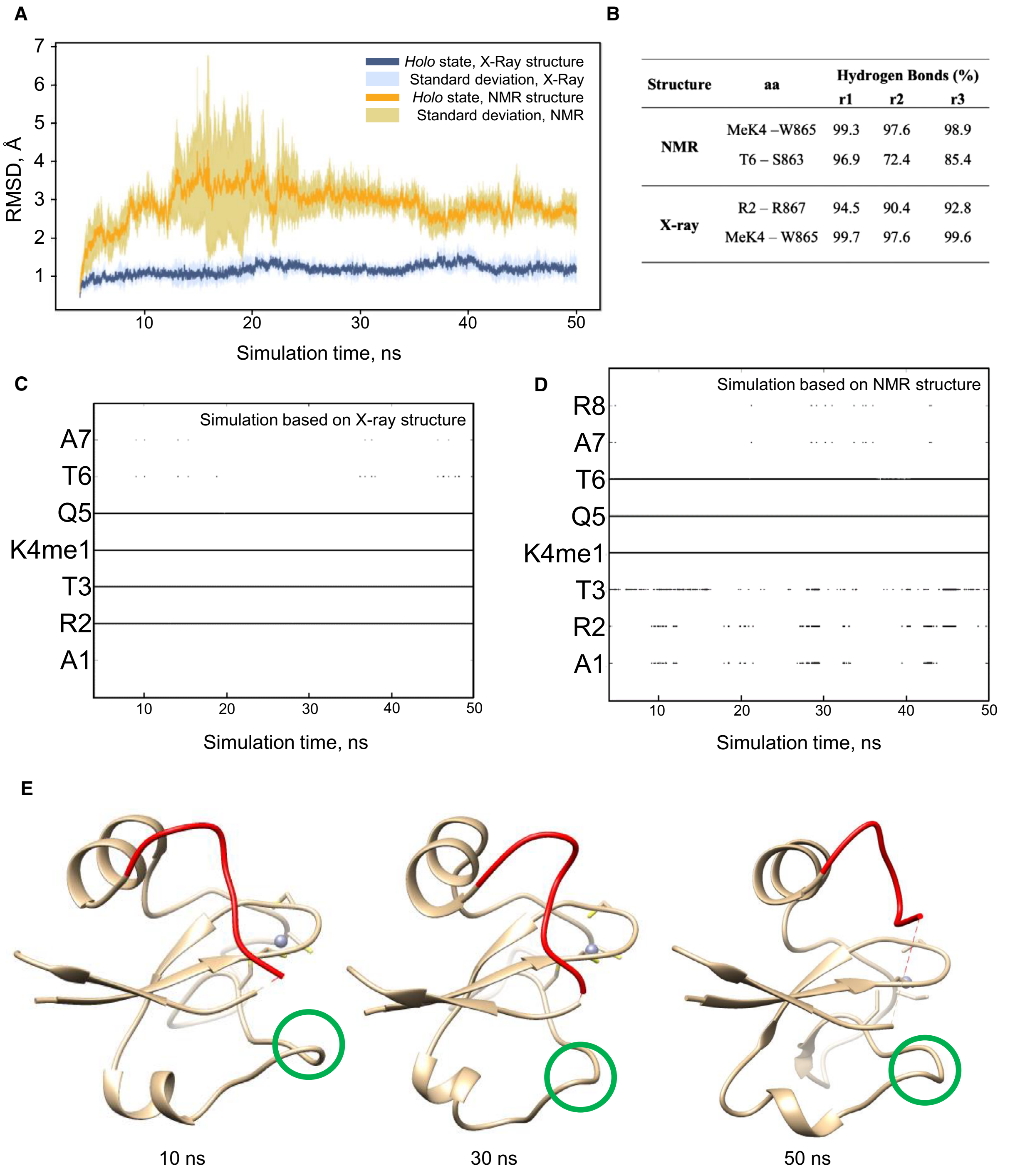


Fig. 4. In-solution structure of CW42-H3K4me1 complex. (A) Structural comparison of the CW42 in the free (PDB code: 2L7P, in beige) and bound state (in light blue, PDB code: 6QXZ). The bound state is represented as the medoid structure, conformer 15. RMSD between the *apo*- and *holo*-structures is 1.6 Å. The ligand is presented in red, and the side chain of the H3K4me1 resides between W865 and W874 of CW domain. Cartoons are rendered with 0.35 transparency setting in PYMOL to increase visibility of key elements. (B) Section of panel A, highlighting the key C-terminal residues forming the hydrophobic pocket – I915, L919, I921 and Q923, clustering around the methylated lysine. N916 is also indicated, but this residue does not contact the ligand. (C) Structural superposition of the CW domain in the bound state – NMR structure (in light blue) and X-ray (in magenta, PDB code: 5YVX). The RMSD value for the core residues 863–908 is 0.913 Å. The H3 residues are positioned differently in the NMR (red) and crystal structure (orange). Ligand residues Q5, T6 and A7 are indicated. The overall backbone of the structures is similar except for the α1-helix, where C_αs is displaced by about 4 Å. (D) Space-fill representation of the ligand-binding site, showing key interactions between the protein in light blue and the ligand in red. (E) As in D, but for the crystal structure, protein in pink and ligand in orange. Graphical representations of structures were prepared in PYMOL 1.5 (Schrödinger, New York, NY, USA).

both structures indicates a stable intermolecular β-sheet (Fig. 5B). The secondary structure analysis of the ligand in both structures shows some difference in the residues involved in β-sheet augmentation through the simulation (residues A2 to Q5 and T3 to A6 of the X-ray and NMR structures, respectively, Fig. 5C,D). The same trend has been observed in the MD simulation replicates. These intermolecular β-sheet interactions are reinforced by hydrogen bonds between the ligand and the CW domain side chains.

In the crystal structure, the ligand is oriented differently for residues Q5-A7, probably due to the misorientation of the α1-helix (Fig. 4C). This is, in turn, a likely consequence of the E917A crystallization mutant



Structure	aa	Hydrogen Bonds (%)		
		r1	r2	r3
NMR	MeK4 –W865	99.3	97.6	98.9
	T6 – S863	96.9	72.4	85.4
X-ray	R2 – R867	94.5	90.4	92.8
	MeK4 – W865	99.7	97.6	99.6

Fig. 5. Induction of β-sheet and rearrangement of C-terminal coil upon ligand binding. (A) Average RMSD evolution (dark colour) and standard deviation (light colour) of the NMR WT CW42-H3K4me1 structure (orange) and crystal structure of the E917A mutant (blue) during the last 46 ns of simulations. (B) Hydrogen bond occupancy between the main chain of amino acids (aa) involved in the intermolecular β-sheet through the last 46 ns of MD simulation performed on the NMR representative structure and the X-ray structure for each replicate (r1, r2 and r3). These hydrogen bonds are present in the initial structures and maintained through MD simulations. (C) and (D) Secondary structure analysis of the ligand along MD simulations performed on the X-ray structure (C) and the NMR structure (D). The results are shown for one replicate of each structure. Points/lines indicate β-strand conformation at a given time during the last 46 ns of the MD simulation. (E) Rearrangement of the C-terminal coil along the simulation. Snapshots from 10, 30 and 50 ns show the C-terminal coil (in red) in interaction with the ligand. The A879-S889 coil that also shifts up to interact with the ligand is encircled in green. Graphical representations of structures were prepared in Chimera.

and the shortened C-terminal part. Although the sequence of the crystal structure ends at I921, just after the α1-helix, both our affinity data for shortened domains (CW33/37) and NMR data suggest that this part of the domain is relevant for binding. Tellingly, there are numerous NOE cross-peaks indicative of stable links from this coil to both the cores of the CW42 domain and the bound ligand (Fig. 3C). For instance, as many as three ligand contacts are mediated by I921, and six are mediated by Q923. In the NMR structure, Q923 resides within a coil, absent in the crystal structure, that appears as an ensemble of fluctuating conformations. Our MD simulations indicate that there is a tendency for this coil to move towards the N-terminal part of the ligand, and together with the η1 region interacts with the ligand but from the opposite side (Fig. 5E). The C-terminal coil's rearrangement with respect to the ligand is observed from around 10 ns and is maintained until the end of the simulation. This observation has been confirmed by the replicates.

Complexation modulates the flexibility of key binding elements

Molecular motions are important for protein function in general and ligand binding in particular [29,30]. We have observed in this study that CW42 responds to binding both at a global level and at a more detailed level. To characterize the motional changes triggered by binding, we compared the local internal motions in the *apo*- and the *holo*-states of the protein using NMR. Steady-state heteronuclear ^{1}H-^{15}N NOE values, and R_1 and R_2 relaxation rates are sensitive to high-frequency motions (10^8–10^{12} s^{-1}) occurring at ps-ns timescale, with R_2 also having contributions from much slower processes occurring at μs-ms timescale [31]. The analysis of these parameters in the free and bound state provides information about the protein local backbone mobility change upon ligand binding (Fig. 6A-C). Overall, all the residues show NOE values near 0.9, indicating backbone motions at the ns scale. Outliers exist in the β2 sheet, and η1 and η3 loop regions. The η1 region and the post-α1-helix flexible loop, including Q923, undergo changes restricting motions upon binding (Figs 4A, 6A and 5E). The R_1 parameter is generally lower for the *holo*-state, indicating an overall stabilization, while the R_2 parameter shows outliers in the D886-R890 interval, as well as M910 and L919.

