Decoding the centromeric nucleosome through CENP-N

Satyakrishna Pentakota (1,*), Keda Zhou (2,*), Charlotte Smith (1), Stefano Maffini (1), Arsen Petrovic (1), Garry P. Morgan (3), John R. Weir (1,&), Ingrid R. Vetter (1), Andrea Musacchio (1,4,#) & Karolin Luger (2,5,#)

(1) Department of Mechanistic Cell Biology, Max-Planck Institute of Molecular Physiology, Otto-Hahn-Straße 11, 44227 Dortmund, Germany

(2) Department of Chemistry and Biochemistry, University of Colorado at Boulder; 3415 Colorado Avenue, Boulder, CO – 80303

(3) Department of MCDB, University of Colorado at Boulder; 3415 Colorado Avenue, Boulder, CO – 80303

(4) Centre for Medical Biotechnology, Faculty of Biology, University Duisburg-Essen, Universitätsstrasse, 45141 Essen, Germany

(5) Howard Hughes Medical Institute

*Equal contribution

Correspondence: karolin.luger@colorado.edu, andrea.musacchio@mpi-dortmund.mpg.de

& Present address: Friedrich Miescher Laboratory, Spemannstrasse 39, 72076 Tuebingen, Germany

Keywords: PYRIN, CENP-A, CENP-C, CENP-L, CENP-N, centromere, kinetochore, nucleosome, chromosome segregation, epigenetics, electron microscopy, X-ray crystallography, biochemical reconstitution, cell division, cell cycle

Short title: Structure of the CENP-N:CENP-A nucleosome complex

Word count: Main text 2268, Summary 279

Centromere protein (CENP) A, a histone H3 variant, is a key epigenetic determinant of chromosome domains known as centromeres. Centromeres nucleate kinetochores, multi-subunit complexes that capture spindle microtubules to promote chromosome segregation during mitosis. Two kinetochore proteins, CENP-C and CENP-N, recognize CENP-A in the context of a rare CENP-A nucleosome. Here, we reveal the structural basis for the exquisite selectivity of CENP-N for centromeres. CENP-N uses charge and space complementarity to decode the L1 loop that is unique to CENP-A. It also engages in extensive interactions with a 15-base pair segment of the distorted nucleosomal DNA double helix, in a position predicted to exclude chromatin remodelling enzymes. Besides CENP-A, stable centromere recruitment of CENP-N requires a coincident interaction with a newly identified binding motif on nucleosome-bound CENP-C. Collectively, our studies clarify how CENP-N and CENP-C decode and stabilize the non-canonical CENP-A nucleosome to enforce epigenetic centromere specification and kinetochore assembly.

Introduction

Accurate segregation of chromosomes from a mother cell to its two daughters during cell division is a prerequisite for healthy cell physiology and for the transmission of the genetic information across generations (Santaguida and Amon, 2015). Specialized, conserved molecular machinery dedicated to this crucial function has been identified in the majority of eukaryotic organisms studied to date (Drinnenberg et al., 2016; van Hooff et al., 2017). The purpose of this machinery is to generate stable linkages between chromosomes, the carriers of genetic information, and the mitotic spindle, the microtubule-based structure devoted to the segregation of chromosomes into the daughter cells.

In the last two decades, substantial progress in our understanding of the molecular features of the chromosome segregation apparatus has been made. A crucial role in this process is played by centromeres, specialized chromatin domains whose defining mark in almost all known eukaryotes is the enrichment of centromeric protein A (CENP-A, also known as CenH3), which replaces histone H3 in nucleosomes (Fukagawa and Earnshaw, 2014; Musacchio and Desai, 2017). The primary function of centromeres is to provide a

platform for the assembly of macromolecular complexes known as kinetochores, whose task in turn is the physical capture of microtubules of the mitotic spindle. Kinetochores contain approximately 30 core subunits, normally subdivided in centromere-proximal and microtubule-proximal groups. The microtubule-proximal subunits (outer kinetochore), which are directly implicated in microtubule binding, are usually denoted as the KMN assembly, from the name of three sub-complexes, the Knl1, Mis12, and Ndc80 complexes (Musacchio and Desai, 2017). The centromere-proximal subunits (inner kinetochore), which are also organized in sub-complexes, are collectively identified as the constitutive centromere associated network (CCAN) because they appear to reside at centromeres for the entire cell cycle (Cheeseman and Desai, 2008; Foltz et al., 2006; Izuta et al., 2006; Obuse et al., 2004; Okada et al., 2006) (Figure 1A).

The ability of CENP-A to nucleate kinetochores depends on its incorporation into nucleosomes (CENP-A nucleosomes) with histones H2A, H2B, and H4. *In vitro*, these interact specifically and selectively with two CCAN components, CENP-C and CENP-N (Carroll et al., 2010; Carroll et al., 2009; Guo et al., 2017; Guse et al., 2011; Hoffmann et al., 2016; Klare et al., 2015; Nagpal et al., 2015; Samejima et al., 2015; Weir et al., 2016). Binding of these proteins to CENP-A nucleosomes has been shown to require two regions where the CENP-A sequence diverges significantly from that of histone H3, the L1 loop and the C-terminal tail (Carroll et al., 2009; Fachinetti et al., 2013; Guo et al., 2017; Kato et al., 2013; Logsdon et al., 2015). An evolutionary conserved motif of CENP-C, present in one or two copies in different organisms, is sufficient for recognition of CENP-A *in vitro*. This motif interacts primarily with a solvent-exposed acidic patch on the H2A and H2B subunits of the CENP-A nucleosome and also decodes the divergent C-terminal tail of CENP-A (Guo et al., 2017; Kato et al., 2013). The two copies of this motif in human CENP-C are referred to as the central motif (or domain) and the CENP-C motif (Figure 1A). While at least the central motif has been shown to be required for efficient centromere retention of newly incorporated CENP-A (Guo et al., 2017), neither motif appears to be strictly necessary for centromere localization of CENP-C in human cells (Guo et al., 2017), likely because CENP-C contains binding sites for additional CCAN subunits that can stabilize its centromere localization even in the absence of a direct interaction with CENP-A (Guo et al., 2017; Hinshaw and Harrison, 2013; Klare et al., 2015; McKinley et al., 2015; Nagpal et al., 2015; Weir et al., 2016). The specific succession of binding sites within CENP-C, a protein that secondary structure prediction algorithms identify as being largely

intrinsically disordered, has led to suggest that it acts as a blueprint in the establishment of the inner to outer kinetochore axis, with an N-terminal motif involved in stabilizing the outer kinetochore, a middle region involved in stabilizing the inner kinetochore CCAN complex, and a C-terminal region involved in interactions with the centromeric chromatin (Figure 1A) (Gascoigne et al., 2011; Kato et al., 2013; Klare et al., 2015; McKinley et al., 2015; Przewloka et al., 2011; Screpanti et al., 2011).

CENP-N forms a constitutive complex with CENP-L (designated CENP-LN complex), which in turn interacts with the CENP-HIKM complex and with CENP-C (Guo et al., 2017; Hinshaw and Harrison, 2013; Klare et al., 2015; McKinley et al., 2015; Weir et al., 2016). Binding of CENP-N requires the exposed L1 loop of CENP-A and may also reach into the neighbouring DNA (Carroll et al., 2010; Carroll et al., 2009; Fang et al., 2015; Guo et al., 2017). The structural basis of the interaction of CENP-N with the CENP-A nucleosome, however, has remained elusive. Furthermore, it is unclear whether this interaction is sufficient for the recruitment of CENP-N to the kinetochore, or whether additional interactions with CCAN subunits are also required. Here, we addressed both issues. First, we combined X-ray crystallography and cryo electron microscopy (EM) to gain a high-resolution view of the CENP-N:CENP-A nucleosome complex, and identified and validated the main determinants of this interaction. Second, we defined the determinants of a physical interaction of CENP-LN with CENP-C and demonstrated that kinetochore recruitment of CENP-N requires the coincident presence of CENP-A and CENP-C at kinetochores. Our studies have important implications for kinetochore assembly and epigenetic specification of centromeres.

Results

Crystal structure of CENP-N$^{1-235}$

Human CENP-N, a 339-residue protein (Figure 1B), interacts directly with CENP-L (Hinshaw and Harrison, 2013; Weir et al., 2016). When immobilized on solid phase and challenged with CENP-A or H3 nucleosome core particles (NCPs), CENP-LN interacted specifically with CENP-A^{NCP} (Figure 1C). As shown previously (Carroll et al., 2009), the CENP-A binding region of the CENP-LN complex lies within the N-terminal region of CENP-N, because a stable fragment encompassing residues 1-212 of human

CENP-N (CENP-N$^{1-212}$) also bound selectively to CENP-A^{NCPs} but not H3^{NCPs} (Figure 1D).

To address the structural features of CENP-N and the basis of its interaction with the CENP-A^{NCP}, we therefore focused our structural analysis on N-terminal constructs of CENP-N (Figure 1B). We obtained well diffracting crystals of the CENP-N$^{1-235}$ construct and determined its crystal structure at 2.8 Å resolution (Table 1). CENP-N$^{1-235}$ consists of two closely juxtaposed domains that interact through an extended interface to form a single structural unit (Figure 2A-B). The first domain (residues 1-77) consists of a 5-helix bundle, whereas the second domain (residues 78-212, cyan in Figure 1A) consists of a 6-stranded anti-parallel β-sheet sandwiched between α-helices (Figure 2C-D). There is no clear density beyond residue ~210, indicating that the structure is disordered after this point. Fold-recognition by DALI (Holm and Rosenstrom, 2010) identified similarity of the first domain to PYRIN domains (PYDs; a superposition is shown in Figure 2 Supplement 1A-B). PYDs are 'death fold' family domains implicated in protein-protein interactions relevant to inflammation and apoptosis (Ratsimandresey et al., 2013). They have not been previously implicated in interactions with DNA or chromatin.

Iml3 and Chl4 are fungal orthologs of CENP-L and CENP-N, respectively. We referred to a previously reported crystal structure of the full-length Iml3 protein bound to the C-terminal region of Chl4 (Iml3:Chl4^{C}, PDB ID 4JE3) (Hinshaw and Harrison, 2013) to deduce the structural organization of the human CENP-LN complex. Iml3 consists of an N-terminal domain (shown in green in Figure 2 Supplement 2A) and a C-terminal domain (the 'insert' domain shown in yellow; the topology of Iml3 is shown in Figure 2 Supplement 2B). Iml3 hetero-dimerizes with Chl4 through a subdomain within the insert domain (Figure 2 Supplement 2B) (Hinshaw and Harrison, 2013). Due to strong sequence similarity of Iml3 and CENP-L throughout their length (not shown), the structure of Iml3 provides an excellent model for the structure of CENP-L. Importantly, although our crystal structure does not encompass the C-terminal region of CENP-N, the sequence of the latter is strongly related to that of Chl4^{C} (Figure 2 Supplement 3A), which was captured in complex with Iml3 in the Iml3:Chl4^{C} structure, indicating that they are also structurally related (Figure 2C). Indeed, as already observed (Guo et al., 2017; Hinshaw and Harrison, 2013), the C-terminal region of CENP-N (CENP-N^{230-C}) was sufficient to interact with CENP-L (Figure 1 Supplement 1). Thus, the structure of CENP-N$^{1-235}$ reported here and that of the Iml3:Chl4^{C} complex are complementary, and

together provide an almost comprehensive view of the CENP-L^{Iml3}:CENP-N^{Chl4} complex (Figure 2 Supplement 3A and Figure 2 Supplement 3B).

Besides identifying the N-terminal domain of CENP-N$^{1-235}$ as a PYRIN domain, DALI also identified an unanticipated structural homology of the second domain of CENP-N$^{1-235}$ with the N-terminal domain of Iml3^{CENP-L} (Figure 2 Supplement 3C). We therefore refer to these domains of CENP-N and CENP-L as CLN-HD (for CENP-L and CENP-N homology domain). Structural similarities of the CLN-HD suggest that CENP-N and CENP-L are evolutionary related. However, sequence identity of the two domains, even after structural superposition, is minimal, likely explaining why structural similarity had not been predicted (Figure 2 Supplement 3D). CENP-L, or its complex with CENP-N^{230-C}, did not interact with CENP-A^{NCPs} or H3^{NCPs} (Figure 1E and Figure 1 Supplement 1). Thus, CENP-L and CENP-N, even if partly structurally related, have clearly distinct functions. In conclusion, the structure of CENP-N contains an N-terminal Pyrin domain, a central CLN-HD, and a C-terminal CENP-L dimerization domain, while CENP-L contains an N-terminal CLN-HD, interrupted immediately before the C-terminal helix by an insertion that contains a region required for CENP-N dimerization.

Cryo-EM analysis of the CENP-N:CENP-A nucleosome complex

Using cryo electron microscopy (cryo-EM), we obtained a three-dimensional reconstruction of CENP-N$^{1-289}$ bound to CENP-A^{NCPs} at ~4.0 Å (Figure 3A-B, Figure 3 Supplement 1, and Table 2). We built an atomic model of the CENP-N:CENP-A^{NCP} complex by fitting into the EM density high-resolution models of the CENP-A histone core (PDB ID 3AN2) (Tachiwana et al., 2011), combined with DNA derived from a nucleosome reconstituted with the 145-bp 601 DNA sequence (PDB ID 3LZ0; Vasudevan et al., 2010), and the newly determined crystal structure of CENP-N$^{1-235}$. Both manual and automatic fitting strategies produced unequivocal fits, allowing the first visualization of the interaction of CENP-N with the CENP-A nucleosome (Figure 3 Supplement 2).

The CENP-A nucleosome appears to be stabilized by its interaction with CENP-N (Guo et al., 2017). There is clear density for 139 of the 145 bp of DNA and for the N-terminal helix of CENP-A (Figure 3C and Figure 3 Supplement 2A), two features reported to be largely disordered and thus invisible in the crystal structure of the CENP-A^{NCP} (PDB ID

3AN2) (Tachiwana et al., 2011). CENP-N, whose structure changes very little upon binding to the CENP-A nucleosome, is positioned on top of the L1 loop of CENP-A (also called RG loop for the presence of a conserved arginine-glycine motif at the loop's apex) and contacts approximately 15-bps of the adjacent DNA gyre (Figure 3A). There is clear density only until CENP-N^{1-289} residue ~210, indicating that the following approximately 80 C-terminal residues (at the opposite end of the nucleosome interaction interface) may be flexible. Of the ~2400 $Å^2$ of CENP-A^{NCP} and CENP-N surface area that become buried in the complex, ~1400 $Å^2$ are at the CENP-N:DNA interface, where both CENP-N domains form extensive interactions with DNA from bp -21 to -35 relative to the 2-fold axis, or superhelical location [SHL] -2 to -3. There is a marked accumulation of positively charged residues on this DNA binding interface (Figure 3D). Four loops in the CLN-HD straddle the DNA double helix over ~8 bp, and the consecutive 7 bp are bound by the PYRIN domain, which is positioned to insert an arginine (R44) into the compressed minor groove in an arrangement that is reminiscent of the minor groove arginines inserted by the histones (Figure 3E). The highly conserved P17 in the PYRIN domain positions the main chain of CENP-N to latch on to the phosphate backbone of the DNA, with interactions made through the side chains of K15, R42, K45, K81, and R194. There are also likely insertions of CENP-N side chains into two minor grooves (besides R44, also K148, M167, R170) and the intervening major groove at SHL -3 [R196, see also (Carroll et al., 2009)] (Figure 3E). In agreement with the presence of a large interaction interface with nucleosomal DNA, CENP-N bound more tightly to CENP-A^{NCPs} but retained substantial binding affinity for $H3^{NCPs}$ in electrophoretic mobility shift assays (EMSAs) (Figure 3 Supplement 3A-B). Likely, this residual binding to H3NCPs in the EMSAs, which emerged less clearly in solid phase binding assays (Figure 1C), reflects emphasis on electrostatic interactions under the low salt conditions of the EMSA assays (150 mM NaCl for complex formation, followed by further dilution upon loading onto the gel), compared to the GST binding assays (300 mM NaCl), as also discussed in the context of Figure 4 Supplement 2.