To further exploit these data, the three relaxation parameters were combined with the structure of the complex using the Lipari–Szabo model-free formalism [31]. Output parameters of this analysis are the order parameter, S^2, reflecting the amplitude of the internal motions on the ns timescale, the effective correlation time for the internal motions, τ_e, and the conformational exchange rate on the μs to ms timescale, R_{ex} (Fig. 6D-F). Overall, the order parameter values, S^2, indicate a quite flexible protein, especially for the *apo*-state. Even its most stable parts have an S^2 value between 0.9 and 0.8, somewhat lower than what is usual for folded proteins and closer to proteins with fluctuating structures [32,33]. Differences between the *apo*- and *holo*-states are found in the loop regions of the protein, post-η1 in particular, but also in the β2-sheet and the α1-helix and its posthelical coil. Overall, the S^2 values suggest a consolidation of the fold upon binding. Residues V882-S889 of the η1 region are restricted upon binding (Fig. 6D). The values of the local correlation time, τ_e, are rather low throughout the protein (within 0.8 ns), indicating overall protein flexibility (Fig. 6E). The R_{ex} parameter, where available, suggests that *apo*-CW undergo conformational exchange on the μs-ms timescale, often associated with conformational shifts related to function [34]. The majority of observed rates are below 2 s^{-1} (Fig. 6F). There are notable outliers, again to be found in the β2-sheet, and near η1 and η3. These residues, with larger values than 2 s^{-1}, are R875, I877, G883, D886, E887, D898, M910, E917, L919 and A926. For these residues and in the α1-helix, we generally observe higher R_{ex} values for the *apo*-form, indicating a slowing down of conformational fluctuations also at the ms-μs timescales. Of these residues, only E887 and L919 make direct contact with the ligand in at least one of the available *holo*-structures, suggesting that lock–key type formalism is not sufficient to understand this binding process.

In contrast, binding through conformational selection may explain why we observe these outliers. Such binding mechanisms postulate that the *apo*-state is flexible and fluctuating and that a small population of the bound conformation exists in equilibrium, also when the ligand is not present [35]. When the ligand is present, binding occurs by stabilizing the pre-organized conformation corresponding to the bound state [36]. Although the preceding Lipari–Szabo model-free analysis implicates dynamic elements in the binding event, the timescales associated with conformational selection are better assessed using relaxation–dispersion NMR experiments. In brief, this approach isolates the contribution of ms-s conformational exchange towards R_2 relaxation [37,38]. We performed these experiments at 600 and 850 MHz, and performed global data fits

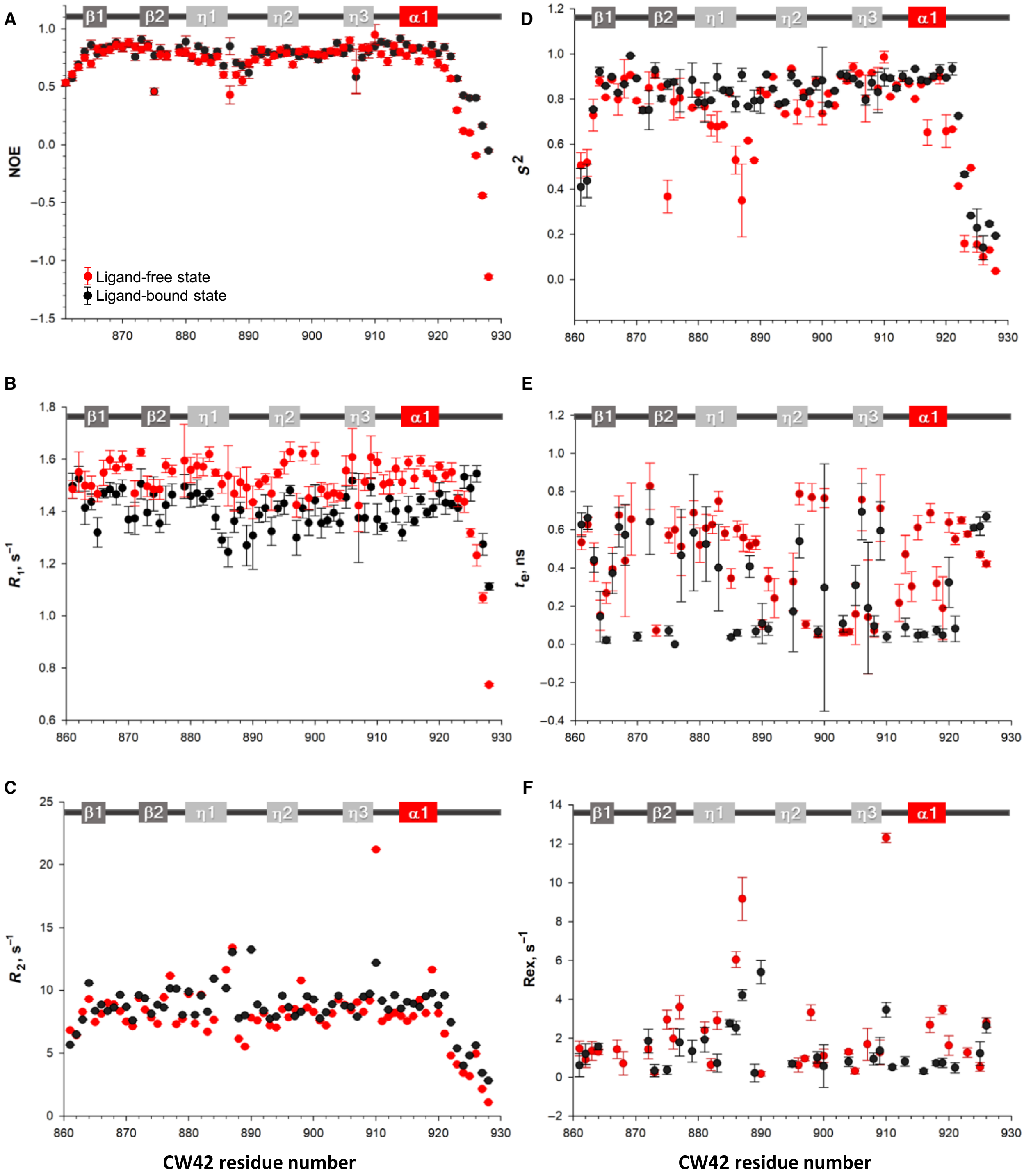
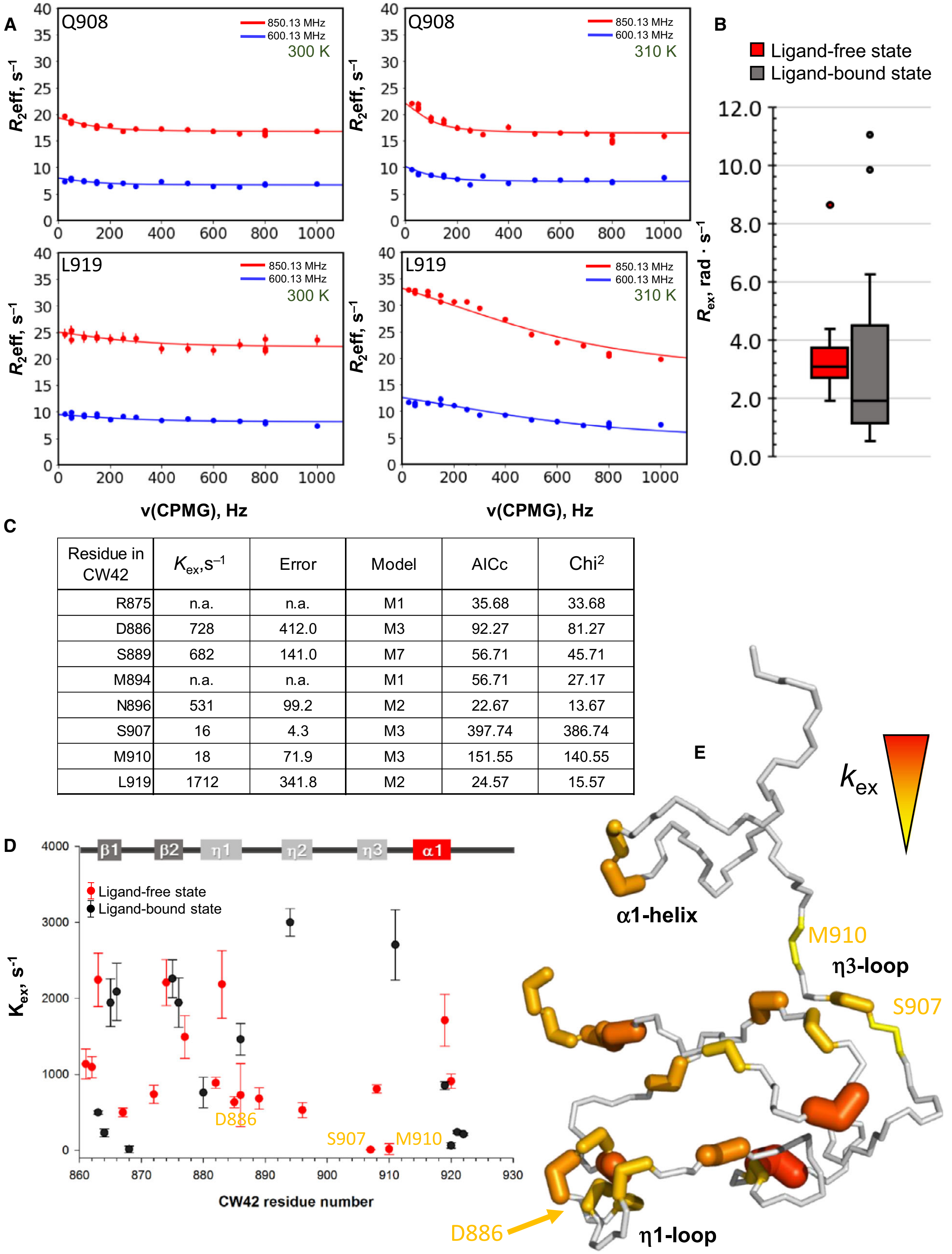