The CENP-N:CENP-A interface

The substantial interface with DNA explains why CENP-N does not bind CENP-A:H4 tetramers lacking DNA (Carroll et al., 2009). However, while DNA binding clearly contributes to the binding affinity of this interaction, it is unlikely to contribute to the

discrimination of CENP-A^{NCPs} from H3^{NCPs}, because CENP-N bound selectively to CENP-A^{NCPs} even when the CENP-A^{NCPs} and H3^{NCPs} contained the same DNA sequence (Figure 1B-C). Conversely, the structure clearly suggests why recognition of the L1 loop is crucial for discrimination (Black et al., 2004; Carroll et al., 2010; Carroll et al., 2009). CENP-N binds the L1 loop through a continuous interface comprising the α1 helix in CENP-N^{PD} and the β3-β4 loop in CENP-N^{CLN-HD}. Several of the infrequent conserved solvent-exposed residues of CENP-N (identified by a green dot in Figure 2D), including E3, E7, R11, K143, P145, N146, and K148 reside in this interface. Y147, which is less conserved, contributes to the stabilization of the relative arrangements of the CENP-N^{PD} and CENP-N^{CLN-HD}, which is largely unchanged in the nucleosome-bound and free structures of CENP-N. Insertion of the side chain of M1 into the hydrophobic core contributes to the stabilization of the α1 helix. The interaction with the CENP-A L1 loop engages a triad of residues, E3^{CENP-N}, E7^{CENP-N}, and R11^{CENP-N}, whose side chains emerge from the same face of the α1 helix looking towards the L1 loop (Figure 3F).

The CENP-A residues R80^{CENP-A} and G81^{CENP-A} form a 2-residue insertion that is the most conspicuous difference between the L1 loops in CENP-A and H3 (Figure 3F, Figure 3 Supplement 4A-B). The insertion is crucial, because it allows R80^{CENP-A} to form hydrogen bonds with both E3^{CENP-N} and E7^{CENP-N}, while absence of a side chain at G81^{CENP-A} allows the CENP-A loop to insert deeply into a cleft formed between the two CENP-N domains, where the side chain of Y147^{CENP-N} packs tightly against V82^{CENP-A}. In EMSAs, mutation of R80 and G81 to alanine partly ablated the preference of CENP-N for CENP-A^{NCPs} (Figure 3 Supplement 3A-B). The side chain of R11^{CENP-N}, a residue previously shown to be important for the CENP-N:CENP-A^{NCP} interaction (Carroll et al., 2009), on the other hand, is squeezed between the loop 1 region of CENP-A and the loop 2 region of H4, where it may be involved in a double salt bridge with E74^{H4} and E7^{CENP-N} (Figure 3F and Figure 3 Supplement 5).

Mutational validation of the CENP-N:CENP-A^{NCP} structure

We generated a collection of single and double alanine point mutants to probe the role of individual CENP-N residues in the interaction with the CENP-A^{NCP}. In pull-down assays *in vitro*, we found essentially complete loss of binding with an alanine (A) mutant of R11 (Figure 4A), and substantial reductions of binding with alanine mutants of E7 or Y147, at

the CENP-A L1 interface, or of K15 or K45, at the interface with DNA (Figure 4 Supplement 1A). Combining mutations of Y147 with either K15 or K45 almost completely disrupted CENP-A^{NCP} binding (Figure 4A), in line with the idea that recognition of the L1 loop and of the DNA jointly contribute to the binding affinity of CENP-N for the CENP-A nucleosome. CENP-N targeting to centromeres in U2OS cells reflected the observations made *in vitro*, with R11A single mutant and the K15A-Y147A and K45-Y147A double mutants appearing severely impaired in the ability to target centromeres (Figure 4B-C), and other single mutants suffering intermediate effects on binding to centromeres (Figure 4 Supplement 1B).

In competition gel shift assays, CENP-N mutants at the CENP-A L1 loop interface (L1 mutants), including R11A and two double mutants (E3A-E7A and K143A-Y147A), lost the ability of $CENP\text{-}N^{wt}$ to discriminate between CENP-A and H3-nucleosomes (Figure 4 Supplement 2A-B). We have shown in Figure 4A that $CENP\text{-}N^{R11A}$ does not bind $CENP\text{-}A^{NCPs}$ in solid phase binding assays. The residual interaction of this mutant with $CENP\text{-}A^{NCPs}$ or $H3^{NCPs}$ in EMSAs likely reflects the extensive binding interface of CENP-N for nucleosomal DNA (whose effects are emphasized under low salt conditions, as already indicated for $CENP\text{-}N^{wt}$ in the context of Figure 3 Supplement 3). In line with this interpretation, we find that in the EMSAs the L1 mutants of CENP-N bind $H3^{NCPs}$ indistinguishably from $CENP\text{-}N^{wt}$, whereas the same mutants bind to $CENP\text{-}A^{NCPs}$ considerably worse than $CENP\text{-}N^{wt}$ (Figure 4 Supplement 2). Collectively, these results further emphasize the importance of the L1 loop of CENP-A in selective recognition by CENP-N.

Identification of a CENP-C region involved in CENP-LN binding

As discussed in the Introduction, CENP-C, an intrinsically disordered protein, provides a defined succession of binding sites for different kinetochore components (Klare et al., 2015) (Figure 1A). CENP-C, CENP-LN, and CENP-HIKM, another kinetochore sub-complex located in the vicinity of CENP-A, form a 7-subunit complex (designated CHIKMLN) that binds $CENP\text{-}A^{NCPs}$ cooperatively, i.e. with increased binding affinity in comparison to any of the individual subunits or sub-complexes (McKinley et al., 2015; Weir et al., 2016). Within this assembly, $CENP\text{-}C^{2\text{-}545}$ binds the CENP-LN complex *in*

vitro [Figure 5 Supplement 1A and (Hinshaw and Harrison, 2013; McKinley et al., 2015; Nagpal et al., 2015; Weir et al., 2016)].

We set out to exploit biochemical reconstitution and our improved structural understanding of the CENP-LN complex to query the importance of this interaction for kinetochore assembly in humans. Trimming of CENP-C^{2-545} identified CENP-$C^{225-364}$ as a minimal CENP-LN interaction domain (Figure 5A), in line with a recent study (Guo et al., 2017). Neither CENP-L nor CENP-N^{1-235} bound to CENP-C^{2-545} (Figure 5 Supplement 1B-C). However, the CENP-LN^{230-C} dimer bound CENP-C in the absence of nucleosomes (Figure 5B). Thus, CENP-$C^{225-364}$ binds at or near the CENP-LN dimer interface, possibly also exploiting structural ordering of these regions upon dimerization. CENP-$C^{225-364}$ contains a handful of conserved residues, some of which were previously shown to mediate an interaction with the CENP-HIKM complex (Klare et al., 2015) (Figure 5C). We probed an additional conserved linear motif in CENP-$C^{225-364}$ (residues 302-306) for its potential role in CENP-N binding. A 5-alanine mutant of residues 302-306 (identified as CENP-C^{5A}) failed to interact with CENP-NL, identifying this region of CENP-C as the CENP-LN binding motif (Figure 5D and Figure 5 Supplement 2). Importantly, CENP-$C^{2-545-5A}$ did not interact with CENP-LN, but retained binding to CENP-A^{NCPs} and CENP-HIKM (Figure 5E-F). In isothermal titration calorimetry (ITC) experiments, CENP-LN^{230-C} bound CENP-$C^{225-364}$ with a dissociation constant (K_D) of 1 μM, but it showed no binding to CENP-$C^{225-364-5A}$ (Figure 5G-H).

CENP-LN binding motif of CENP-C is required for kinetochore recruitment of CENP-N

The availability of a CENP-LN binding mutant of CENP-C gave us an opportunity to ask if the interaction of CENP-LN with CENP-A, besides being necessary, is also sufficient for kinetochore recruitment of CENP-N. For this, we depleted CENP-C by RNAi and replaced it with exogenous wild type (wt) or mutant (5A) copies. Depletion of CENP-C prevented kinetochore localization of CENP-N, showing that nucleosome binding is not sufficient for CENP-N to reach kinetochores at the low cellular concentration of these proteins. Exogenously expressed wild type CENP-C promoted CENP-N recruitment, while CENP-C^{5A} failed to promote it (Figure 6A-B). Thus, the CENP-LN binding site of CENP-C, while not crucial for CENP-C recruitment to

kinetochores, is instead crucial for CENP-N recruitment. Overall, these observations indicate that the CENP-LN complex reads the presence of two features of kinetochores, the presence of CENP-A and the presence of CENP-C, both of which are necessary for its efficient recruitment.

Discussion

The histone H3 variant CENP-A is an essential feature of centromeres and has two main functions. First, it is required for kinetochore assembly through its direct interactions with inner kinetochore subunits that can then seed the assembly of this large macromolecular assembly. Second, it is a landmark that determines the stability of centromere chromatin identity through cell division. Interactions of CENP-C and CENP-N with CENP-A^{NCPs} are the only known direct and specific points of contact of the kinetochore with the centromere and are therefore the crucial effectors through which CENP-A implements its role (Carroll et al., 2010; Carroll et al., 2009; Kato et al., 2013).

While the structural basis of CENP-C binding to CENP-A had been described (Kato et al., 2013), how CENP-N binds CENP-A had remained elusive. Here, we have filled this important gap, as shown schematically in Figure 6D-E. CENP-N nucleosome binding differs from that observed with RCC1 and other nucleosome binders that engage primarily an exposed acidic patch on histones H2A and H2B (Makde et al., 2010). However, in its outline it resembles the interaction of the ATPase domain of SWI2/SNF2 chromatin remodeler with H3^{NCPs} (Farnung et al., 2017; Liu et al., 2017; Narlikar et al., 2013), with the important difference that SWI2/SNF2 closely approaches H3 without making significant direct contacts with it, whereas CENP-N interacts directly with CENP-A (Figure 3 Supplement 6A-D). There are also similarities with the nucleosome binding mechanism of the bromo-adjacent homology (BAH) domain of Sir3 (PDB ID 3TU4) (Armache et al., 2011), but the latter interacts predominantly with the H4 N-terminal tail, through recognition of K16^{H4}, and with the acidic patch on H2A-H2B, and much less extensively with DNA (Figure 3 Supplement 6E-F). In the CENP-N:CENP-A^{NCP} complex, the normally disordered N-terminal tail of H4 is ordered until R23 and interacts weakly with the CENP-N loop connecting β3 with β4 (Figure 3 Supplement 2D). The reported mono-methylation of K20 of H4 in the CENP-A

nucleosome (Hori et al., 2014) may further modulate this interaction. In summary, the SWI2/SNF2 and BAH modes of nucleosome binding are predominantly based on interactions with DNA or with the histones, respectively, while CENP-N shows a balance of both. The considerable interaction of CENP-N with DNA is a remarkable and unexpected feature of the complex structure.

CENP-C and CENP-N can interact concomitantly with the same CENP-A nucleosome (Carroll et al., 2010), as also confirmed in recent studies (Guo et al., 2017; Weir et al., 2016). The central motif and the CENP-C motif of CENP-C, which confer CENP-A recognition ability *in vitro*, interact through a 'arginine anchor' with the acidic patch of H2A and H2B, and also decode the divergent C-terminal tail of CENP-A (Kato et al., 2013). These determinants of CENP-C binding on CENP-A are located adjacent to, but not overlapping with, the CENP-N binding footprint. Indeed, when modelled on the CENP-N:CENP-A^{NCP} structure according to the position it adopts in its structure with the nucleosome (PDB ID 4X23) (Kato et al., 2013), the CENP-C motif can be accommodated without steric clashes (Figure 6C,E). Thus, CENP-C and CENP-N interact with CENP-A through complementary interfaces. In the context of a larger CCAN complex, these CENP-A binding motifs cooperate to increase the overall binding affinity for CENP-A (Guo et al., 2017; Weir et al., 2016).

It has been proposed that CENP-N is significantly stabilized upon binding to the CENP-A nucleosome (Guo et al., 2017). Our study did not identify a clear structural basis for this phenomenon, as we failed to identify significant conformational changes in CENP-N in isolation (crystal structure) or in its complex with CENP-A^{NCP}. It has also been proposed that CENP-C reshapes and rigidifies the CENP-A nucleosome and that it modulates the DNA termini to make them match the loose wrap observed at centromeres (Falk et al., 2015). Importantly, these effects of CENP-C binding on the CENP-A nucleosome do not appear to be required for the selective (over H3) interaction of CENP-N, because selectivity for CENP-A was retained in the absence of CENP-C [this study and (Carroll et al., 2009; Weir et al., 2016)]. We also note that the DNA termini appear to be well defined in our structure of the CENP-N:CENP-A^{NCP} complex, contrarily to what was observed in the structure of the isolated CENP-A^{NCP} (PDB ID 3AN2). At present, we cannot definitively conclude whether the stabilization of the termini is due to CENP-N binding to the CENP-A nucleosome, as we have not yet been able to obtain a high-resolution EM structure of the CENP-A nucleosome in

385 isolation for comparison. It is possible that the cryogenic conditions used for our
386 structural work stabilize a specific conformation of the complex.

387 In most organisms, centromere identity is not specified by the centromere's DNA
388 sequence, but rather by the enrichment of CENP-A at a defined chromatin domain. *De*
389 *novo* formation of stably inherited centromeres at previously non-centromeric sites (neo-
390 centromeres) provides clear evidence in favour of this idea. Thus, rather than being
391 genetically (i.e. DNA-sequence) specified, centromeres are epigenetically specified, with
392 the pre-existing enrichment of CENP-A being a necessary condition for continued
393 deposition of new CENP-A at the same site through the generations. There is therefore
394 considerable interest in the molecular mechanisms that promote new CENP-A
395 deposition at centromeres during the cell cycle, and in the mechanisms that promote the
396 stabilization and persistence of CENP-A after its incorporation at centromeres.

397 Conserved machinery for new CENP-A deposition, including the specialized CENP-A
398 chaperone HJURP (Scm3 in *S. cerevisiae*) and an adaptor complex consisting of the Mis18
399 and M18BP1 subunits, has been described in recent years (Dunleavy et al., 2009; Foltz et
400 al., 2009; Fujita et al., 2007; Hayashi et al., 2004; Pidoux et al., 2009; Sanchez-Pulido et
401 al., 2009; Williams et al., 2009). Additional machinery, in particular chromatin
402 remodelling enzymes harnessing a source of energy to evict H3, is likely involved in the
403 reaction but has not been univocally identified. This machinery is recruited to
404 centromeres early during the cell-cycle and is believed to promote the replacement of
405 histone H3 with new CENP-A (Dunleavy et al., 2011; Jansen et al., 2007; Schuh et al.,
406 2007). Likely, the existing CENP-A nucleosome acts as a template in this reaction, as the
407 abundance of CENP-A nucleosomes at a given centromere is, at least in first
408 approximation, constant through subsequent cell divisions (French et al., 2017; Hori et
409 al., 2017; Jansen et al., 2007). This implies that the same number of new CENP-A
410 nucleosomes is incorporated after each cell division as that of originally present CENP-A
411 nucleosomes, suggesting that the deposition machinery targets for H3 eviction and
412 replacement with CENP-A an H3 nucleosome that is likely in close proximity of the
413 CENP-A nucleosome (Musacchio and Desai, 2017). While the mechanistic details of
414 CENP-A deposition remain partly unclear, there is now substantial evidence that
415 recruitment of the CENP-A deposition machinery requires CENP-C and possibly other
416 CCAN factors (Dambacher et al., 2012; Moree et al., 2011; Shono et al., 2015). While
417 CENP-N has not been directly implicated in CENP-A deposition, our observation that

CENP-N occupies a region of the nucleosome required for binding by chromatin remodelling enzymes of the SWI2/SNF2 family suggests that CENP-N may protect centromeric nucleosomes from remodelling and eviction, thereby contributing to its stability. Indeed, both CENP-C and CENP-N contribute to the stabilization of newly incorporated CENP-A in centromeric chromatin (Guo et al., 2017). New CENP-N deposition at centromeres occurs in late S phase (Fang et al., 2015; Hellwig et al., 2011), and may trigger a stabilization of centromere organization required for successful kinetochore assembly.

In summary, our analysis of the mechanisms of the interactions of the CENP-NL complex with CENP-A and CENP-C represents a step forward in the molecular dissection of the almost universally conserved functions of CENP-A in eukaryotes, which are required for accurate chromosome segregation and, more generally, for the success of cell division and the propagation of life.

Methods

Production of recombinant proteins in insect cells GST-CENP-L and its complexes with CENP-N fragments were produced in insect cells. Specifically, the coding sequence expressing 3C cleavable GST-tagged CENP-L was sub-cloned into MCS2 of the MultiBac vector pFL (Bieniossek et al., 2012), where the coding sequences of CENP-N or CENP-$N^{230\text{-}C}$ were sub-cloned into MCS1 of pFL for co-expression. All constructs were further transformed into EMBacY cells and subsequently transfected into Sf9 cells in order to produce baculovirus. Baculovirus was amplified in Sf9 cells and used to infect Tnao38 cells. Tnao38 cells expressing GST-CENP-L:CENP-N, GST-CENP-L:CENP-$N^{230\text{-}C}$, or GST-CENP-L virus were cultured for 72 h and isolated (Weir et al., 2016). Briefly, cells were resuspended in lysis buffer containing 50 mM HEPES (pH 7.5), 300 mM NaCl, 10% glycerol, 4 mM 2-mercaptoethanol, 1 mM $MgCl_2$ pH 7.5 in presence of Benzonase. Resuspended cells were lysed by sonication and centrifuged at 100,000g at 4 °C for 1 h. Cleared lysates were incubated with pre-equilibrated GSH-Sepharose beads (Amintra) at 4°C for 2h. After extensive washing with lysis buffer, the GST fusion proteins were eluted in lysis buffer supplemented with 20 mM reduced glutathione. The eluted proteins were concentrated in a 10 K Da Amicon-Ultra-15 Centrifugal filter either alone or in the presence of GST-tagged 3C protease. Concentrated proteins were subsequently loaded onto a Superdex 200 16/600 column equilibrated in 20 mM HEPES

pH 7.5, 2.5% glycerol, 300 mM NaCl and 1 mM TCEP. The corresponding peak fractions were collected and concentrated in a 10 KDa MWCO concentrator and then flash frozen in liquid N2, and stored at -80 °C until further use.

Full length CENP-N was cloned into a pACEbac1 vector. CENP-N^{NT} constructs (containing the N-terminal constructs 1-289 and 1-238) were generated by introducing a sequence encoding 6His followed by two stop codons at the designated positions. CENP-N^{NT} constructs were expressed in SF21 insect cells (Invitrogen). Pellets from 1L of cell culture (~2 million cells/ml) were lysed in 100 ml lysis buffer (50 mM sodium phosphate pH 7.5, 300 mM NaCl, 20 mM imidazole and 10% glycerol). Nickel NTA beads (1.5 ml) were incubated with the lysed cells overnight at 4ºC. Beads were washed with wash buffer (20 mM Tris.Cl pH 7.5, 300 mM NaCl, 30 mM imidazole and 10% glycerol), and eluted with elution buffer (20 mM Tris.Cl pH 7.5, 300 mM NaCl, 150 mM imidazole and 10% glycerol). The concentrated eluate was loaded onto an S200 gel filtration column in buffer containing 20 mM Tris.Cl pH 7.5, 300 mM NaCl, 10% glycerol and 1 mM DTT for the final purification.

Production of recombinant proteins in bacteria Fragments encoding different constructs of CENP-N were sub-cloned from a cDNA into pST50Tr-DHFRHIS for expression of recombinant C-terminally polyhistidine-tagged products (Tan et al., 2005). To produce GST tagged CENP-N, a GST encoding sequence was sub-cloned in pST50Tr-DHFRHIS in frame with the sequence coding for CENP-N^{1-212}. Mutant CENP-N^{1-212} constructs were created by site-directed mutagenesis using the QuikChange kit (Agilent Technologies). Escherichia coli (DE3) cells harbouring vectors expressing CENP-N^{1-212}-His or CENP-N^{1-235}-His were grown in TB media supplemented with 100 μg ampicillin at 37 °C at an OD_{600} of 0.8-1.0. Then the temperature was reduced to 20 °C and protein expression was induced with 0.3 mM IPTG for 16 h. Cells were harvested at 4600 g for 15 minutes. Bacterial pellets were resuspended in lysis buffer containing 50 mM HEPES (pH 7.5), 300 mM NaCl, 10% glycerol, 2 mM 2-mercaptoethanol, 10 mM $MgCl_2$ and 5 mM Imidazole. Resuspended cells were lysed by sonication and cleared by centrifugation at 100,000 g at 4ºC for 30 minutes. The cleared lysate was incubated with cOmplete™ his tag beads (Roche) and incubated at 4°C for 2 hr. After extensive washing, CENP-N^{1-212}-His or CENP-N^{1-235}-His were eluted in lysis buffer supplemented with 300 mM Imidazole. A 6 ml ResourceS cation exchange column was pre-equilibrated in 15% buffer B (20 mM HEPES, 1 M NaCl, 10% glycerol, 10 mM $MgCl_2$ pH 7.5, 1 mM

TCEP) and 85% buffer A (20 mM HEPES, 10% glycerol, 10 mM $MgCl_2$ pH 7.5, 1 mM TCEP). CENP-N^{1-212}-His or CENP-N^{1-235}-His were diluted with buffer A to reach a final concentration of 150 mM NaCl, loaded onto the ResourceS column and eluted with a linear gradient of buffer B from 150 to 1000 mM NaCl. Fractions containing CENP-N^{1-212} His or CENP-N^{1-235} His were concentrated in a 10 kDa MWCO and loaded onto a Superdex200 16/600 column equilibrated in 20 mM HEPES pH 7.5, 2.5% glycerol, 300 mM NaCl and 1 mM TCEP. The corresponding peak fractions were collected and concentrated in a 10 kDa MWCO concentrator and then flash frozen in liquid N_2, and stored at -80 °C until further use. The expression and purification procedure was identical for wild-type or mutants GST-CENP-N sequences, except that the concentration of NaCl in the lysis buffer was raised to 500 mM NaCl (instead of 300 mM).

CENP-C fragments encoding CENP-C^{2-545} or CENP-$C^{225-364}$ were obtained from codon-optimized CENP-C cDNA and subcloned in pGEX-6P-2rbs for 3′ fusion to the sequence encoding GST. Mutant CENP-C constructs were created by site-directed mutagenesis using the QuikChange kit (Agilent Technologies). Expression and purification for all CENP-C constructs and mutants was carried out as described (Klare et al., 2015). All constructs were sequence-verified.

Nucleosome reconstitution Plasmids for the production of *X. laevis* H2A, H2B, H3 and H4 were a gift from D. Rhodes. Plasmids for the production of human CENP-A:H4 histone tetratmer were a gift from A.F. Straight. 145-bp DNA (601-Widom) wrapped around CENP-A or H3 octamers was a gift from C.A. Davey. Purification of CENP-A or H3 containing NCPs were performed as previously described (Guse et al., 2011; Guse et al., 2012; Weir et al., 2016). For cryo-EM studies, nucleosomes were reconstituted from recombinant human H2A, H2B, CENP-A, and H4. Nucleosomes were reconstituted using the salt dialysis method (Dyer et al., 2004). '601' DNA (145 bp), H2A:H2B dimer, and CENP-A:H4 (or H3:H4) dimer were mixed at molar ratios of 1 to 2.4 to 2.4 in buffer containing 20 mM Tris.Cl pH 7.5, 2M NaCl, and 1 mM EDTA. Sample was transferred to a dialysis tube (D-tube Dialyzer, EMD Millipore), and dialyzed against RB high (Reconstitution Buffer high: 20 mM Tris.Cl pH7.5, 2 M NaCl, 1 mM EDTA and 1 mM DTT) for 4 hours at 4°C. Gradient dialysis was set up as described (Dyer et al., 2004). A peristaltic pump (flow rate 1.5 ml/minute) was used for replacing RB high with RB low (Reconstitution Buffer low: 20 mM Tris.Cl pH7.5, 0.25 M NaCl, 1

517 mM EDTA and 1 mM DTT) for 18 hours. Finally, the sample was dialyzed against
518 storage buffer (20 mM Tris.Cl pH7.5, 20 mM NaCl, 1 mM EDTA and 1 mM DTT) for 8
519 hours.

520 **GST pull-down assays** GST pulldown assays were performed as previously described
521 (Klare et al., 2015) with minor modifications. Briefly, pre-blocked GSH Sepharose beads
522 were incubated in pull-down buffer consisting of 10 mM HEPES pH 7.5, 2.5% glycerol,
523 1 mM TCEP and either 300 mM NaCl for GST-LN-CENP-C assays or 150 mM NaCl
524 for GST-LN-NCPs assays. GST-CENP-LN or GST-CENP-N^{1-212} as baits (at a 1 μM
525 overall concentration) were incubated with either NCPs or CENP-C as prey at a 3 μM
526 concentration. The bait was loaded onto 10 μl pre-blocked beads before adding the prey.
527 The reaction volume was topped up to 40 μl with buffer and incubated at 4°C for 1 h.
528 Beads were spun down at 500*g* for 3 min. The supernatant was removed and beads were
529 washed twice with 200 μl pull-down buffer supplemented with 0.05% Triton-X-100.
530 Supernatant was removed completely, samples boiled in 15 μl Laemmli sample loading
531 buffer and run on a 14% SDS–PAGE gel. Bands were visualized with Coomassie
532 brilliant blue staining. For pre-blocking of GSH sepharose beads, 750 μl of GSH
533 sepharose beads were washed twice with 1 ml washing buffer (20 mM HEPES pH 7.5,
534 200 mM NaCl) and incubated in 1 ml blocking buffer (20 mM HEPES pH 7.5, 500 mM
535 NaCl, 500 μg/μl BSA) overnight at 4°C with rotation. Beads were washed five times
536 with 1 ml washing buffer and resuspended in 500 μl washing buffer to have a 50/50
537 slurry of beads and buffer.

538 **Analytical SEC analysis** Analytical SEC was performed on either Superdex 200 5/150
539 or Superose 6 5/150 columns in a buffer containing 20 mM HEPES (pH 7.5), 300 mM
540 NaCl, 2.5% glycerol, 2 mM TCEP, pH 7.5 on an ÄKTAmicro system. Proteins were
541 mixed at 6 μM in a total volume of 50 μl, incubated for one hour on ice, spun for 15
542 minutes in a bench-top centrifuge before each injection. All samples were eluted under
543 isocratic conditions at 4°C in SEC buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 2.5%
544 glycerol, 2 mM TCEP) at a flow rate of 0.1 ml/min. Elution of proteins was monitored
545 at 280 nm. Fractions (100 μl) were collected and analyzed by SDS–PAGE and
546 Coomassie blue staining.

547 **CENP-N:CENP-A nucleosome complex preparation** Complexes containing CENP-
548 A nucleosome and CENP-N^{NT} (CENP-N:CENP-A nucleosome complex) were

generated by mixing CENP-A nucleosome and CENP-N^{NT} at a molar ratio of 1:3, then dialyzed against 20 mM Tris.Cl pH 7.5, 50 mM NaCl, 1 mM EDTA and 1 mM DTT. The complexes were concentrated to approximately 2.5 μM using a Millipore concentrator (MW 50 kD cut-off).

EMSA assays Nucleosomes were mixed with CENP-N^{NT} constructs in buffer containing 20 mM Tris.Cl, pH 7.5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT and 0.1% CHAPS for 30 minutes at room temperature. The mixture was analysed by 5% native PAGE, and the gel was stained with SYBR Gold. The gel was imaged at 488 nm (for SYBR Gold staining or Alexa488 labelled H2B in the nucleosome) and 647 nm (for Atto N 647 labelled H2B in the nucleosome) using a Typhoon imager. Intensity at 647 nm was used for binding curves, with the number of replicates indicated. The intensity of the nucleosome band (not the shifted bands) was quantified, and normalized to the nucleosome sample without CENP-N.

Isothermal titration calorimetry All protein samples were loaded onto a Superdex 200 10/300 column equilibrated in 20 mM HEPES pH 7.5, 2.5% glycerol, 300 mM NaCl and 1 mM TCEP prior to ITC runs. ITC measurements were performed at 25°C on an ITC200 micro calorimeter (GE Healthcare). In each titration, the protein in the cell (at a 5-8 μM concentration) was titrated with 19 x 2 μl injections (at 180 sec intervals) of protein ligand (at 50-80 μM concentration). The injections were continued beyond saturation to allow for determination of heats of ligand dilution. Data were fitted by least-square procedure to a single site binding model using ORIGIN software package (MicroCal).

Mammalian plasmids Plasmids for stable cell lines were generated in pCDNA5/FRT/TO-EGFP-IRES, a modified version of the pCDNA5/FRT/TO vector (Invitrogen, Carlsbad, CA). The starting plasmid for EGFP expression was made by PCR amplifying the EGFP sequence of the pEGFP-C1 plasmid (Takara Bio Inc.) and cloning it into the pcDNA5/FRT/TO vector previously modified to carry an internal ribosomal entry site (IRES) sequence creating the pcDNA5/FRT/TO EGFP-IRES vector (Petrovic et al., 2010). Mammalian expression plasmids used in this study to express N-terminally tagged CENP-C WT or the PEST mutant were derived from the pCDNA5/FRT/TO-EGFP-IRES and used either for genomic integration or transient expression of human CENP-C proteins. A modified version of the pCDNA5/FRT/TO-EGFP-IRES plasmid where EGFP was removed and replaced with mCherry to create

582 the pCDNA5/FRT/TO-mCherry-IRES was used to express all C-terminally tagged
583 CENP-N constructs either via genomic integration or transient expression. To create the
584 EGFP tagged CENP-C protein and CENP-N tagged with mCherry protein full-length
585 proteins were amplified by PCR from full-length human codon-optimized cDNA
586 synthesized by GeneArt (Life Technologies). Using the Gibson cloning method[36] the
587 cDNA sequence was ligated with the PCR amplified pCDNA5/FRT/TO-EGFP-IRES
588 or pCDNA5/FRT/TO-mCherry-IRES to create the final constructs used in this study.
589 Each plasmid was then sequence verified before use.

590 **Cell culture and transfection** U2OS cells, a gift from A. Bird (MPI-Dortmund,
591 Germany), were grown in DMEM (PAN Biotech, Aidenbach, Germany) supplemented
592 with 10% FBS (Clontech, part of Takara Bio group, Shiga, Japan), penicillin and
593 streptomycin (GIBCO, Carlsbad, CA), and 2 mM L-glutamine (PAN Biotech). Transient
594 transfections were performed using pCDNA5-FRT-TO plasmids carrying either CENP-
595 N^{wt} or mutants with a C-terminal-mCherry tag transfected into asynchronously growing
596 cells and expressed for 24 h before preparation for immunofluorescence analysis. To
597 analyse whether the CENP-C^{5A} mutant abrogates kinetochore localization of CENP-N,
598 we used Flp-In T-REx HeLa cells (a gift from SS Taylor, University of Manchester,
599 Manchester, England, UK) to generate stable doxycycline-inducible cell lines, which were
600 maintained in DMEM (PAN Biotech, Aidenbach, Germany) with 10% tetracycline-free
601 FBS (Clontech) supplemented with 50 μg/ml Zeocin (Invitrogen) and 2 mM L-
602 glutamine (PAN Biotech). Flp-In T-REx HeLa expression cell lines were generated as
603 previously described (Krenn et al., 2012). FlpIn T-REx HeLa cells stably expressing
604 mCherry-CENP-N were grown on coverslips pre-coated with poly-D-Lysine (Millipore,
605 15 μg/ml). The day after cells were seeded, expression of mCherry-CENP-N was
606 induced using 0.2–0.5 μg/ml doxycycline (Sigma, St. Louis, MO). Further, cells were
607 transiently transfected with GFP-CENP-C^{WT} or GFP-CENP-C^{5A} mutant and CENP-C
608 siRNA for 72 h. When required 16 hours before fixation cells were synchronized
609 overnight in 0.33 μM nocodazole (Sigma–Aldrich) to assess mitotic localization.
610 Endogenous CENP-C was depleted by siRNA (target sequence: 5′-
611 GGAUCAUCUCAGAAUAGAA-3′ obtained from ThermoFisher), transfected into
612 cells using Lipofectamine RNAiMAX Transfection Reagent (ThermoFisher) as per the
613 manufactures instructions for 72 h.

614 **Immunofluorescence** Either U2OS or FlpIn T-REx HeLa cells were fixed using 4%

paraformaldehyde in phosphate buffered saline (PBS), permeablised using 0.25 % Triton X-100 in PBS and blocked in 3 % BSA/PBS. U2OS cells were stained for with the following antibodies: anti-CENP-C [rabbit polyclonal antibody SI410, diluted 1:1000 (Trazzi et al., 2009); and CREST/anti-centromere antibody (Antibodies Inc. 15-234-0001), diluted 1:100]. FlpIn T-REx HeLa cells were stained for GFP (GFP-Boost, Chromotek gba-488, 1:500), mCherry (RFP-Boost, Chromotek, rba-594, 1:500), CENP-C [rabbit polyclonal antibody SI410; 1:1000 (Trazzi et al., 2009)], CREST/anti-centromere antibodies (Antibodies, Inc., Davis, CA, 1:100), or CENP-A (Mouse, Gene Tex GTX13939, 1:500) diluted in blocking buffer for 2-4 h. Donkey anti-human Alexa Fluor 405 and donkey anti-rabbit Alexa Fluor 488, donkey anti-human (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA), as well as donkey anti-mouse (Invitrogen) were used as secondary antibodies. DNA was stained with 0.5 μg/ml DAPI (Serva), and coverslips were mounted with Mowiol mounting media (Calbiochem). U2OS cells were imaged with a Deltavision Elite System (GE Healthcare, UK) equipped with an IX-71 inverted microscope (Olympus, Japan), a PLAPON 60x/1.42NA objective and a pco.edge sCMOS camera (PCO-TECH Inc., USA). Images were acquired as Z-sections (using the softWoRx software from Deltavision) and converted into maximal intensity projections TIFF files for illustrative purposes while FlpIn T-REx HeLa Cells were imaged at room temperature using a spinning disk confocal device on the 3i Marianas system equipped with an Axio Observer Z1 microscope (Zeiss), a CSU-X1 confocal scanner unit (Yokogawa Electric Corporation, Tokyo, Japan), Plan-Apochromat 63× or 100×/1.4NA Oil Objectives (Zeiss), and Orca Flash 4.0 sCMOS Camera (Hamamatsu). Images were acquired as z-sections at 0.2 μm. Images were converted into maximal intensity projections, exported, and converted into 8-bit. Quantification of kinetochore signals was performed on unmodified 16-bit z-series images using Imaris 7.3.4 32-bit software (Bitplane, Zurich, Switzerland). Measurements were exported in Excel (Microsoft) and graphed with GraphPad Prism 6.0 (GraphPad Software, San Diego California USA). For quantification in Figure 3b-c, after background subtraction, all signals were normalized to CREST and the values obtained were then normalized by the mean of the CENP-N-mCherryWT construct. Quantifications are based on one experiment where a range of 7 to 10 cells and 177 to 393 kinetochores per condition were analysed.