Fig. 6. Relaxation NMR data and Lipari–Szabo model-free dynamic analysis for CW42 in its free (●) and bound to H3K4me1 (•) states. (A) Steady-state ^{1}H-^{15}N NOEs. (B) R_1 relaxation rates. (C) R_2 relaxation rates. Model-free parameters derived for the free and bound states: (D) order parameter S^2, (E) local correlation time τ_e and (F) conformational exchange rate Rex. The CW42 secondary structure is indicated at the top of each panel. Errors are estimated by Monte Carlo simulations (CI, 95%).

using the NESSY software made by Bieri *et al* for this purpose [37]. In NESSY, relaxation dispersion profiles are fitted to models identifying protein motions related to no exchange (i.e. no movement at this timescale), slow-exchange or the fast-exchange limit, for each residue with a backbone amide. The program picks



Residue in CW42	K_{ex},s^{-1}	Error	Model	AICc	Chi^2
R875	n.a.	n.a.	M1	35.68	33.68
D886	728	412.0	M3	92.27	81.27
S889	682	141.0	M7	56.71	45.71
M894	n.a.	n.a.	M1	56.71	27.17
N896	531	99.2	M2	22.67	13.67
S907	16	4.3	M3	397.74	386.74
M910	18	71.9	M3	151.55	140.55
L919	1712	341.8	M2	24.57	15.57

Fig. 7. The ms dynamics of CW. (A) Examples of relaxation dispersion data accumulated at 600 MHz and 850 MHz field strengths (at 300 and 310K), and then globally fitted in NESSY by Monte Carlo simulation. The fitting procedure attempts to fit three models and is selected using χ^2 and AICc tests measuring the goodness of fits. The models M1-M3 allowed are no exchange (M1), two states in the fast-exchange limit exchange (M2) and two states in slow exchange (M3). (B) Box plots of the *apo*- and *holo*-distribution of all R_{ex} output values found by the NESSY fits. Circles designate outliers, whiskers are the highest and lowest nonoutlier values in the data sets, and upper and lower box border is the third and first data quartile, respectively. The data median is indicated by the black bar. Mean values when outliers are removed are 3.1 ± 0.64 rad·s^{-1} and 2.0 ± 1.55 rad·s^{-1} for the *apo*- and *holo*-situation, respectively. The difference between the two data sets is significant ($P < 0.05$, Student's *t*-test, one-tailed, heteroscedastic). (C) Examples of extracted K_{ex} values from fits performed by NESSY, along with the model selected and their χ^2 and AICc scores. Errors are estimated by Monte Carlo simulations (CI, 95%). (D) Plot of determined K_{ex} values for the *apo*- and *holo*-forms of CW42. Values with errors exceeding 500 s^{-1} are not included in this plot. (E) K_{ex} values associated with the *apo*-state plotted onto the unbound NMR structure of CW (2L7P). Residues with quicker motions are drawn using thicker stick representation and more intense red colour. Residues exhibiting slow-exchange behaviour are indicated in yellow and thinner stick representation. For full, tabulated summaries of NESSY output, see Table S4 and S5. The graphical representation of the structure was prepared in PYMOL 1.5 (Schrödinger, New York, NY, USA).

models using an approach avoiding overfitting based on χ^2 and AICc goodness-of-fit scoring functions [39]. Examples of dispersion curves for the *apo*-state at 300K and 310K are shown in Fig. 7A. In the NESSY models, R_{ex} is an output parameter that can be interpreted as the contributions of relatively slow protein motions towards the total R_2 relaxation behaviour. As the ligand binds, there is a significant (*t*-test, one-sided, heteroscedastic, $P < 0.05$) lowering of R_{ex} values (Fig. 7B). We interpret this as an overall quenching of this type of motions upon binding, a behaviour that is expected for binding through conformational selection [40]. For fast-exchange limit (M2) and slow-exchange (M3) residues, tabulated examples of residues displaying relaxation dispersion behaviours consistent with different models of exchange behaviour are presented in Fig. 7C, along with extracted K_{ex} values representing rates of conformational exchange. A plot of K_{ex} values for the *apo*- and *holo*-states, where residues with large errors (more than 500 s^{-1}) removed, is presented in Fig. 7C. All NESSY-selected models, along with their output values and associated χ^2 scores, can be viewed in Table S4 and S5.

In the following analysis, we focus on the residues that exhibit slow exchange, as this type of behaviour is an indication of minor populations that may be relevant for binding [40]. Three residues, D886, S907 and M910, are in slow exchange (M3), according to the NESSY selection. D886 belongs to the η1 loop that shifts towards the ligand upon binding (Fig. 4A), and its actual K_{ex} value is more similar to fast-exchanging residues fitting the M2 model (Fig. 7C). S907 and M910 are, interestingly, located in the η3 loop which is known to affect binding, is fairly conserved, but also shows ASHH2-specific variations (Fig. 1A). Because of this, we still include S907 in our analysis even though its associated AICc and χ^2 values were notably high for S907. We also show, *vide infra*, that S907 has significant ($P < 0.05$) effect on the binding of H3K4me1, and abolishes binding of ligands with K4me2 and K4me3 altogether. The rate of conformational exchange, K_{ex}, is very slow and similar for these two residues, around 16–18 s^{-1} (Fig. 7C). The behaviour and location of the slow-exchanging residues in the η3 loop which leads up to the α1-helix (Fig. 7E) is suggestive of a mechanism where the flexibility of the loop allows the α1-helix to sample the binding conformation, which is then consolidated if the correct ligand is present.

The coils flanking the α1-helix are mediators of binding and flexibility

The findings presented above are consistent with a role for protein conformational sampling in binding. We, therefore, returned to our MD simulations and compared the root mean square fluctuation (RMSF) and the radii of gyration (R_g) of the bound and unbound states. In the simulations, the R_g values of the *apo*-state and *holo*-state overlap at a time, suggesting that the *apo*-state can sample the bound conformation (Fig. 8A). RMSF calculations remove the time dimensions in the simulations and allow this measurement for local flexibility to be plotted onto the domain backbone. Overall, the domain fluctuates from 50% to 30% less in the bound form than in the free form (Figs 8B and S3A). The η1, η2 and η3 regions where molecular rearrangement takes place upon binding appear as outliers with increased and unaffected flexibility. The MD and NMR dynamics data generally match; both approaches indicated a restriction of the *holo*-structure as well as showing outliers in the same regions. Both results corroborate the initial low-resolution characterization of ligand binding (Figs 1 and 2).

Our simulations also indicate a difference between the two *holo*-structures as determined by NMR and