Crystallization of CENP-N$^{1-235}$ His and structure determination Initial crystallization hits of CENP-N$^{1-212}$His6 (His6, hexahistidine tag) or CENP-N$^{1-235}$His6

649 were obtained in sitting drop crystallization experiments at c.a. 6 mg/ml in a 96 well
650 format using a Mosquito protein crystallization robot (TTP Labtech) at 4°C. Crystals
651 grew in various commercial screens including Qiagen Nextal PEGSII conditions B11
652 and B12 and Qiagen Nextal PEG conditions H6 and H8 within 24-48 hours, reaching
653 maximum size in 5-7 days. CENP-N^{1-235}His6 crystals were further optimized in 24-well
654 plates via hanging drop method using a two-dimensional grid screen varying PEG3350
655 (from 6%-16%) and pH (from 6.6-7.2). Selenomethionine (SeMet) derivatives of CENP-
656 N^{1-235}His6 were crystallized under similar conditions. Crystals were cryo-cooled in a
657 mother liquor solution containing 20-25% (v/v) glycerol. All data was collected at 100K
658 using a Pilatus 6M detector either at the X10SA beamline at the SLS in Villigen,
659 Switzerland, or at the P11 beamline of PETRA in Hamburg, Germany. All data sets were
660 integrated and scaled using XDS and XSCALE (Kabsch, 2010).

661 Both native and selenomethionine (SeMet) crystals grew in space group $P4_1$ with two
662 molecules per asymmetric unit and a relatively similar packing, although the c-axis is
663 approximately 4 Å shorter in the latter. Phasing with PHENIX (Adams et al., 2010)
664 located 12 of the 14 possible SeMet sites (in each monomer a C-terminal Met is
665 disordered, and M167 has very weak anomalous density). Merging of SeMet datasets
666 from two different crystals was essential to improve the anomalous signal. The quality of
667 the phases allowed autobuilding of (mostly) alpha helices into the electron density in
668 spite of the relatively low resolution (3.3 Å). Molecular replacement with PHASER
669 (Collaborative Computational Project, 1994) successfully placed this initial model into
670 the native dataset (conservative resolution at I/sigma=3 of 2.89 Å, data used to 2.74 Å).
671 The sequence was assigned with the help of the anomalous peaks. Refinement with
672 REFMAC (Collaborative Computational Project, 1994) and PHENIX resulted in a
673 model with very good Ramachandran geometry (98% residues in favoured regions, 0%
674 outliers) and R_{work}/R_{free} values of 21.6 and 26.0%, respectively. Data to 2.74 Å were used
675 despite high R factors, since they improved the convergence and quality of the
676 refinement.

677 Monomers A and B in the asymmetric unit are quite similar except for a minor hinge-
678 bending between the two subdomains of each monomer, and except for loops 137-142
679 and 164-174. In monomer B, the 137-142 loop is pulled approximately 3 Å away from
680 the remainder of the molecule by a symmetry contact, and in turn pulls on the
681 neighbouring loop causing a 2.6-Å movement of the backbone of K117. This loop has
682 no crystal contacts in monomer A. Similarly, residues 164-174 of monomer B pack

against a symmetry related molecule, most likely causing the observed (relative) stabilization of M167 in monomer B and backbone shifts of up to 7.3 Å relative to monomer A. Residues 166-168 have very weak density especially in chain A, corresponding to a weak and multiple anomalous density of SeM167, indicating multiple conformations of this loop. The coordinates of CENP-N have been submitted to the protein data bank (PDB) with code 6EQT.

CryoEM grid preparation and microscopy Quantifoil 2/2 grids (Quantifoil Micro Tools GmbH) were used for the CENP-N^{NT}:CENP-A^{NCP} complex (2.5 μM concentration). The grids were glow discharged (EMItec) at 40 mA for 20 seconds. 4 μl sample was applied onto the grid before plunge freezing into Ethane, using a Vitrobot (FEI, MARK IV). Blot time was 4 seconds. All grids were stored in liquid nitrogen before imaging. CENP-N^{NT}:CENP-A^{NCP} complex was imaged at nominal magnification of 29000x on a FEI Titan Krios (300 kV), equipped with a Gatan K2 Summit direct detector. Pixel size was 1.02 Å. The movies were captured in super resolution mode with electron dose rate at 10 electrons per pixel per second for 8 seconds and 0.2 seconds per frame. Defocus range was -1.0 to -2.5 μm.

Single particle analysis of cryoEM images Motioncor2 was used for the alignment of images (motion correction) (Zheng et al., 2017). GCTF was applied for Constant transfer function (CTF) estimation (Zhang, 2016). Images were evaluated manually by inspecting their power spectra. Particles were manually picked for the initial 2D classification (10 class averages) in RELION 2.05 (Fernandez-Leiro and Scheres, 2017). Initial 2D class averages were used for particle auto-picking as described in Relion2.0 tutorials. Particles from auto-picking were extracted and sorted for reference free 2D classification. In 2D classification, 200 class averages were generated. Noisy class averages were discarded. Particles from the retained class averages were used for reference free 3D reconstruction in cryoSPARC (Punjani et al., 2017). The low pass filtered map from the 3D reconstruction in cryoSPARC was used as reference for 3D classification in RELION 2.05 (Fernandez-Leiro and Scheres, 2017). Four 3D classes were created after 3D classification. Particle images from the class at high resolution were used for 3D refinement in Relion 2.1b1. The map was then sharpened in "post-process", using a mask file created by Relion2.1b1. The local resolution of the map was estimated by Relion (Kucukelbir et al., 2014). Anisotropy was analysed by 3DFSC calculation (Tan et al., 2017).

CryoEM structure modelling The original 3D refined map and post-processed map were used for the model fitting and refinement. DNA was taken from a high-resolution crystal structure of a nucleosome with 601 DNA [Protein Data Bank accession code 3LZ0 (Vasudevan et al., 2010)], CENP-A containing histone core was taken from Protein Data Bank accession code 3AN2 (Tachiwana et al., 2011). The crystal structure of CENP-N (amino acids 1-212) was used to fit into the remaining density using UCSF Chimera (Pettersen et al., 2004) (rigid body without allowing flexibility). Based on map density, the model was iteratively modified and locally refined in Coot (Emsley et al., 2010). The final model was subjected to real space refinement in PHENIX (Adams et al., 2010).

Acknowledgements

We thank Chuan Hong, Rick Huang, and Zhiheng Yu at the HHMI Janelia Farm CryoEM Facility for help in microscope operation and data collection, Bridget Carragher and Clint Potter at NRAMM for data collection and advice, staff at the University of Colorado Boulder EM Services Core Facility, the beamline staff of X10SA at the Swiss Light Source Paul Scherrer Institute, Villigen (Switzerland), and of P11 at the PETRA synchrotron, Hamburg (Germany), for support, and our colleagues of MPI Dortmund for help with the data collection, Christos Gatsogiannis for advice on the EM data, Kerstin Klare for initial contributions to the project, and Doro Vogt, Annika Take, and Sabine Wohlgemuth for excellent technical assistance. A.M. acknowledges funding by the the Horizon 2020 ERC agreement RECEPIANCE (AdG 669686), and the DFG's Collaborative Research Centre (CRC) 1093. K.L. received funding from NIH (GM067777), and from the Howard Hughes Medical Institute. Some of this work was performed at the National Resource for Automated Molecular Microscopy located at the New York Structural Biology Center, supported by grants from the NIH National Institute of General Medical Sciences (GM103310) and the Simons Foundation (349247). The authors declare no competing financial interests.

Author Contributions

SP and KZ expressed and purified proteins and performed analytical biochemical experiments, including experiments with CENP-N mutants. SP, with help from AP, performed crystallization trials and ITC assays. CS and SM performed fluorescence

localization experiments in U2OS and HeLa cells and related data processing. IRV determined the crystal structure of CENP-N and contributed to the interpretation of the EM maps. KZ, with GPM and KL determined the EM structure. In their respective laboratories, KZ and KL, and SP, IRV, AP, and AM, conceived the mutational analysis of CENP-N, which was executed by KZ and SP. JRW provided instrumental intellectual and practical guidance in the initial training of SP. AM and KL coordinated the respective working teams and their collaboration. AM drafted the paper with substantial contributions from SP, KZ, IRV, and KL. All authors read, commented, and approved the manuscript.

References

Adams, P.D., P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, and P.H. Zwart. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 66:213-21.

Armache, K.J., J.D. Garlick, D. Canzio, G.J. Narlikar, and R.E. Kingston. 2011. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 A resolution. *Science.* 334:977-82.

Bieniossek, C., T. Imasaki, Y. Takagi, and I. Berger. 2012. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci.* 37:49-57.

Black, B.E., D.R. Foltz, S. Chakravarthy, K. Luger, V.L. Woods, Jr., and D.W. Cleveland. 2004. Structural determinants for generating centromeric chromatin. *Nature.* 430:578-82.

Carroll, C.W., K.J. Milks, and A.F. Straight. 2010. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J Cell Biol.* 189:1143-55.

Carroll, C.W., M.C. Silva, K.M. Godek, L.E. Jansen, and A.F. Straight. 2009. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat Cell Biol.* 11:896-902.

Cheeseman, I.M., and A. Desai. 2008. Molecular architecture of the kinetochore-microtubule interface. *Nat Rev Mol Cell Biol.* 9:33-46.

Collaborative Computational Project, N. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 50:760-3.

Dambacher, S., W. Deng, M. Hahn, D. Sadic, J. Frohlich, A. Nuber, C. Hoischen, S. Diekmann, H. Leonhardt, and G. Schotta. 2012. CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. *Nucleus.* 3:101-10.

Drinnenberg, I.A., S. Henikoff, and H.S. Malik. 2016. Evolutionary Turnover of Kinetochore Proteins: A Ship of Theseus? *Trends Cell Biol.* 26:498-510.

Dunleavy, E.M., G. Almouzni, and G.H. Karpen. 2011. H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase. *Nucleus.* 2:146-57.

Dunleavy, E.M., D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, Y. Nakamura, Y. Daigo, Y. Nakatani, and G. Almouzni-Pettinotti. 2009. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell.* 137:485-97.

Dyer, P.N., R.S. Edayathumangalam, C.L. White, Y. Bao, S. Chakravarthy, U.M. Muthurajan, and K. Luger. 2004. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* 375:23-44.

Emsley, P., B. Lohkamp, W.G. Scott, and K. Cowtan. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 66:486-501.

Fachinetti, D., H.D. Folco, Y. Nechemia-Arbely, L.P. Valente, K. Nguyen, A.J. Wong, Q. Zhu, A.J. Holland, A. Desai, L.E. Jansen, and D.W. Cleveland. 2013. A two-step mechanism for epigenetic specification of centromere identity and function. *Nat Cell Biol.* 15:1056-66.

Falk, S.J., L.Y. Guo, N. Sekulic, E.M. Smoak, T. Mani, G.A. Logsdon, K. Gupta, L.E. Jansen, G.D. Van Duyne, S.A. Vinogradov, M.A. Lampson, and B.E. Black. 2015. Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science.* 348:699-703.

Fang, J., Y. Liu, Y. Wei, W. Deng, Z. Yu, L. Huang, Y. Teng, T. Yao, Q. You, H. Ruan, P. Chen, R.M. Xu, and G. Li. 2015. Structural transitions of centromeric chromatin regulate the cell cycle-dependent recruitment of CENP-N. *Genes Dev.* 29:1058-73.

Farnung, L., S.M. Vos, C. Wigge, and P. Cramer. 2017. Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature.* 550:539-542.

Fernandez-Leiro, R., and S.H.W. Scheres. 2017. A pipeline approach to single-particle processing in RELION. *Acta Crystallogr D Struct Biol.* 73:496-502.

Foltz, D.R., L.E. Jansen, A.O. Bailey, J.R. Yates, 3rd, E.A. Bassett, S. Wood, B.E. Black, and D.W. Cleveland. 2009. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell.* 137:472-84.

Foltz, D.R., L.E. Jansen, B.E. Black, A.O. Bailey, J.R. Yates, 3rd, and D.W. Cleveland. 2006. The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol.* 8:458-69.

French, B.T., F.G. Westhorpe, C. Limouse, and A.F. Straight. 2017. Xenopus laevis M18BP1 Directly Binds Existing CENP-A Nucleosomes to Promote Centromeric Chromatin Assembly. *Dev Cell.* 42:190-199 e10.

Fujita, Y., T. Hayashi, T. Kiyomitsu, Y. Toyoda, A. Kokubu, C. Obuse, and M. Yanagida. 2007. Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. *Dev Cell.* 12:17-30.

Fukagawa, T., and W.C. Earnshaw. 2014. The centromere: chromatin foundation for the kinetochore machinery. *Dev Cell.* 30:496-508.

Gascoigne, K.E., K. Takeuchi, A. Suzuki, T. Hori, T. Fukagawa, and I.M. Cheeseman. 2011. Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell.* 145:410-22.

Guo, L.Y., P.K. Allu, L. Zandarashvili, K.L. McKinley, N. Sekulic, J.M. Dawicki-McKenna, D. Fachinetti, G.A. Logsdon, R.M. Jamiolkowski, D.W. Cleveland, I.M. Cheeseman, and B.E. Black. 2017. Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. *Nat Commun.* 8:15775.

837 Guse, A., C.W. Carroll, B. Moree, C.J. Fuller, and A.F. Straight. 2011. In vitro
838 centromere and kinetochore assembly on defined chromatin templates. *Nature*.
839 477:354-8.

840 Guse, A., C.J. Fuller, and A.F. Straight. 2012. A cell-free system for functional
841 centromere and kinetochore assembly. *Nat Protoc*. 7:1847-69.

842 Hayashi, T., Y. Fujita, O. Iwasaki, Y. Adachi, K. Takahashi, and M. Yanagida. 2004.
843 Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at
844 centromeres. *Cell*. 118:715-29.

845 Hellwig, D., S. Emmerth, T. Ulbricht, V. Doring, C. Hoischen, R. Martin, C.P. Samora,
846 A.D. McAinsh, C.W. Carroll, A.F. Straight, P. Meraldi, and S. Diekmann. 2011.
847 Dynamics of CENP-N kinetochore binding during the cell cycle. *J Cell Sci*.
848 124:3871-83.

849 Hinshaw, S.M., and S.C. Harrison. 2013. An Iml3-Chl4 heterodimer links the core
850 centromere to factors required for accurate chromosome segregation. *Cell Rep*.
851 5:29-36.

852 Hoffmann, S., M. Dumont, V. Barra, P. Ly, Y. Nechemia-Arbely, M.A. McMahon, S.
853 Herve, D.W. Cleveland, and D. Fachinetti. 2016. CENP-A Is Dispensable for
854 Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly.
855 *Cell Rep*. 17:2394-2404.

856 Holm, L., and P. Rosenstrom. 2010. Dali server: conservation mapping in 3D. *Nucleic*
857 *Acids Res*. 38:W545-9.

858 Hori, T., W.H. Shang, M. Hara, M. Ariyoshi, Y. Arimura, R. Fujita, H. Kurumizaka, and
859 T. Fukagawa. 2017. Association of M18BP1/KNL2 with CENP-A Nucleosome
860 Is Essential for Centromere Formation in Non-mammalian Vertebrates. *Dev Cell*.
861 42:181-189 e3.

862 Hori, T., W.H. Shang, A. Toyoda, S. Misu, N. Monma, K. Ikeo, O. Molina, G. Vargiu, A.
863 Fujiyama, H. Kimura, W.C. Earnshaw, and T. Fukagawa. 2014. Histone H4 Lys
864 20 monomethylation of the CENP-A nucleosome is essential for kinetochore
865 assembly. *Dev Cell*. 29:740-9.

866 Izuta, H., M. Ikeno, N. Suzuki, T. Tomonaga, N. Nozaki, C. Obuse, Y. Kisu, N.
867 Goshima, F. Nomura, N. Nomura, and K. Yoda. 2006. Comprehensive analysis
868 of the ICEN (Interphase Centromere Complex) components enriched in the
869 CENP-A chromatin of human cells. *Genes Cells*. 11:673-84.

870 Jansen, L.E., B.E. Black, D.R. Foltz, and D.W. Cleveland. 2007. Propagation of
871 centromeric chromatin requires exit from mitosis. *J Cell Biol*. 176:795-805.

872 Kabsch, W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta*
873 *Crystallogr D Biol Crystallogr*. 66:133-44.

874 Kato, H., J. Jiang, B.R. Zhou, M. Rozendaal, H. Feng, R. Ghirlando, T.S. Xiao, A.F.
875 Straight, and Y. Bai. 2013. A conserved mechanism for centromeric nucleosome
876 recognition by centromere protein CENP-C. *Science*. 340:1110-3.

Klare, K., J.R. Weir, F. Basilico, T. Zimniak, L. Massimiliano, N. Ludwigs, F. Herzog, and A. Musacchio. 2015. CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. *J Cell Biol.* 210:11-22.

Krenn, V., A. Wehenkel, X. Li, S. Santaguida, and A. Musacchio. 2012. Structural analysis reveals features of the spindle checkpoint kinase Bub1-kinetochore subunit Knl1 interaction. *J Cell Biol.* 196:451-67.

Kucukelbir, A., F.J. Sigworth, and H.D. Tagare. 2014. Quantifying the local resolution of cryo-EM density maps. *Nat Methods.* 11:63-5.

Liu, X., M. Li, X. Xia, X. Li, and Z. Chen. 2017. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature.* 544:440-445.

Logsdon, G.A., E.J. Barrey, E.A. Bassett, J.E. DeNizio, L.Y. Guo, T. Panchenko, J.M. Dawicki-McKenna, P. Heun, and B.E. Black. 2015. Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. *J Cell Biol.* 208:521-31.

Lu, A., V. Kabaleeswaran, T. Fu, V.G. Magupalli, and H. Wu. 2014. Crystal structure of the F27G AIM2 PYD mutant and similarities of its self-association to DED/DED interactions. *J Mol Biol.* 426:1420-7.

Makde, R.D., J.R. England, H.P. Yennawar, and S. Tan. 2010. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature.* 467:562-6.

McKinley, K.L., N. Sekulic, L.Y. Guo, T. Tsinman, B.E. Black, and I.M. Cheeseman. 2015. The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. *Mol Cell.* 60:886-98.

Moree, B., C.B. Meyer, C.J. Fuller, and A.F. Straight. 2011. CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *J Cell Biol.* 194:855-71.

Musacchio, A., and A. Desai. 2017. A Molecular View of Kinetochore Assembly and Function. *Biology (Basel).* 6.

Nagpal, H., T. Hori, A. Furukawa, K. Sugase, A. Osakabe, H. Kurumizaka, and T. Fukagawa. 2015. Dynamic changes in CCAN organization through CENP-C during cell-cycle progression. *Mol Biol Cell.* 26:3768-76.

Narlikar, G.J., R. Sundaramoorthy, and T. Owen-Hughes. 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell.* 154:490-503.

Obuse, C., H. Yang, N. Nozaki, S. Goto, T. Okazaki, and K. Yoda. 2004. Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. *Genes Cells.* 9:105-20.

Okada, M., I.M. Cheeseman, T. Hori, K. Okawa, I.X. McLeod, J.R. Yates, 3rd, A. Desai, and T. Fukagawa. 2006. The CENP-H-I complex is required for the efficient

incorporation of newly synthesized CENP-A into centromeres. *Nat Cell Biol.* 8:446-57.

Petrovic, A., S. Pasqualato, P. Dube, V. Krenn, S. Santaguida, D. Cittaro, S. Monzani, L. Massimiliano, J. Keller, A. Tarricone, A. Maiolica, H. Stark, and A. Musacchio. 2010. The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *J Cell Biol.* 190:835-52.

Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 25:1605-12.

Pidoux, A.L., E.S. Choi, J.K. Abbott, X. Liu, A. Kagansky, A.G. Castillo, G.L. Hamilton, W. Richardson, J. Rappsilber, X. He, and R.C. Allshire. 2009. Fission yeast Scm3: A CENP-A receptor required for integrity of subkinetochore chromatin. *Mol Cell.* 33:299-311.

Przewloka, M.R., Z. Venkei, V.M. Bolanos-Garcia, J. Debski, M. Dadlez, and D.M. Glover. 2011. CENP-C is a structural platform for kinetochore assembly. *Curr Biol.* 21:399-405.

Punjani, A., J.L. Rubinstein, D.J. Fleet, and M.A. Brubaker. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* 14:290-296.

Ratsimandresey, R.A., A. Dorfleutner, and C. Stehlik. 2013. An Update on PYRIN Domain-Containing Pattern Recognition Receptors: From Immunity to Pathology. *Front Immunol.* 4:440.

Samejima, I., C. Spanos, L. Alves Fde, T. Hori, M. Perpelescu, J. Zou, J. Rappsilber, T. Fukagawa, and W.C. Earnshaw. 2015. Whole-proteome genetic analysis of dependencies in assembly of a vertebrate kinetochore. *J Cell Biol.* 211:1141-56.

Sanchez-Pulido, L., A.L. Pidoux, C.P. Ponting, and R.C. Allshire. 2009. Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell.* 137:1173-4.

Santaguida, S., and A. Amon. 2015. Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nat Rev Mol Cell Biol.* 16:473-85.

Scheres, S.H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol.* 180:519-30.

Schuh, M., C.F. Lehner, and S. Heidmann. 2007. Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. *Curr Biol.* 17:237-43.

Screpanti, E., A. De Antoni, G.M. Alushin, A. Petrovic, T. Melis, E. Nogales, and A. Musacchio. 2011. Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol.* 21:391-8.

Shono, N., J. Ohzeki, K. Otake, N.M. Martins, T. Nagase, H. Kimura, V. Larionov, W.C. Earnshaw, and H. Masumoto. 2015. CENP-C and CENP-I are key connecting factors for kinetochore and CENP-A assembly. *J Cell Sci.* 128:4572-87.

Tachiwana, H., W. Kagawa, T. Shiga, A. Osakabe, Y. Miya, K. Saito, Y. Hayashi-Takanaka, T. Oda, M. Sato, S.Y. Park, H. Kimura, and H. Kurumizaka. 2011. Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature*. 476:232-5.

Tan, S., R.C. Kern, and W. Selleck. 2005. The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. *Protein Expr Purif*. 40:385-95.

Tan, Y.Z., P.R. Baldwin, J.H. Davis, J.R. Williamson, C.S. Potter, B. Carragher, and D. Lyumkis. 2017. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods*. 14:793-796.

Trazzi, S., G. Perini, R. Bernardoni, M. Zoli, J.C. Reese, A. Musacchio, and G. Della Valle. 2009. The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. *PLoS One*. 4:e5832.

van Hooff, J.J., E. Tromer, L.M. van Wijk, B. Snel, and G.J. Kops. 2017. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep*. 18:1559-1571.

Vasudevan, D., E.Y. Chua, and C.A. Davey. 2010. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol*. 403:1-10.

Weir, J.R., A.C. Faesen, K. Klare, A. Petrovic, F. Basilico, J. Fischbock, S. Pentakota, J. Keller, M.E. Pesenti, D. Pan, D. Vogt, S. Wohlgemuth, F. Herzog, and A. Musacchio. 2016. Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature*. 537:249-253.

Williams, J.S., T. Hayashi, M. Yanagida, and P. Russell. 2009. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol Cell*. 33:287-98.

Zhang, K. 2016. Gctf: Real-time CTF determination and correction. *J Struct Biol*. 193:1-12.

Zheng, S.Q., E. Palovcak, J.P. Armache, K.A. Verba, Y. Cheng, and D.A. Agard. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*. 14:331-332.

Figure legends

Figure 1. The interaction of CENP-N with nucleosomes

A) Schematic of crucial CCAN and KMN subunits discussed in the text. The Knl1-Mis12-Ndc80 (KMN) complex is the main microtubule receptor at the kinetochore. Other interactions are discussed in the main text. The question mark indicates that the precise determinants for the recruitment of CENP-LN to CENP-C and for the interaction of CENP-N with the CENP-A nucleosome have not been identified. **B**) Schematic depicting constructs described in the manuscript. **C-E**) Solid phase binding assays where the indicated GST fusion proteins were immobilized on glutathione-sepharose beads (at a final concentration of 1 μM) and incubated with 3 μM of the indicated nucleosome core particles. After incubation (see Methods), beads were centrifuged, washed, and bound proteins visualized by SDS-PAGE and Coomassie staining.

Figure 1 Supplement 1

Further interactions of CENP-NL

Solid phase binding assay demonstrating that CENP-LN$^{230\text{-}C}$ does not interact with CENP-A nucleosome. The indicated GST fusion proteins were immobilized on glutathione-sepharose beads (at a final concentration of 1 μM) and incubated with 3 μM of the indicated nucleosome core particles. After incubation (see Methods), beads were pelleted, washed, and bound proteins visualized by SDS-PAGE and Coomassie staining.

Figure 2. Crystal structure of the CENP-A binding region of CENP-N

A) Cartoon model of CENP-N$^{1\text{-}235}$ with secondary structure and domain organization. **B**) Close-up of the boxed region in A. **C**) Topology diagram of CENP-N. The topology of the Pyrin and CLN-HD domains was directly derived from the crystal structure of CENP-N$^{1\text{-}235}$ reported here. The topology of the CENP-L binding domain was derived from the crystal structure of the Chl4 fragment in the complex of the Chl4$^{CENP\text{-}N}$:Iml3$^{CENP\text{-}L}$ yeast homolog (Hinshaw and Harrison, 2013). **D**) Multiple sequence alignment of CENP-N from the indicated species with secondary structure. Green, blue, and orange dots indicate solvent-exposed, semi-buried, and buried side chains, respectively. Positions with conserved residues are displayed red; positions with

conserved side chain charge are boxed. **E**) Schematic summarizing domain organization of CENP-L, CENP-N, and their dimerization.

Figure 2 Supplement 1

Structural relatedness of CENP-N N-terminal domain with PYRINs

A) Cartoon model of the "death fold" Pyrin domain of AIM2 (Lu et al., 2014). **B**) Structural superposition of the AIM2 Pyrin domain (red) with the CENP-N Pyrin domain.

Figure 2 Supplement 2

Topology of Iml3 and CENP-L

A) Overall domain organization of Iml3^{CENP-L} with cartoon model illustrating the CLN-HD in green and the CENP-L insert in yellow. The latter includes the Chl4^{CENP-N}-binding domain, which is conserved in humans, as shown in panel E (Hinshaw and Harrison, 2013). **B**) Topology diagram of CENP-L, based on the structure of the Iml3^{CENP-L} yeast homolog (Hinshaw and Harrison, 2013) and its comparison with the crystal structure of CENP-N after identification of structural homology by program DALI.

Figure 2 Supplement 3

Overall organization of the CENP-LN complex

A) Structure-based sequence alignment generated with program Chimera of the C-terminal region of CENP-N (model) and the C-terminal region of Chl4^{CENP-N} (structure) with β-strands shown in green and α-helices in yellow (secondary structure predictions in case of the model), and of the C-termini of CENP-L (model) and Iml3 (PDB structure 4KR1) with the same colour scheme. **B**) Schematics of the overall organization of the CENP-LN complex, and complementarity of the CENP-N$^{1-212}$ and 4JE3 structures. **C**) Structural superposition of the CLN-HD of Iml3^{CENP-L} (green) on the corresponding middle domain of CENP-N$^{1-235}$. **D**) Structure-based sequence alignment generated with program Chimera of CENP-N, CENP-L, and their yeast homologs Iml3^{CENP-L} and Chl4^{CENP-N}, with β-strands shown in green and α-helices in yellow, including the two structures (CENP-N and Iml3) and sequences and predicted secondary structures for the

two models (Chl4 and CENP-L).

Figure 3. The CENP-N:CENP-A^{NCP} complex

A) Cartoon model of the CENP-A^{NCP} with bound CENP-N$^{1-235}$, determined by cryo-EM. **B**) Surface representation of the complex. In A and B, the L1 loop of CENP-A is displayed in red. **C**) Comparison of the DNA ends in the crystal structure of the CENP-A nucleosome (Tachiwana et al., 2011) and in the structure of the CENP-A:CENP-N complex. **D**) Electrostatic potential at the CENP-N DNA binding interface with contour levels +/- 4 k_BT/e (k_B, Boltzmann constant; T, absolute temperature; e, the magnitude of electron charge, calculated with the APBS Pymol plugin). **E**) Interaction of CENP-N with backbone, minor groove, and major groove of DNA with close-up views of selected interactions. **F**) Interactions at the CENP-A L1 loop and comparison with superimposed H3.

Figure 3 Supplement 1

Additional EM data and analysis

A) Collected micrograph after CTF correction at -2.5 μm defocus (nominal). Scale bar = 100 nm. **B**) Representative 2D class averages from RELION (Scheres, 2012) 2D classification. **C**) Fourier Shell Correction (FSC) curve for the maps. **D**) Estimated local resolution for the CENP-N$^{1-289}$:CENP-A^{NCP} by RELION. The unit for colour scale is Å. **E**) Euler angular distribution of the 3D reconstruction. **F**) Histogram and directional FSC plot for unmasked map. Sphericity = 0.901 out of 1, global resolution = 4.08 Å.

Figure 3 Supplement 2

EM maps

A) Representative areas of the map of the CENP-N$^{1-289}$:CENP-A^{NCP} complex contoured at 3σ shows very clear density for the DNA ends and for the N-terminal helix of CENP-A, starting from residue G46. The grey arrow points to a zoomed-in region close to the DNA end. **B**) The zoomed-in region. **C**) Overall fit of the CENP-N$^{1-235}$ crystallographic model in the EM density. **D**) The H4 N-terminal region is clearly ordered starting at residue R23, possibly due to packing against the β3-β3 loop of CENP-N. **E**) Illustrative

example of model fit in the map around H2A. **F**) Illustrative example of model fit in the map around H2B.

Figure 3 Supplement 3

EMSA assays

A) EMSAs (electrophoretic mobility shift assays) with the indicated NCPs and CENP-N^{1-289}. **B**) Quantification of data in A. Data were fitted in Graphpad using the specific binding model with Hill coefficient fitting.

Figure 3 Supplement 4

Essential features of CENP-A

A) Cartoon model of the CENP-A nucleosome (this work). CENP-A is shown in light blue and in a zoomed-in view. **B**) Alignment of CENP-A and H3 sequences from three species showing divergence in the L1 loop region.

Figure 3 Supplement 5

Comparison of CENP-A and H3 and interface with CENP-N

Close-up view of the CENP-A:CENP-N interface with EM density map. The superposition with H3 (dark blue) illustrates that H3 would be unable to form productive contacts with CENP-N. V82 of CENP-A packs against the side chain of Y147 of CENP-N.

Figure 3 Supplement 6

Comparison of nucleosome binding modes

A-B) Two views in different orientation of the structure of CENP-N^{1-289}:CENP-A^{NCP} complex, with CENP-A in light blue and CENP-N in teal and cyan. **C-D**) Two views in different orientations of the SWI2/SNF2:$H3^{NCP}$ complex (Liu et al., 2017) illustrating similarities in the interactions with the nucleosome. The contacts of SWI/SNF with the proteinaceous part of the $H3^{NCP}$ particle are more limited than for CENP-N. **E-F**) Two views in different orientations of the BAH:$H3^{NCP}$ complex (Armache et al., 2011). The

contacts of the BAH domain with DNA are very limited, while there are extensive interactions with the H2A-H2B acidic patch and with the H4 N-terminal tail and with H3.

Figure 4. Validation of the CENP-N:CENP-A^{NCP} complex

A) *In vitro* binding assay probing the interaction of GST-CENP-N$^{1-212}$ immobilized on solid phase with CENP-A^{NCP}. **B**) Fluorescence microscopy analysis comparing localization at human kinetochores (U2OS osteosarcoma cells) of a wild type CENP-N-mCherry fluorescent reporter and of its mutant variants. **C**) Quantification of localization of the mCherry constructs in B normalized to CREST. The same concentrations of transiently transfected plasmids were compared.

Figure 4 Supplement 1

Characterization of CENP-N mutants in solid phase and cell assays

A) *In vitro* binding assay probing the interaction of wild type and mutant GST-CENP-N$^{1-212}$ immobilized on solid phase with CENP-A^{NCP}. **B**) Fluorescence microscopy analysis comparing kinetochore localization of a wild type CENP-N-mCherry fluorescent reporter and of the indicated mutant variants in U2OS cells. The same concentrations of transiently transfected plasmids were compared.