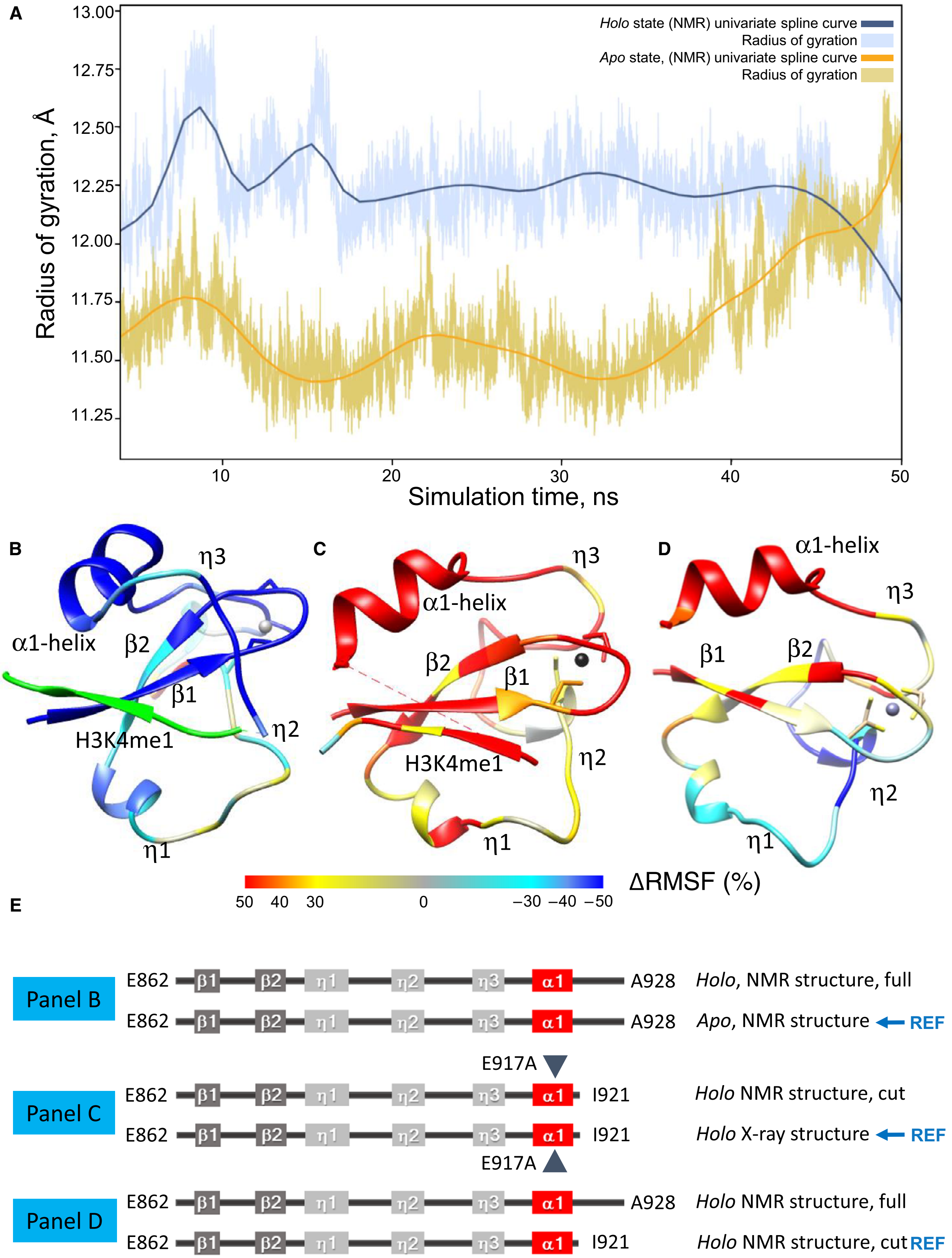

A
Radius of gyration, Å
13.00
12.75
12.50
12.25
12.00
11.75
11.50
11.25
10
20
30
40
50
Simulation time, ns
Holo state (NMR) univariate spline curve
Radius of gyration
Apo state, (NMR) univariate spline curve
Radius of gyration
B
η3
α1-helix
β2
β1
H3K4me1
η2
η1
C
η3
α1-helix
β2
β1
H3K4me1
η2
η1
D
α1-helix
η3
β1
β2
η2
η1
ΔRMSF (%)
50 40 30 0 −30 −40 −50
E
Panel B
E862 β1 β2 η1 η2 η3 α1 A928 Holo, NMR structure, full
E862 β1 β2 η1 η2 η3 α1 A928 Apo, NMR structure REF
E917A
Panel C
E862 β1 β2 η1 η2 η3 α1 I921 Holo NMR structure, cut
E862 β1 β2 η1 η2 η3 α1 I921 Holo X-ray structure REF
E917A
Panel D
E862 β1 β2 η1 η2 η3 α1 A928 Holo NMR structure, full
E862 β1 β2 η1 η2 η3 α1 I921 Holo NMR structure, cut REF

Fig. 8. The flexibility of the domain is influenced by ligand binding and the posthelical coil. (A) The calculated radius of gyration (Å) as a function of simulation time for the *apo* (orange)- and *holo* (blue)-forms of ASHH2 CW for one of the MD simulation replicates (n = 3). The lighter colours show the radii of gyrations, and the darker line (univariate spline curve) highlights their overall trend. (B, C, D) Comparative flexibilities of CW structures. Relative changes in flexibilities are calculated by comparing the RMSF values of a given state to a reference state (per cent change). For all panels, cyan to blue colours represent parts of the complex that are less flexible than the reference state. Yellow to red colours represent the parts of the complex that is more flexible than the reference state. For the extreme red and blue colours, the $|\Delta \text{RMSF}| \geq 50\%$. The results are shown for one replicate of each state. (B) RMSF differences between the *apo*-state as the reference vs the *holo*-state of CW42, using the NMR structures available. Structures used were 2L7P and 6QXZ, modified to match in length, and see panel E. The ligand is represented in light green. (C) RMSF differences between the *holo*-form crystal structure (mutant) as the reference state vs the NMR *holo*-form NMR structure. Structures used were 5YVX and 6QXZ, where the latter NMR structure also has been modified with the E917A mutation and a shortened the C-terminal coil to match of the 5YVX structure, and see panel E. (D) RMSF differences between the full-length NMR *apo*-complex as the reference state vs the same structure without the C-terminal coil. Structures used were 6QXZ, and a version where the latter has the C-terminal coil removed, and see panel E. In this last case, the ligand is not stably locked in the binding site and is therefore not shown. (E) Schematic overview of the structures used in panels B, C and D. The blue triangle indicates the crystallization mutant E917A. The reference model referred above is indicated by REF. Graphical representations of structures were prepared in Chimera.

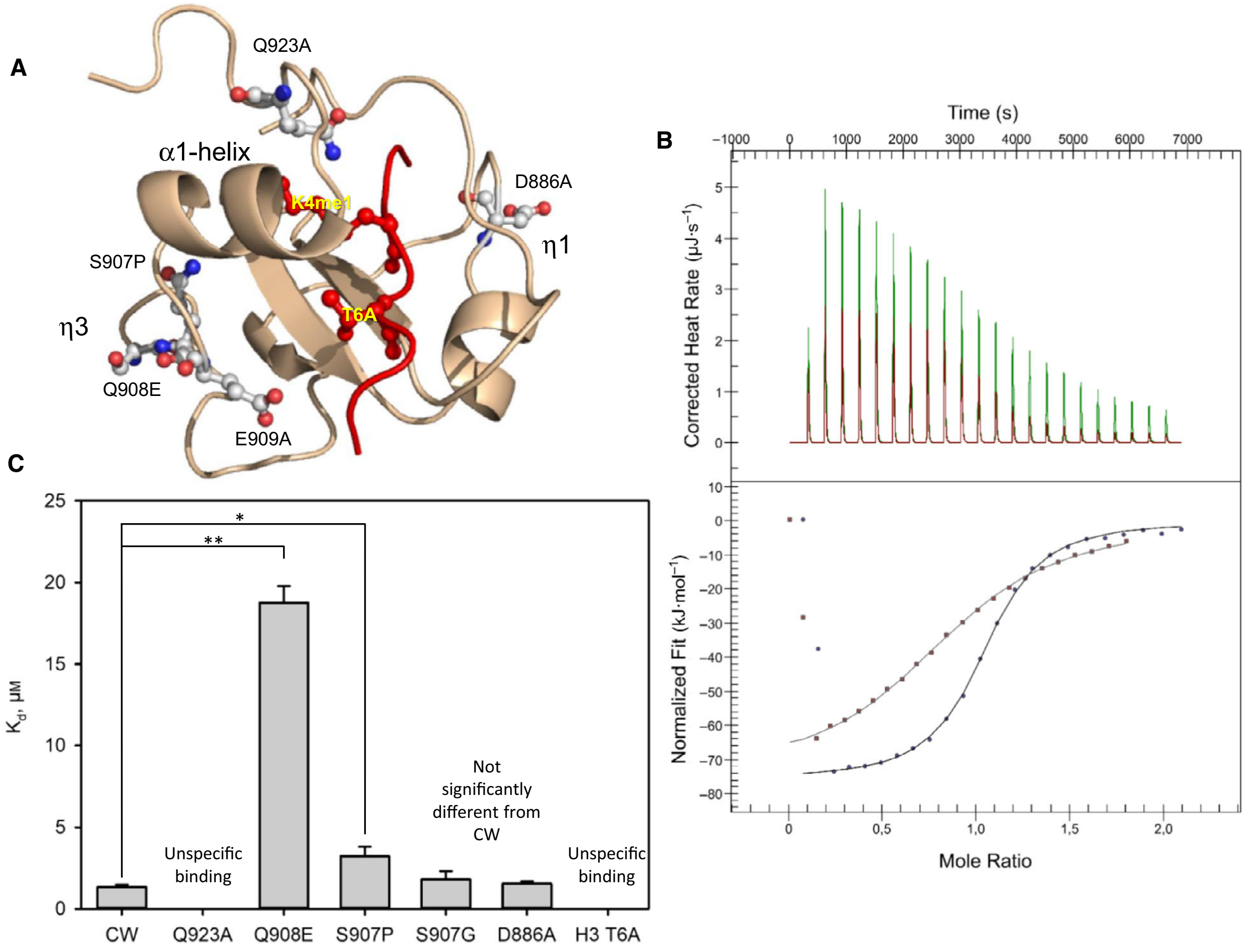
crystallography. However, it was unclear whether this is caused by the inherent differences in the crystal vs the NMR model, or whether the sequence difference is the cause. We, therefore, modified the NMR-structure sequence in such a way that it matched the crystal structure sequence (Fig. 8E), and simulated these two states for 50 ns. The method used to resolve the structure does seem to impact the complex flexibility since the NMR structure fluctuates much more than the crystal structure during MD simulations (30–50% more, Figs 8C and S3B). This suggests that Liu *et al.* structure is restricted to a very limited conformational space.