Figure 4 Supplement 2

Characterization of CENP-N mutants in competition gel shift assays

A-B) Gel shift assays with differentially labelled CENP-A^{NCPs} (red) and H3^{NCPs} (green) at 125 nM in presence of the indicated CENP-N wild type and mutant species (1 μM). The two individual channels are shown below. The nature of all slow-migrating species formed upon binding of CENP-N and its mutants to nucleosomes is unknown, but we suspect they represent higher-order oligomerization. Note that mutants affecting the CENP-A L1 loop-binding interface of CENP-N affect CENP-A^{NCPs} binding more than H3^{NCPs} binding, likely because the latter is solely based on the intact DNA-binding interface of CENP-N (see text for additional details).

Figure 5. Identification of a CENP-N binding site on CENP-C

A) Size exclusion chromatography (SEC) runs of CENP-$C^{225-364}$, CENP-LN complex, and their combination at the indicated loaded concentrations, identified a binding site for CENP-LN in CENP-$C^{225-364}$. Elution fractions were separated by SDS-PAGE and visualized by Coomassie staining. **B**) CENP-LN^{230-C} binds CENP-C^{2-545}, indicating that the CENP-N N-terminal region is not required for CENP-C binding. **C**) Sequence of a segment of the PEST-rich domain CENP-C that contains a binding site for the CENP-HIKM complex (Klare et al., 2015) (residues indicated in salmon). The CENP-N-binding motif is shown in grey. **D**) CENP-LN does not bind CENP-$C^{2-545-5A}$. **E**) CENP-$C^{2-545-5A}$ retains the ability to bind to CENP-A^{NCP}. **F**) CENP-$C^{2-545-5A}$ retains the ability to bind to the CENP-HIKM complex. **G**) Isothermal titration calorimetry (ITC) experiment quantifying the physical interaction of the CENP-L:CENP-N^{230-C} complex with CENP-$C^{225-364}$. **H**) In agreement with the SEC data, CENP-$C^{225-364-5A}$ fails to interact with the CENP-L:CENP-N^{230-C} complex in an ITC experiment.

Figure 5 Supplement 1

Additional size-exclusion chromatography experiments

A) SEC experiment demonstrating an interaction of CENP-C^{2-545} with CENP-LN complex, as shown previously (Weir et al., 2016). **B**) SEC experiment demonstrating that CENP-L does not bind CENP-C^{2-545}. **C**) SEC experiment demonstrating that CENP-N^{1-235} does not bind CENP-C^{2-545}.

Figure 5 Supplement 2

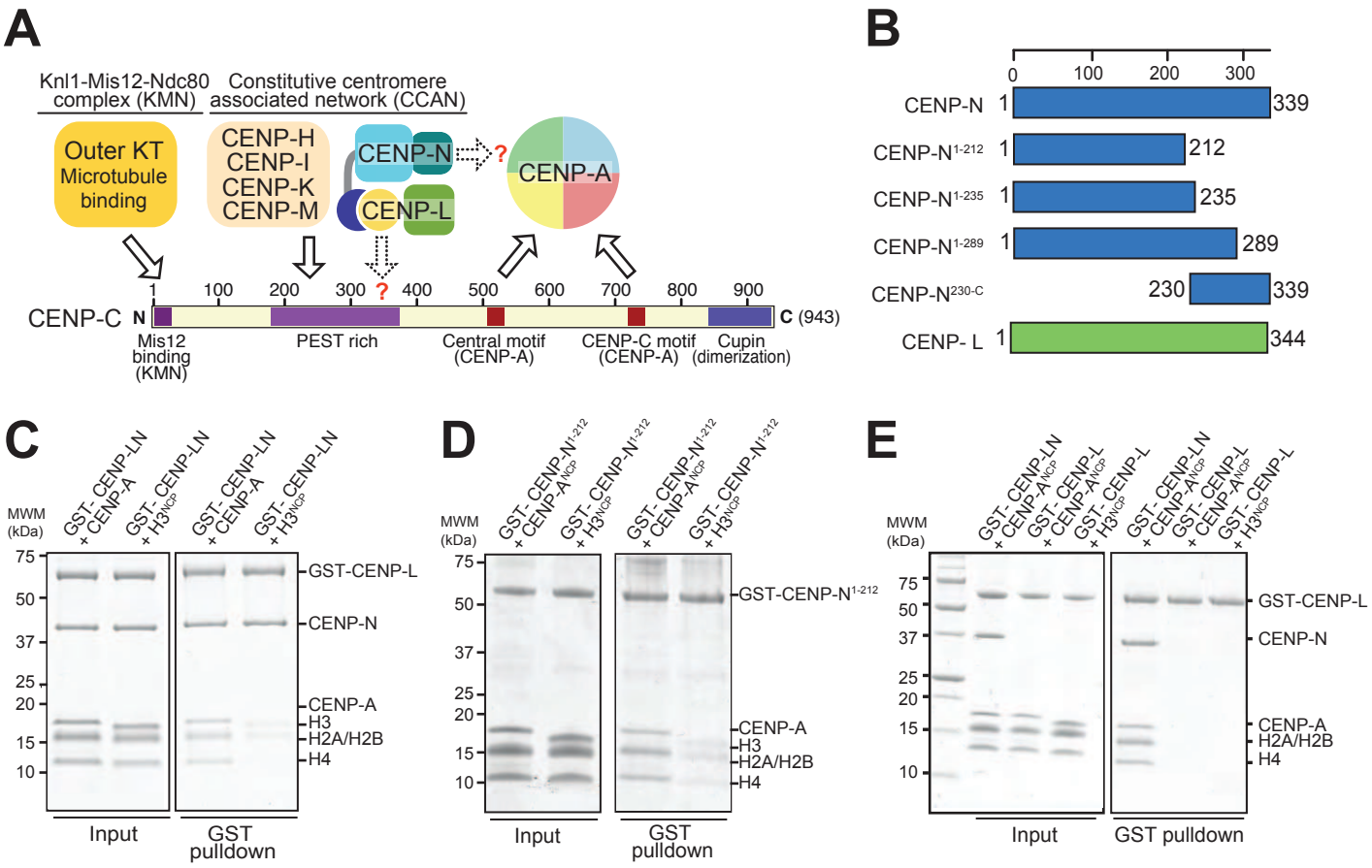
Solid phase binding assays with CENP-C and CENP-C mutant

Solid phase binding assay with immobilized GST-CENP-LN complex and soluble CENP-C^{2-545} or CENP-$C^{2-545-5A}$ mutant.

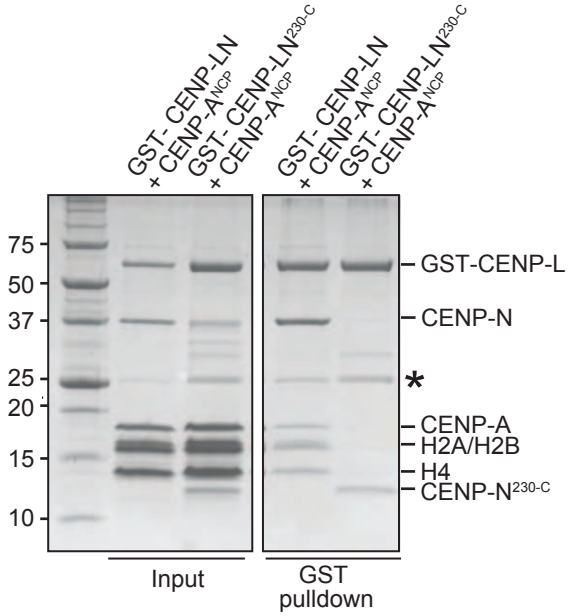
Figure 6. Effective CENP-N localization requires CENP-C

A) Fluorescence microscopy analysis comparing kinetochore localization of a wild type CENP-N-mCherry fluorescent reporter in human HeLa FlpIn TRex cells depleted of CENP-C and, were indicated, further expressing wild type GFP-CENP-C or the 5A

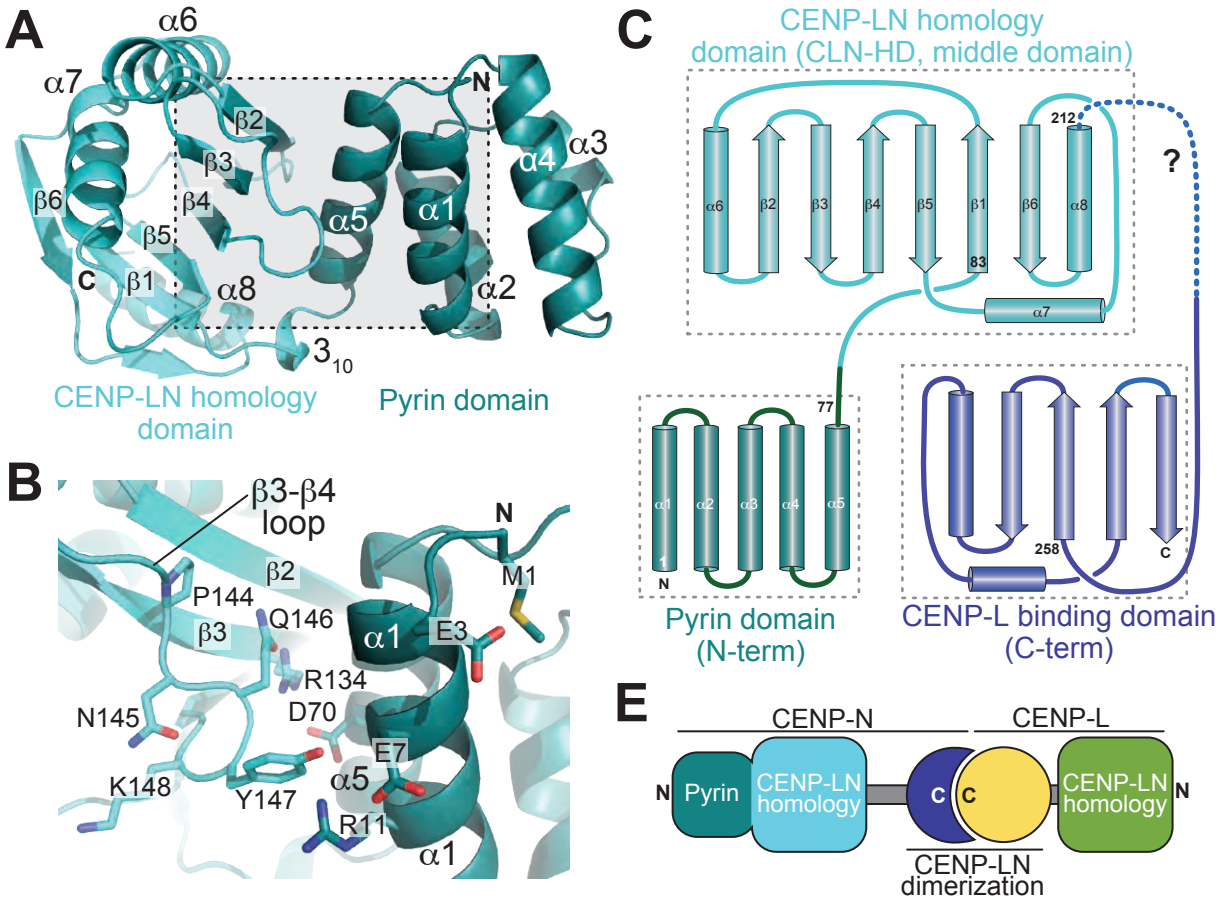
mutant. **B**) Quantification of CENP-C (left) and mCherry-CENP-N (right) levels at kinetochores in mitotic cells following the rescue of CENP-C depletion by either GFP-CENP-C^{WT} or the GFP-CENP-C^{5A} mutant. Graphs show kinetochore fluorescence intensity of the indicated protein (antibodies against CENP-C or mCherry) normalized to CENP-C or mCherry-CENP-N kinetochore levels in the absence of RNAi treatment, respectively. Each graph is representative of two independent experiments. **C**) Surface representation of a composite model built by combining the coordinates of the CENP-C motif (residues 712 to 733 from PDB ID 4X23, describing its interaction with nucleosome) with those of the CENP-N:CENP-A^{NCP} complex. **D**) Schematic of crucial kinetochore interactions, already shown in Figure 1A, but with question marks removed at interactions investigated in the present work. **E**) The grey box, an enlargement of the box in D, summarizes the details of the interactions reported in this work, as well as previous information on the interaction of CENP-C with the CENP-A^{NCP}.



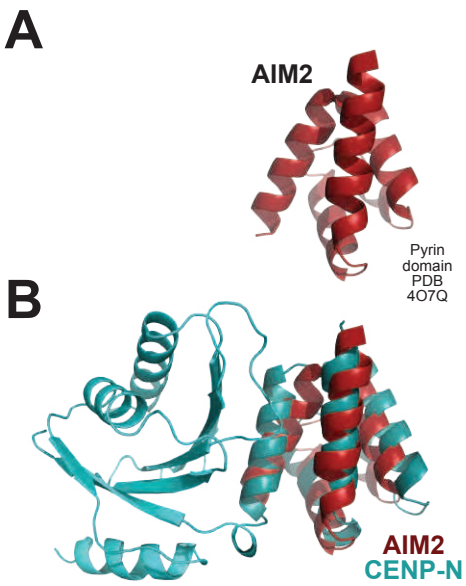
Pentakota, Zhou et al.
Figure 1



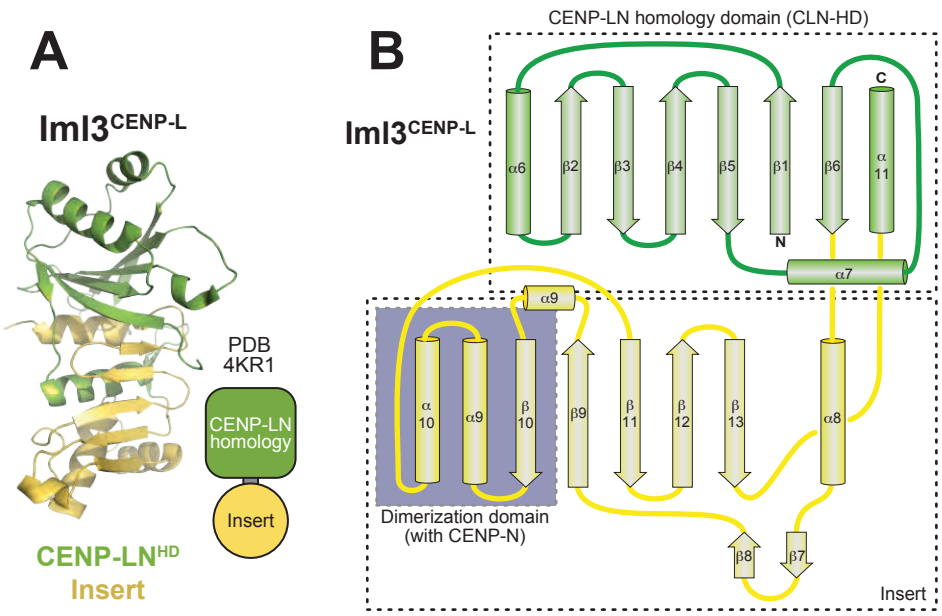
Pentakota, Zhou et al.
Figure 1 Supplement 1