The truncation of constructs scored for H3Kme1 binding affinity indicated that removing the C-terminal adversely affects binding affinity (Fig. 1, constructs CW33 and CW37). The NMR data and MD simulations also supported a role for the C-terminal post-α1-helix coil in binding (Figs 6A,D and 5E). To evaluate the impact of this part of the sequence on the complex flexibility, we compared the RMSF of the complex with and without the C-terminal coil (for a schematic overview of how structures are compared, see Fig. 8E). The results indicate a high ($\Delta \text{RMSF} \geq 50\%$) and a moderate ($30\% < \Delta \text{RMSF} < 50\%$) increase in flexibility in the NMR structure lacking the C-terminal coil, residing in the α1-helix and β-sheet, respectively (Figs 8D, S3C). In contrast, the η1 and η2 regions experience a stabilization upon removal of the C-terminal coil. Moreover, the ligand is not stable within the binding site of the truncated structure simulations, underlying the importance of the C-terminal coil for the ligand stability. Nevertheless, the ligand is stable in the X-ray structure simulations and in the comparative simulation with its NMR counterpart (see Fig. 8E schematics), suggesting that the E917A mutation plays an important, albeit artificial, role in the complex stability.

Our structural and dynamics results, as well as sequence alignment and earlier work, strongly implicate the η1 and η3 loops. A graphical summary of this is presented in Figs 1A and 9A. η3 notably contains conserved residues with variation relatable to the loop flexibility, such as Pro to Ser variations. The *holo*- or *apo*-structures do not show much difference in these places (Fig. 4A), yet the conservation pattern and the MD and NMR results related to mobility suggest that these residues might be involved in regulating the equilibrium position of the α1-helix in the free and bound situation. An effect on H3K4me1 binding, or ability to differentiate between methylation states, would be particularly interesting since these residues are in a coil without directly contacting the ligand, and the backbone trace of the *apo*- and *holo*-forms is essentially the same (Fig. 4A).

We note that Hoppmann *et al.* mutated two residues in this region, Q908A and E909A, and both mutations effectively abolished binding in pull-down assays ([8] and Fig. 9A). In reference to this work and our current dynamics data (derived from NMR, as well as the RMSF analysis), we further probe the involvement of S907 and Q908 in modulating the binding affinities of H3K4me1-3 using Isothermal calorimetry (ITC). For unmutated CW42 interacting with H3K4me1, the binding constant and stoichiometry of interaction were determined to be $K_\text{d} = 1.09 \pm 0.21$ μM and $n = 0.85 \pm 0.05$, respectively. The reaction is enthalpy-driven (ΔH of -91.63 ± 8.14 kJ·mol^{-1}), while the ΔS term is negative ($\Delta S = -192.98 \pm 25.97$ J·mol^{-1}·K) (Fig. 9B-D), in support of the net ordering of the complex reported above. Since the Q908A polar to aliphatic mutation has already been performed by Hoppmann *et al.*, we did a structurally conservative Q908E mutation that converts this polar residue into a charged one. This results in a 17-fold drop in affinity (Fig. 9B, C,

$P << 0.01$). For S907, we note that Ser is an amino acid associated with a high amount of flexibility, only surpassed in this regard by Gly according to amino acid flexibility rankings [41]. We, therefore, mutated this residue to confer both higher (S907G) and lower (S907P) flexibility to the η3 coil preceding the α1-helix



D

H3K4 me1	CW	D886A	S907P	S907G	Q908E	Q923A	CW to H3K4me1-T6A
K_d, μM	1.09 ± 0.21	1.54 ± 0.15	3.22 ± 0.59	1.80 ± 0.51	18.75 ± 1.02	No reproducible isotherms found (n = 4)	No reproducible isotherms found (n = 12)
n	0.85 ± 0.05	1.09 ± 0.01	1.07 ± 0.07	1.02 ± 0.06	0.89 ± 0.04		
ΔH, kJ/mol	−91.63 ± 8.14	−72.85±1.41	−72.00 ± 4.34	−78.92 ± 5.17	−73.88 ± 1.20		
ΔS, J/mol·K	−192.98 ± 25.97	−133.07±4.64	−135.92 ± 15.26	−154.50 ±19.51	−157.30 ± 3.83		

E

H3K4 me2	CW	D886A	S907P	S907G
K_d, μM	For wild type CW affinities for H3K4m2 and H3K4me3, see Figure 1D.	5.15 ± 0.39	No reproducible isotherms found (n = 3)	No reproducible isotherms found (n = 3)
n		0.96 ± 0.03		
ΔH, kJ/mol		−84.05 ± 0.57		
ΔS, J/mol·K		−180.63±2.57		

F

H3K4 me3	D886, S907P, S907G
K_d, μM	No reproducible isotherms found (n = **3**)
n	
ΔH, kJ·mol^{-1}	
ΔS, J·mol^{-1}·K	

Fig. 9. Mutants in coils and their effects on the binding properties of the CW42-H3K4me1 complex. (A) Ribbon representation of the CW42-H3K4me1 complex. The ligand is shown as sticks in red; mutated residues in this study and Hoppmann *et al.* study are shown in element-specific colouring. There are no direct contacts between the mutated residues and the ligand, except for Q923. The graphical representation of the structure was prepared in PYMOL 1.5 (Schrödinger, New York, NY, USA). (B) Representative ITC data interacting with H3K4me1. Top panel displays representative corrected heat rates plotted against time for CW (—) and the CW Q908E mutant (—) titrated against the ligand. The bottom panel represents the normalized peak areas (CW, •; Q908E, ■) plotted vs the ligand/protein mole ratio. (C) Bar plot representation of mean K_ds where error bars represent one standard deviation ($n = 3$). Statistical significance of pairwise differences as indicated (*t*-test, *, $P < 0.05$, **, $P < 0.01$). (D) Derived thermodynamic parameters for CW42 and mutants binding to H3K4me1. All values are averages based on three determinations, with errors given as standard deviations. N is the binding stoichiometry. (E) Derived thermodynamic parameters for CW42 and mutants binding to H3K4me2. All values are averages based on three determinations, with errors given as standard deviations. N is the binding stoichiometry. (F) Attempts to determine binding and thermodynamic parameters for CW42 mutants binding to H3K4me3 failed after the indicated number of parallels.

and observed that affinity was lowered somewhat. For the restricting S907P mutation, K_d increased almost threefold from 1.09 ± 0.21 μM to 3.22 ± 0.29 μM (*t*-test, $P < 0.05$), while S907G resulted in no significant increase (1.80 ± 0.51 μM). However, when we also investigate the ability of H3K4me2 and H3K4me3 to bind these mutations, we were unable to produce binding isotherms, suggesting that these flexibility-modulating mutations are involved in allowing CW to differentiate between methylation states (Fig. 9E,F). In contrast, the D886A mutation, positioned in the η1 loop that moved towards the ligand upon binding in the structure (Fig. 4A), does not affect H3K4me1 binding much (Fig. 9C) and is still able to bind H3K4me2 at reduced (~ 5-fold) affinity (Fig. 9E).

To examine the effect of the post-C-terminal coil on binding, we also designed a Q923A mutant, which formed NOE contacts with the ligand (Fig. 3C), and performed ITC. These affinity measurements showed that the mutant destroyed specific binding, making it impossible to produce reliable isotherms (Fig. S2), confirming the importance of this residue for binding. We also investigated the effect of altering the H3 peptide. Beyond the central K4me1, the importance of residues 1–3 has been determined by Liu *et al.* [14]. However, the structure presented here suggests that T6 is involved in specific contacts with the ligand site, including L919 and W865 (Fig. 4D). When performing a T6A amino acid substitution in the H3K4me1 peptide, ITC measurements failed to produce reliable binding isotherms (Fig. S2). It may also be relevant that H3 can be phosphorylated at T6 [42]. Such a PTM modification would destroy complementarity (Fig. 4D), and likely abolish binding.