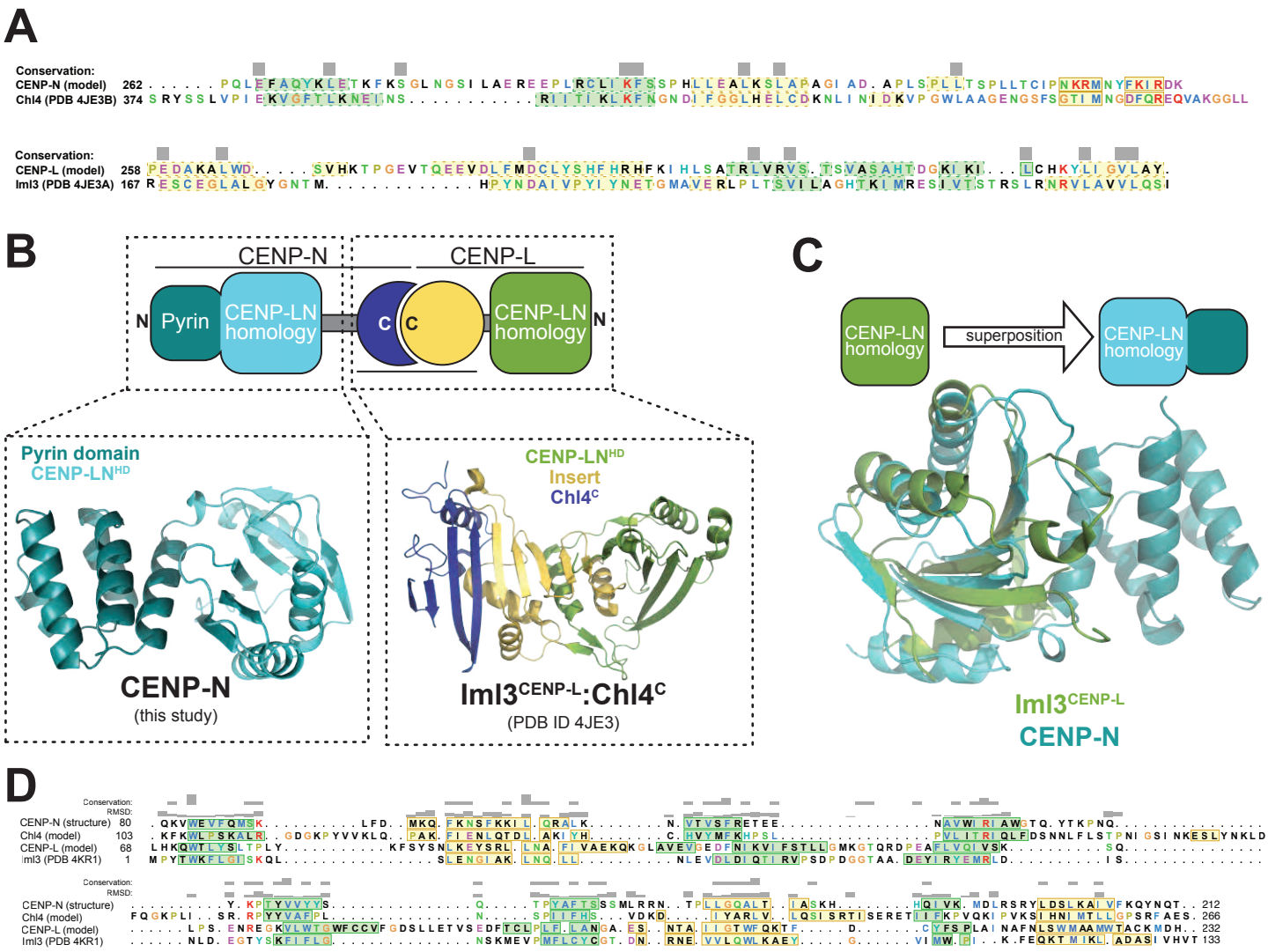
Pentakota, Zhou et al.
Figure 2



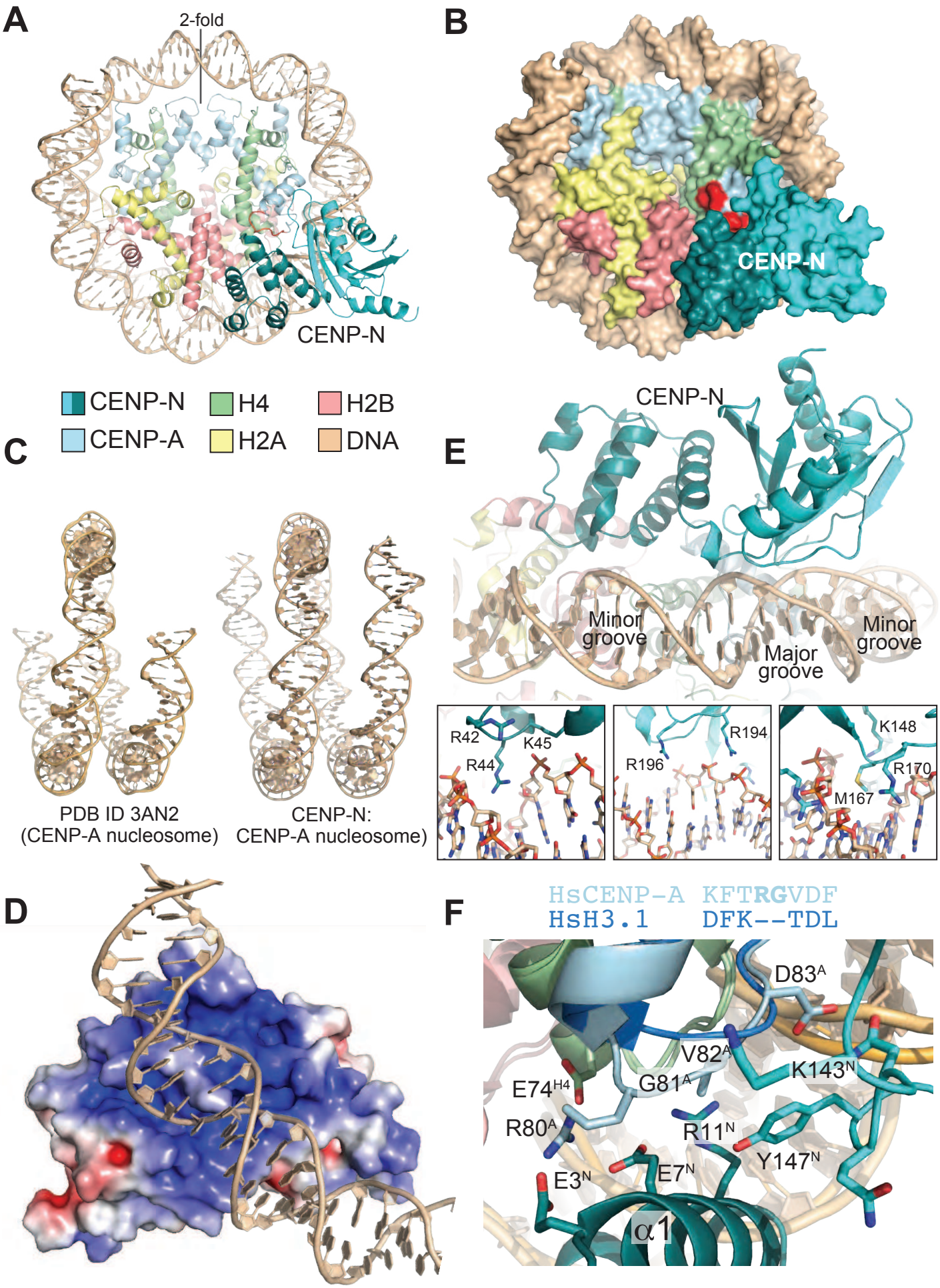
Pentakota, Zhou et al.
Figure 2 Supplement 1



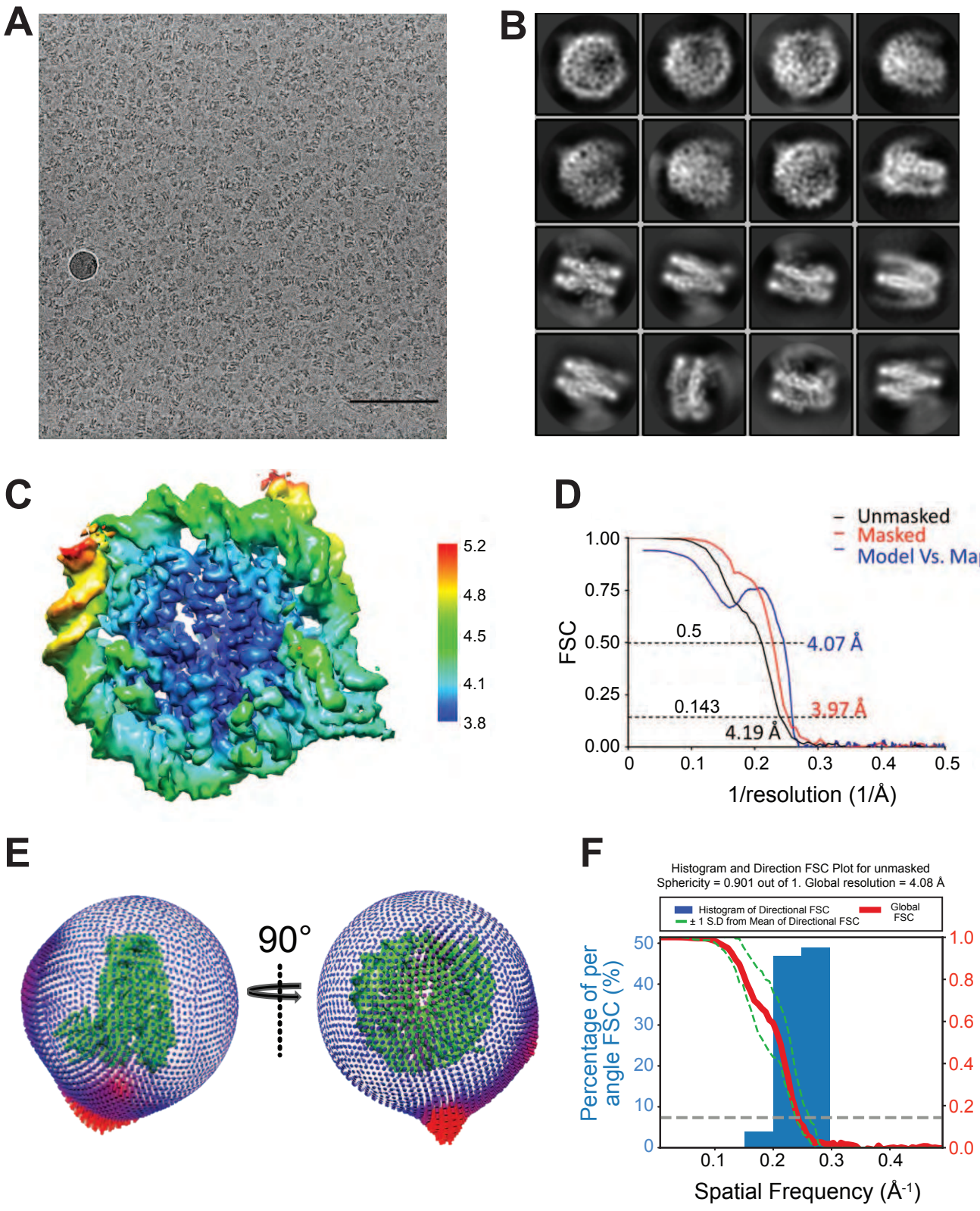
Pentakota, Zhou et al.
Figure 2 Supplement 2



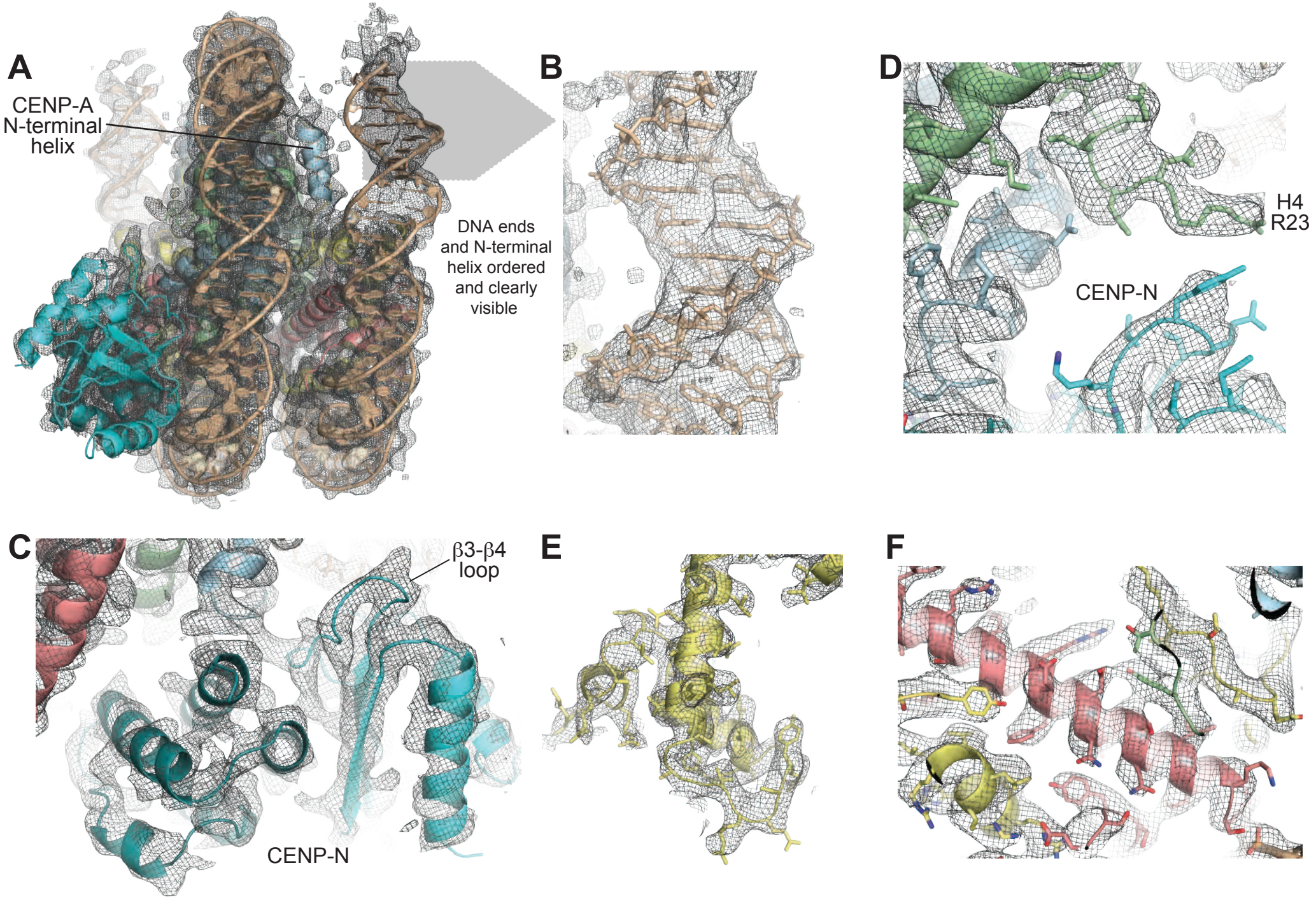
Pentakota, Zhou et al.
Figure 2 Supplement 3



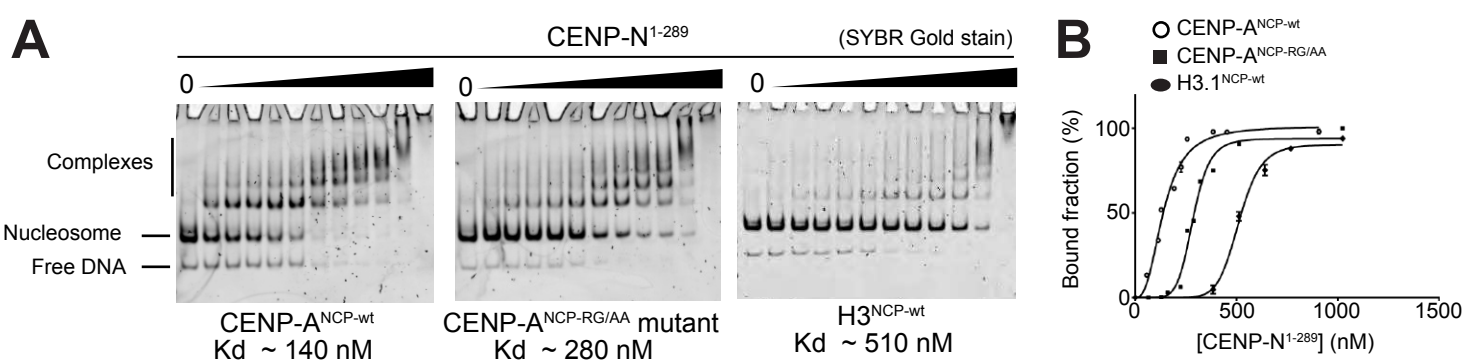
Pentakota, Zhou et al.
Figure 3



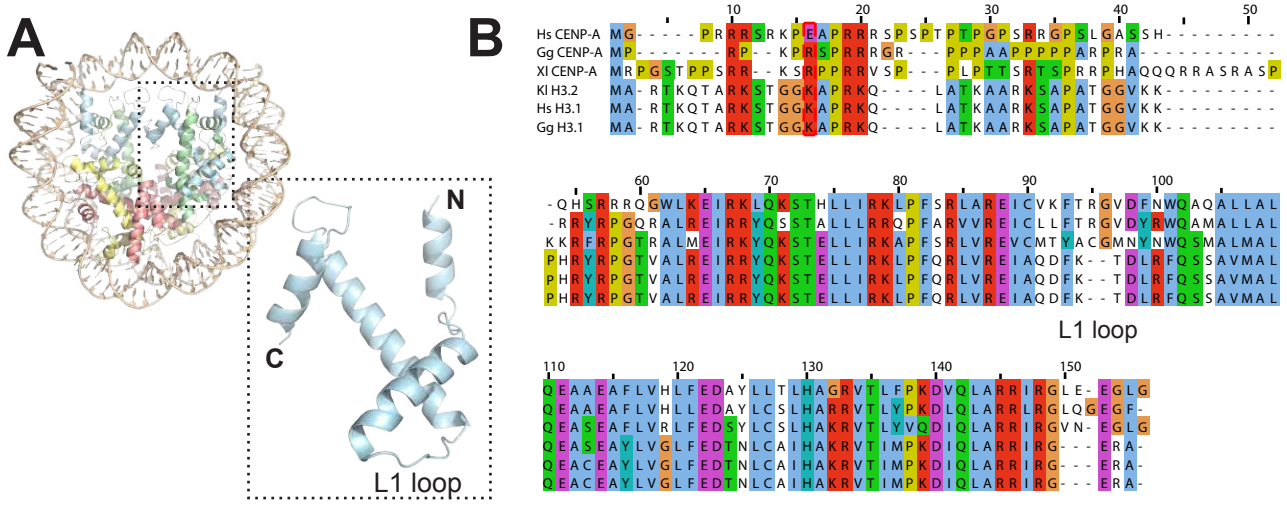
Pentakota, Zhou et al.
Figure 3 Supplement 1