Discussion

Several low-resolution techniques used to assess CW42 to H3K4me1 interaction suggested that a reorganization, compaction and overall slowing down of dynamics takes place upon binding. However, only a limited amount of reorganization was apparent when comparing the *apo*- and *holo*-structures. Investigation of the dynamic behaviour of the bound and unbound states using NMR and MD provided a more comprehensive picture. The *apo*-state is relatively flexible on the ns timescale (Figs 6A,D and 8B), with several hotspots (η1 and η3) also showing tendencies for dynamics on the μs-ms timescale (Fig. 6F). Using relaxation dispersion experiments, we were able to extend our view of the domain's dynamics to the ms-s timescale, where in particular the η3 loop displayed indications of concerted slow exchange at a rate of about 16–18 s^{-1} (Fig. 7C). Significantly, η3 mutations at the S907, Q908 and E909 positions adversely affect binding, especially for H3K4me2 and H3K4me3 (Fig. 9E,F). The MD simulations exhibit a variation in the *apo*-form's RMSD across the simulation that is consistent with a dynamic loosening and compaction of the structure at equilibrium. Although each replicate spans 50 ns only, the repeats all show the same tendencies, and the NMR dynamics spanning ns-ms timescales corroborate this. In sum, this suggests that the equilibrium *apo*-state is less compact than determined structures indicate, and may sample a compact, less flexible *holo*-state.

Compaction behaviour like this has been linked to disorder–order transitions [43]. Our ITC data show that there is a significant entropic cost (−192 $J{\cdot}mol^{-1}{\cdot}K$) associated with binding. The entropic cost must be at least partially related to ordering of flexible elements, as ligand binding should lead to entropically favourable desolvation of the hydrophobic residues of the ligand and the binding pocket. While the *apo*-state of CW is certainly folded, there are enough mobile elements for a disorder-to-order transition to occur. Such binding-induced ordering events are relatable to both induced fit and conformational selection mechanisms of binding, where two fairly flexible entities mutually explore conformational space conductive to binding [35,40]. The availability of these states at ambient conditions is related to both the folding behaviour and flexibility of the protein [44], and both NMR and MD data all point towards CW42 having sufficient flexibility to populate a spectrum of conformations at timescales ranging from ns to ms (Figs 3B and 6-8). For the η3 region, we also find limited evidence that the *apo*-CW exchanging behaviour samples the same state. Throughout these analyses, the η1 and η3 coils consistently display interesting behaviour, and residues therein also show effects on CW binding behaviour when mutated. Taken together, this is consistent with a CW binding behaviour that is suggestive of conformational selection.

There also appears to be an element of induced fit [45], where equilibrium in the bound states of both the ligand and CW42 is shifted towards new structural elements, most notably the β-sheet augmentation observed in the MD simulations (Fig. 5). In addition to the contacts made by the side chain of K4me1, β-sheet augmentation may help dictate the ligand sequence specificity towards the ASHH2 CW domain, as is the case for a number of other instances [46]. The side chains of R2, K4me1 and T6 all orient towards CW42 in the β-sheet and contribute to the final complementarity of the bound state by intercalating as shown in Fig. 4. Conversely, T3, Q5 and A7 are oriented away. This recognition, based on the alternating orientation of side chains, is similar to that reported for the MORC3 CW domain, where β-augmentation is also part of the binding mechanism [21]. Yet, these zipper-like fits are not enough for recognition and binding, as the unmethylated ligand will not bind. Conformational selection may offer an explanation. A lack of or incorrect methylation on K4 will not trigger β-sheet augmentation, and therefore, the zipper-like complementarity will not arise from the ensemble of conformations. The mediator of this could be changes in the α1-helix equilibrium position over the K4me1 side chain, mediated by the η3 loop, followed by rearrangement of coils (the η1 and the C-terminal). In particular, the η1 loop, although fitting models for slow exchange in the relaxation dispersion analysis (Fig. 7B), displays quite fast dynamics, and the D886A mutation is not as crucial for binding of the methylated ligands (Fig. 9). One interpretation of this would be that the first step of binding is mediated by the slowest category of exchange, related to η3 and α1-helix dynamics, and is followed by a more rapid consolidation step mediated by the η1 loop.

Our mutation studies on the coils flanking the α1-helix do show that it is possible to affect binding without directly contacting the ligand or the binding site. Since coils are flexible entities, this suggests that they play the role of tensile regulators of the complementary fit between the ligand, and I915, L919 and Q923. As far as we know, the involvement and functional importance of flexibility in recognizing and binding histones has not been suggested before for CW domains. There are, however, relevant precedents in the literature. Functional flexibility has been reported for acetyltransferases acting on histones [47], as well as bromo-domains binding such acetylation sites [48]. Flexibility and fluctuating conformations may shed light on how ASHH2 CW differentiates between methylation states while at the same time effectively searching the histones. Conformational selection and induced fit mechanisms have been reported to be important for search and dock tasks, such as effectively scanning DNA for sequence-specific DNA methylations [49]. Being able to sample a range of similar PTM states along histones tails before settling down and activating the full enzyme, rather than locking down at the first favourable interaction, would be an attractive property for any protein acting within the complex environment of chromatin.

Although this study concerns the properties of the CW domain, it is relevant to also discuss results in the context of the function of the full-length multidomain ASHH2 protein. Conformational selection is also implicated in regulation mechanisms [35]. Although a speculation, the CW domain could act not only as a passive reader that docks the full enzyme correctly but also play a role in activation and regulation. Extrapolating from the CW structure presented here suggests that the domains of C-terminal from CW are oriented away from H3; however, the movement of the C-terminal of CW also suggests that a rearrangement takes place that could position the SET domain optimally. The β-augmentation could stiffen both CW and the H3 tail as part of this positioning. Our study, although restricted to the ASHH2 CW domain, suggests that protein flexibility, as well as conformational selection, plays an active role in the function of the ASHH2 CW domain. Future work on the structural biology of both nucleosome and chromatin remodellers would benefit from employing a theoretical framework and methodology that allow for the detection and assessment of functional disorder and flexibility.

Material and methods

Materials

The H3 tail mimicking peptides were synthetized by LifeTein (H3K4me1, ARTKme1QTARY). For NMR samples, the peptides were also synthesized with specific stable isotope labelling sites as follows: A(^{15}N,^{13}C)RTKme1QTA(^{15}N,^{13}C)RY; AR(^{15}N,^{13}C)TKme1QTARY; and ARTKme1QTAR(^{15}N,^{13}C)Y. All peptides had 95% purity as assessed by mass spectrometry. D_2O, ^{15}N-enriched (99%) NH_4Cl and ^{13}C-enriched (99%) glucose were purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA), and SVCP-Super-3-103.5 NMR tubes were acquired from Norell Inc. (Morganton, NC, USA). Unless otherwise specified, samples were buffered by the T7 solution (25 mM Tris/HCl pH 7.0, 150 mM NaCl, 1 mM TCEP). Buffer components were acquired from Sigma-Aldrich. CW constructs were prepared using ligation-independent cloning in the KpnI/SacI restriction sites of pET-49b vector (Novagen/Merck, Darmstadt, Germany), and the protein and all mutants were expressed and purified as described [24]. For more details, see Supporting Information: Cloning of CW constructs, site-directed mutagenesis, protein expression and purification. Protein and peptide concentrations for all types of samples were determined using UV-Vis spectroscopy (NanoDrop: absorption at 280 nm, extinction coefficient 19730 $\text{M}^{-1}\cdot\text{cm}^{-1}$ for CW constructs, 1490 $\text{M}^{-1}\cdot\text{cm}^{-1}$ for H3K4meX peptides).

Intrinsic tryptophan fluorescence

Affinity measurements

The approach for determining K_ds was adapted from Ref. [50,51] and is further described in Supporting Information: Intrinsic tryptophan fluorescence affinity measurements. Briefly, for each combination of CW construct and H3K4meX, a titration with constant protein concentration and variable ligand (typically 0.0–8.2 µM, up to 15.9 for low-affinity binders) concentration was performed. Stocks were prepared so that the cuvette concentrations for parallel runs always were within 2.0–2.4 µM. The intrinsic tryptophan fluorescence was monitored at wavelengths where the intensity response was greatest in each case (typically 319–322 nm). Δ-intensity values at the wavelengths where the protein-only contribution is subtracted were used as the observable, F_{PL}, when the dissociation constant K_d was determined using a nonlinear least-squares fit to this equation.

$$F_{OBS} = F_P P_o + (F_{PL} - F_P) \times \left\{ ((K_d + P_o + L_o) - (K_d + P_o + L_o)^2 - 4P_o L_o)^{1/2} \right\} \quad (1)$$

F_P and F_{PL} are fluorescence of the protein and protein–ligand complex, respectively, while P_o and L_o are the total concentrations of the protein and the ligand. The shape of the curves, rather than the absolute measurement values, affects the K_ds, which was determined at least three times for each CW construct–ligand combination. Final affinity values and their errors are means of these determinations, and their standards deviations, respectively.

Thermal denaturation measurements

The ratio of the fluorescence intensities recorded at 335 and 355 nm as a function of temperature was monitored. This ratio is a useful proxy for measuring the unfoldedness of a protein [25] and can be used to fit a 4-parameter sigmoidal curve with T_m as an output parameter. For each sample and temperature point, data were acquired from 310 to 400 nm, and temperature ranged from 4 to 90 °C (5–10 °C stepwise increases). Between each measurement, a 5-min wait was introduced for thermal equilibration. The 100-µL quartz cuvette was equipped with a lid to prevent sample evaporation. For more details, see Supporting Information: Thermal denaturation monitored by intrinsic tryptophan fluorescence.

NMR spectroscopy

Data collection

Data were collected at 25 °C on an 850 MHz Bruker Avance III HD Spectrometer fitted with a ^{1}H/^{13}C/^{15}N TCI CryoProbe and a SampleJet with temperature control for storing samples in between runs (set to 4 °C). Samples were prepared in NMR buffer consisting of 20 mM potassium phosphate, 50 mM NaCl and 1 mM DTT adjusted to pH 6.4, and all NMR experiments related to the backbone and side-chain assignment performed for this study (summarized in Table S2) were collected, processed and analysed as described previously [24]. Protein diffusion measurements were performed on protons at 25 °C using stimulated echo, bipolar gradients and 3-9-19 pulse train for solvent suppression. ^{1}H-^{15}N NOE values, and ^{15}N longitudinal ($R_1 = 1/T_1$) and transverse ($R_2 = 1/T_2$) relaxation rates were acquired using sequences in Table S2. Local backbone dynamics was determined using model-free Lipari–Szabo formalism. CPMG relaxation dispersion experiments were acquired at 600-and 850-MHz fields at 25 and 35 °C using pulse sequences, series of spin-echo pulse elements and relaxation delays described in Table S2. Analysis of the relaxation data was carried out in Bruker Dynamics Center 2.5.3 (Bruker BioSpin, Billerica, MA, USA) and NESSY [37]. For more details, see Supporting Information: Heteronuclear NOE, relaxation measurements and Model free analysis of backbone local dynamics.

Structural calculation, refinement and validation of the CW42-H3K4me1 complex

Assignment of ARTKme1QTARY in complex with CW42 was carried out using 2D-filtered ^{1}H-^{1}H TOCSY and 2D-filtered ^{1}H-^{1}H NOESY spectra. The peptide assignment was used to establish the intermolecular NOE connectivities with CW42 (BMRB ID: 27251, [24]). Intra- and intermolecular NOE cross-peaks were assigned manually using the CARA program v 1.9.1.2 [52]. The structure of the CW42-H3K4me1 complex was then calculated using CYANA v. 3.97 (L.A.Systems, Inc.,Tokyo, Japan) [53] , based on distance constraints converted from the NOESY peak lists and torsion angles obtained from the secondary chemical shifts in TALOS N [54]. Two hundred conformers were calculated using

CYANA with 15 000 simulated annealing steps. To ensure that the Zn^{2+} ion tetrahedral geometry was correctly represented, the position of the Zn^{2+} ion was restricted towards the final stage of calculation by setting the lower and upper distance limits for S^{γ}-Zn^{2+} to 2.2 and 2.4 Å and to 2.9 and 3.4 Å for C^{β}-Zn^{2+} for the residues C868, C871, C893 and C904. The twenty structure conformers of the complex with the lowest target functions were subsequently energy-minimized in implicit solvent (Generalized Born [55]). This was done using the Amber ff14SB force field [56], with modifications for the monomethylated lysine [57] and ZAFF parameters for Zn^{2+} and the Zn-coordinated amino acids [58]. The procedure consisted of 50 000 steps of steepest descent followed by 10 000 steps of conjugate gradient minimization with a 100 nm cut-off for nonbonded interactions. NOE constraints were applied as a square well potential with a force constant of 50 $\mathrm{kcal{\cdot}mol^{-1}{\cdot}Å^{-2}}$. The refined conformations of the CW42-H3K4me1 complex were validated in RCSB validation server OneDep [59] and deposited as PDB entry 6QXZ. For more details, see Table S2.

Molecular dynamics simulations

Molecular dynamics simulations were performed on the following four structures: a representative conformation of the CW42-H3K4me1 complex of the NMR structure reported in this work (PDB ID: 6QXZ, 'NMR full'); the crystal structure complex (PDB ID: 5YVX); the uncomplexed NMR structure (PDB ID: 2L7P); and the CW42-H3K4me1 complex again, but with the E917A mutation specific to the crystal structure and with the C-terminal removed (residues G922–A928, 'NMR Cut'). The latter simulation was performed to evaluate the relevance of these differences and to be able to compare simulations with the same number of atoms. For the NMR structures, residues related to the plasmid or peptide design, and thus not native to the sequence, were removed in the N-terminal prior to simulation. Briefly, the structures were subject to 50-ns MD simulations using NAMD and the CHARMM36 force field [60–62]. In order to obtain a better sampling, a total of three simulations have been run for each system, using different initial velocities. Each system was solvated with TIP3 water molecules and neutralized with chloride and potassium ions using CHARMM, and then energy-minimized prior to three equilibration steps: water simulation with constrained structure (100 ps), simulation with protein backbone and key ligand-binding residues constrained (500 ps), and unconstrained simulation prior to main simulation (500 ps). For details of the equilibration steps, see Supporting Information: Molecular Dynamics Simulations.

Isothermal calorimetry

Isothermal calorimetry was performed for the wild-type (WT) CW42 and the mutants at a stirring rate of 300 r.p.m. at 25 °C. The protein concentrations of CW42 and the mutants were typically in the range of 50–180 μM, and the enthalpy of binding was determined by stepwise titration with 400–1800 μM histone peptide (H3K4meX, ARTKmeXQTARKY, where X denotes the number of methyl groups on the lysine, 1–3). Both the protein and the 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. Corrected heat flow peaks were integrated, plotted and fitted by independent modelling to determine binding parameters using the NanoAnalyze V 2.4.1 software. Experiments were performed in triplicate or more on a Nano ITC from TA Instruments.

Acknowledgements

The authors gratefully acknowledge Research Council of Norway for support through the Norwegian NMR Platform, NNP (infrastructure grant 226244/F50). This work was also supported by Bergen Research Foundation (BFS-NMR-1) and Sparebankstiftinga Sogn og Fjordane (509-42/16). The authors would like to thank Nils Åge Frøystein and Helena Kovacs for useful discussions and expert advice, and the staff at the NNP Bergen node for facilitating the NMR measurements. Furthermore, Marianne Kristin Holmedal is acknowledged for doing early work regarding the role of the α1-helix in binding, Diana Turcu and Carol Issalene for expert technical assistance and Charlotte Bekkhus for sample preparations and performing experiments under guidance. We gratefully acknowledge Steven J. Gamblin for generously supporting the project by providing guidance and laboratory facilities in its initial stages.

Conflict of interest

All authors declare no conflicts of interest.

Author contributions

RA, VDM and EC designed, planned and executed screen for CW constructs. RA, VDM and ØS performed MALS experiments. MB, RA, ØS and ØØF prepared the samples. OD, ØH, JI and JU acquired NMR data. OD performed NMR structural calculations. OD, ØH and JI performed NMR dynamics experiments and analysed the data. MB and ØS performed mutagenesis. MB, ØS, ØØF and SRM designed ITC and other affinity studies. OD, NM and KT refined NMR structure. OD, NM and NR performed molecular dynamic simulations. OD, MB, NM, ØØF, SM, ØS, RA and ØH prepared figures.

OD, MB, NM, ØØF, ØS, RBA, RA and ØH wrote the manuscript. ØH, OD, RA, JU and NR supervised the work. RA and ØH designed the research and wrote grants. All authors interpreted data and read and commented on the manuscript.

References

1 Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* **403**, 41–45.

2 Spotswood HT & Turner BM (2002) An increasingly complex code. *J Clin Invest* **110**, 577–582.

3 Zhang Y & Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* **15**, 2343–2360.

4 Schneider J & Shilatifard A (2006) Histone demethylation by hydroxylation: chemistry in action. *ACS Chem Biol* **1**, 75–81.

5 Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F & Ponting CP (2003) The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28**, 69–74.

6 Sanchez R & Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**, 364–372.

7 He FH, Umehara T, Saito K, Harada T, Watanabe S, Yabuki T, Kigawa T, Takahashi M, Kuwasako K, Tsuda K *et al.* (2010) Structural insight into the zinc finger CW domain as a histone modification reader. *Structure* **18**, 1127–1139.

8 Hoppmann V, Thorstensen T, Kristiansen PE, Veiseth SV, Rahman MA, Finne K, Aalen RB & Aasland R (2011) The CW domain, a new histone recognition module in chromatin proteins. *EMBO J* **30**, 1939–1952.

9 Perry J & Zhao Y (2003) The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. *Trends Biochem Sci* **28**, 576–580.

10 Berg A, Meza TJ, Mahic M, Thorstensen T, Kristiansen K & Aalen RB (2003) Ten members of the Arabidopsis gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. *Nucleic Acids Res* **31**, 5291–5304.

11 Liu Y, Tempel W, Zhang Q, Liang X, Loppnau P, Qin S & Min J (2016) Family-wide characterization of histone binding abilities of human CW domain-containing proteins. *J Biol Chem* **291**, 9000–9013.

12 Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, Husmann D, Zhan Y, Lajoie BR, McCord RP *et al.* (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336**, 1448–1451.

13 Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C *et al.* (2010) Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol Cell* **39**, 222–233.

14 Liu Y & Huang Y (2018) Uncovering the mechanistic basis for specific recognition of monomethylated H3K4 by the CW domain of Arabidopsis histone methyltransferase SDG8. *J Biol Chem* **293**, 6470-6481.

15 Ko JH, Mitina I, Tamada Y, Hyun Y, Choi Y, Amasino RM, Noh B & Noh YS (2010) Growth habit determination by the balance of histone methylation activities in Arabidopsis. *EMBO J* **29**, 3208–3215.

16 Thorstensen T, Grini PE & Aalen RB (2011) SET domain proteins in plant development. *Biochim Biophys Acta* **1809**, 407–420.

17 Zhao Z, Yu Y, Meyer D, Wu C & Shen WH (2005) Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nat Cell Biol* **7**, 1256–1260.

18 Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A & Shen WH (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana. Mol Cell Biol* **28**, 1348–1360.

19 Grini PE, Thorstensen T, Alm V, Vizcay-Barrena G, Windju SS, Jørstad TS, Wilson ZA & Aalen RB (2009) The ASH1 HOMOLOG 2 (ASHH2) histone H3 methyltransferase is required for ovule and anther development in *Arabidopsis*. *PLoS One* **4**, e7817.

20 Li Y, Mukherjee I, Thum KE, Tanurdzic M, Katari MS, Obertello M, Edwards MB, McCombie WR, Martienssen RA & Coruzzi GM (2015) The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants. *Genome Biol* **16**, 79.

21 Andrews FH, Tong Q, Sullivan KD, Cornett EM, Zhang Y, Ali M, Ahn J, Pandey A, Guo AH, Strahl BD *et al.* (2016) Multivalent chromatin engagement and inter-domain crosstalk regulate MORC3 ATPase. *Cell Rep* **16**, 3195–3207.

22 Sandhu KS (2009) Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *J Mol Recognit* **22**, 1–8.

23 Lazar T, Schad E, Szabo B, Horvath T, Meszaros A, Tompa P & Tantos A (2016) Intrinsic protein disorder in histone lysine methylation. *Biol Direct* **11**, 30.

24 Dobrovolska O, Bril'kov M, Odegard-Fougner O, Aasland R & Halskau O (2018) (1)H, (13)C, and (15)N resonance assignments of CW domain of the N-methyltransferase ASHH2 free and bound to the

mono-, di- and tri-methylated histone H3 tail peptides. *Biomol NMR Assign* **12**, 215–220.

25 Martensson LG, Jonasson P, Freskgard PO, Svensson M, Carlsson U & Jonsson BH (1995) Contribution of individual tryptophan residues to the fluorescence spectrum of native and denatured forms of human carbonic anhydrase II. *Biochemistry* **34**, 1011–1021.

26 Macchioni A, Ciancaleoni G, Zuccaccia C & Zuccaccia D (2008) Determining accurate molecular sizes in solution through NMR diffusion spectroscopy. *Chem Soc Rev* **37**, 479–489.

27 Zaric SD (2003) Metal ligand aromatic cation-pi interactions. *Eur J Inorg Chem* **5**, 20–31.

28 Zhang Y, Klein BJ, Cox KL, Bertulat B, Tencer AH, Holden MR, Wright GM, Black J, Cardoso MC, Poirier MG *et al*. (2019) Mechanism for autoinhibition and activation of the MORC3 ATPase. *Proc Natl Acad Sci USA* **116**, 6111–6119.

29 Amaral M, Kokh DB, Bomke J, Wegener A, Buchstaller HP, Eggenweiler HM, Matias P, Sirrenberg C, Wade RC & Frech M (2017) Protein conformational flexibility modulates kinetics and thermodynamics of drug binding, *Nat Commun* **8**, 2276–2289.

30 Yang LQ, Sang P, Tao Y, Fu YX, Zhang KQ, Xie YH & Liu SQ (2014) Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J Biomol Struct Dyn* **32**, 372–393.

31 Kay LE, Torchia DA & Bax A (1989) Backbone dynamics of proteins as studied by N-15 inverse detected heteronuclear nmr-spectroscopy - application to *Staphylococcal Nuclease*. *Biochemistry* **28**, 8972–8979.

32 Alderson TR & Markley JL (2013) Biophysical characterization of alpha-synuclein and its controversial structure. *Intrinsically Disord Proteins* **1**, 18–39.

33 Mazzei L, Dobrovolska O, Musiani F, Zambelli B & Ciurli S (2015) On the interaction of Helicobacter pylori NikR, a Ni(II)-responsive transcription factor, with the urease operator: in solution and *in silico* studies. *J Biol Inorg Chem* **20**, 1021–1037.

34 Boehr DD, McElheny D, Dyson HJ & Wright PE (2010) Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. *Proc Natl Acad Sci USA* **107**, 1373–1378.

35 Csermely P, Palotai R & Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **35**, 539–546.

36 Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM & Kay LE (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590.

37 Bieri M & Gooley PR (2011) Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* **12**, 421.

38 Farber PJ & Mittermaier A (2015) Relaxation dispersion NMR spectroscopy for the study of protein allostery. *Biophys Rev* **7**, 191–200.

39 d'Auvergne EJ & Gooley PR (2003) The use of model selection in the model-free analysis of protein dynamics. *J Biomol NMR* **25**, 25–39.

40 Chakrabarti KS, Agafonov RV, Pontiggia F, Otten R, Higgins MK, Schertler GFX, Oprian DD & Kern D (2016) Conformational selection in a protein-protein interaction revealed by dynamic pathway analysis. *Cell Rep* **14**, 32–42.

41 Huang F & Nau WM (2003) A conformational flexibility scale for amino acids in peptides. *Angew Chem Int Ed Engl* **42**, 2269–2272.

42 Karimi-Ashtiyani R & Houben A (2013) *In vitro* phosphorylation of histone H3 at threonine 3 by Arabidopsis Haspin is strongly influenced by posttranslational modifications of adjacent amino acids. *Mol Plant* **6**, 574–576.

43 Devarakonda S, Gupta K, Chalmers MJ, Hunt JF, Griffin PR, Van Duyne GD & Spiegelman BM (2011) Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1alpha/ERRgamma complex. *Proc Natl Acad Sci USA* **108**, 18678–18683.

44 Halskau O Jr, Perez-Jimenez R, Ibarra-Molero B, Underhaug J, Munoz V, Martinez A & Sanchez-Ruiz JM (2008) Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc Natl Acad Sci USA* **105**, 8625–8630.

45 Okazaki KI & Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci USA* **105**, 11182–11187.

46 Remaut H & Waksman G (2006) Protein-protein interaction through beta-strand addition. *Trends Biochem Sci* **31**, 436–444.

47 Langini C, Caflisch A & Vitalis A (2017) The ATAD2 bromodomain binds different acetylation marks on the histone H4 in similar fuzzy complexes. *J Biol Chem* **292**, 19121.

48 Setiaputra D, Ross JD, Lu S, Cheng DT, Dong MQ & Yip CK (2015) Conformational flexibility and subunit arrangement of the modular yeast Spt-Ada-Gcn5 acetyltransferase complex. *J Biol Chem* **290**, 10057–10070.

49 Estabrook RA & Reich N (2006) Observing an induced-fit mechanism during sequence-specific DNA methylation. *J Biol Chem* **281**, 37205–37214.

50 Martin SR & Schilstra MJ (2008) Circular dichroism and its application to the study of biomolecules. *Method Cell Biol* **84**, 263–293.

51 Martin SR, Schilstra MJ & Siligardi G (2011) Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling. Royal Society of Chemistry, Cambridge.

52 Keller RLJ (2005) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment.

53 Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353–378.

54 Shen Y & Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**, 227–241.

55 Hawkins GD, Cramer CJ & Truhlar DG (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* **100**, 19824–19839.

56 Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE & Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713.

57 Papamokos GV, Tziatzos G, Papageorgiou DG, Georgatos SD, Politou AS & Kaxiras E (2012) Structural role of RKS motifs in chromatin interactions: a molecular dynamics study of HP1 bound to a variably modified histone tail. *Biophys J* **102**, 1926–1933.

58 Peters MB, Yang Y, Wang B, Fusti-Molnar L, Weaver MN & Merz KM (2010) Structural survey of zinc-containing proteins and development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput* **6**, 2935–2947.

59 Young JY, Westbrook JD, Feng Z, Sala R, Peisach E, Oldfield TJ, Sen S, Gutmanas A, Armstrong DR, Berrisford JM *et al.* (2017) OneDep: unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure* **25**, 536–545.

60 Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M & MacKerell AD (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J Chem Theory Comput* **8**, 3257–3273.

61 Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K & Schulten K (1999) NAMD2: greater scalability for parallel molecular dynamics. *J Comput Phys* **151**, 283–312.

62 Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S *et al.* (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545–1614.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Selected distances between the centres of mass of amino acids (aa) of the binding site and K4me1.

Table S2. List of NMR experiments used in this study.

Table S3. Primers used for Ligation Independent Cloning and site-directed mutagenesis of CW42.

Table S4. Model choice and output summary from NESSY for the unbound state of CW42.

Table S5. Model choice and output summary from NESSY for the bound state of CW42.

Fig. S1. K_d determinations of CWs, CW42, CW37 and CW33 for H3K4me2 and H3K4me3.

Fig. S2. Loss of specific binding caused by mutations in the H3-mimicking peptide and the postα1 loop.

Fig. S3. Average RMSF and standard deviation of the CW constructs assessed by MD.