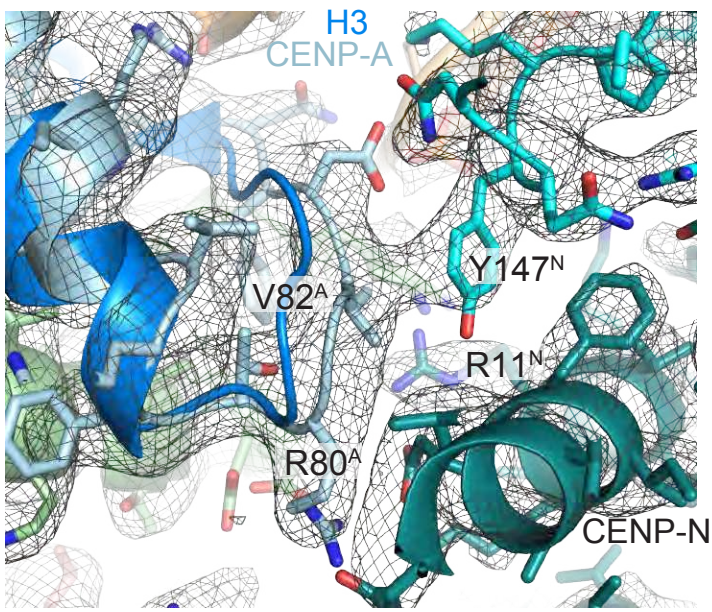
Pentakota, Zhou et al.
Figure 3 Supplement 2



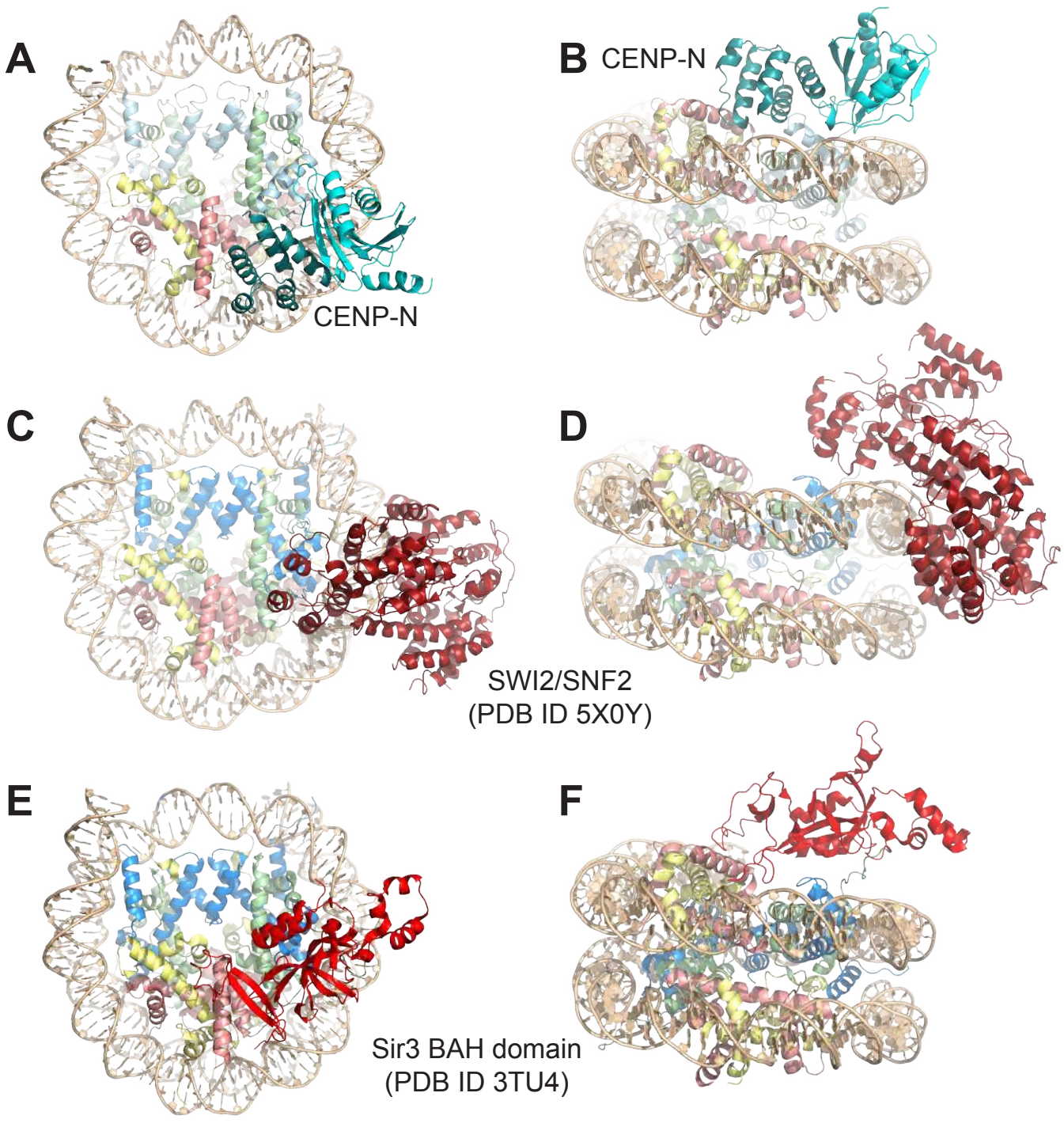
Pentakota, Zhou et al.
Figure 3 Supplement 3



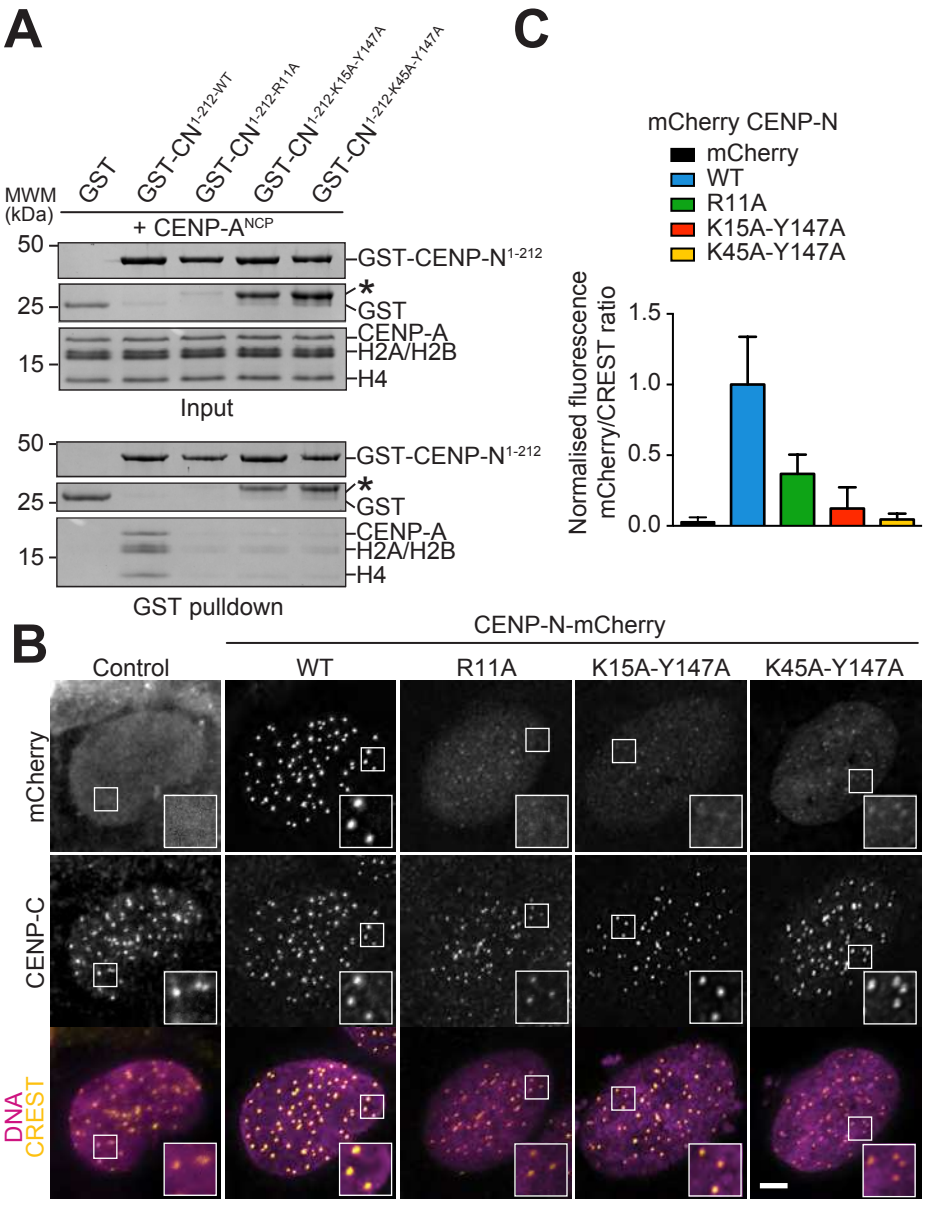
Pentakota, Zhou et al.
Figure 3 Supplement 4



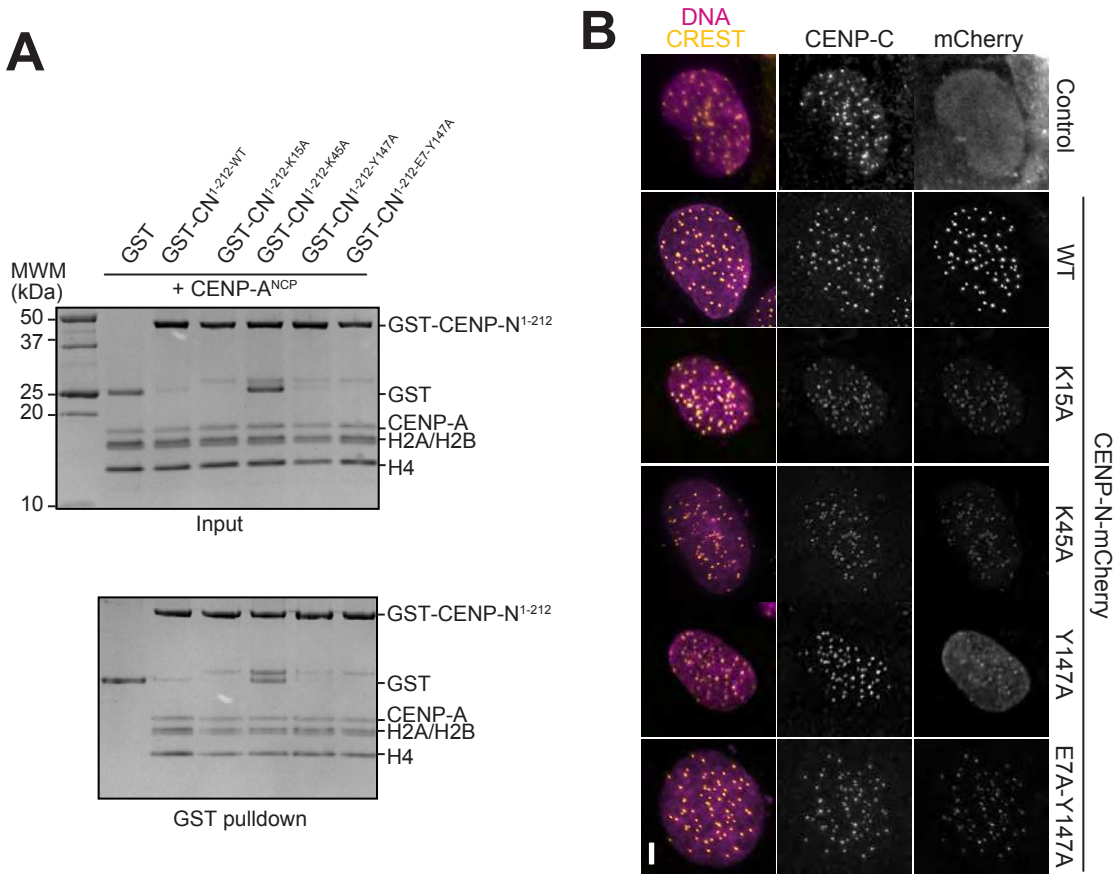
Pentakota, Zhou et al.
Figure 3 Supplement 5



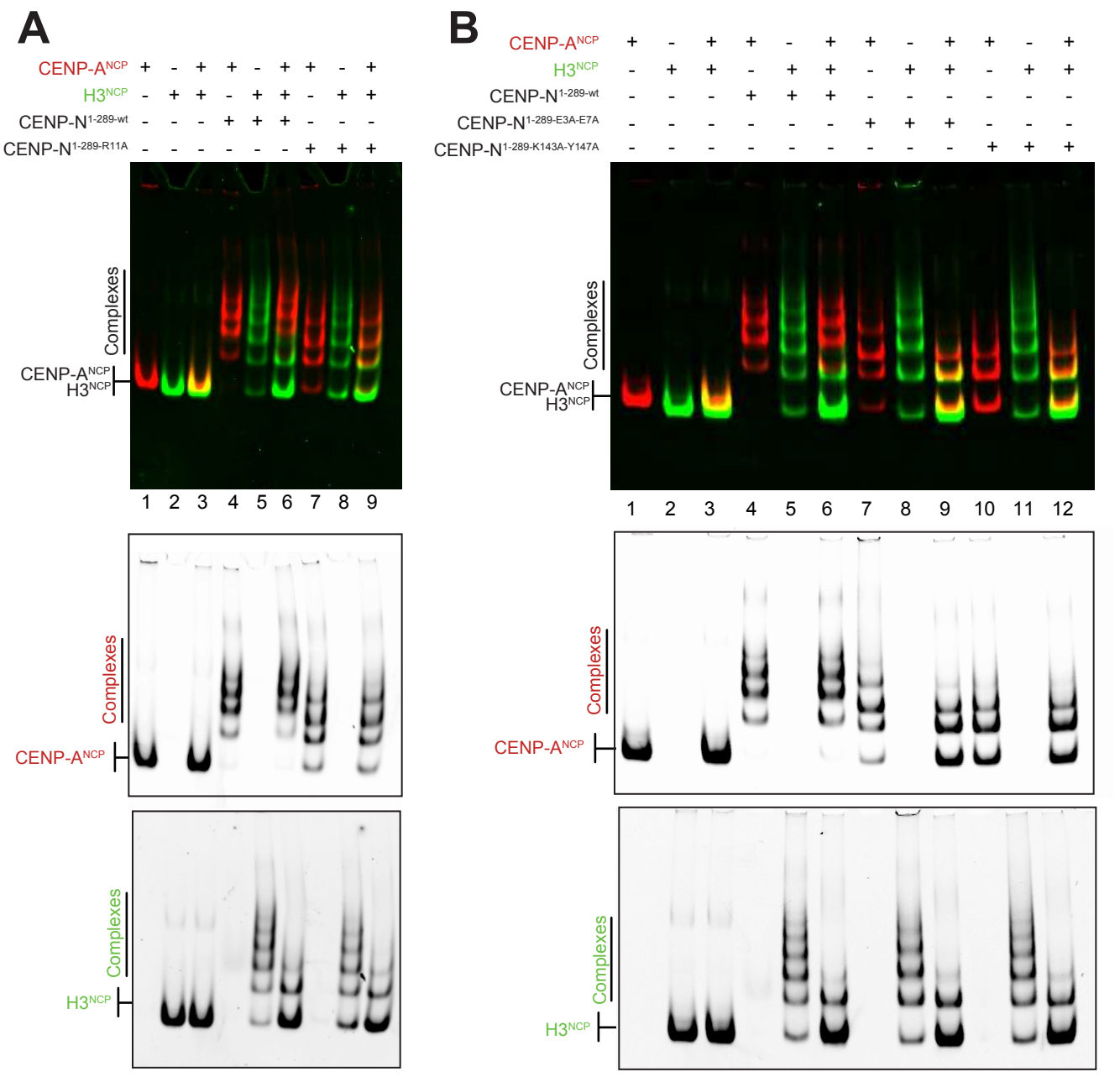
Pentakota, Zhou et al.
Figure 3 Supplement 6



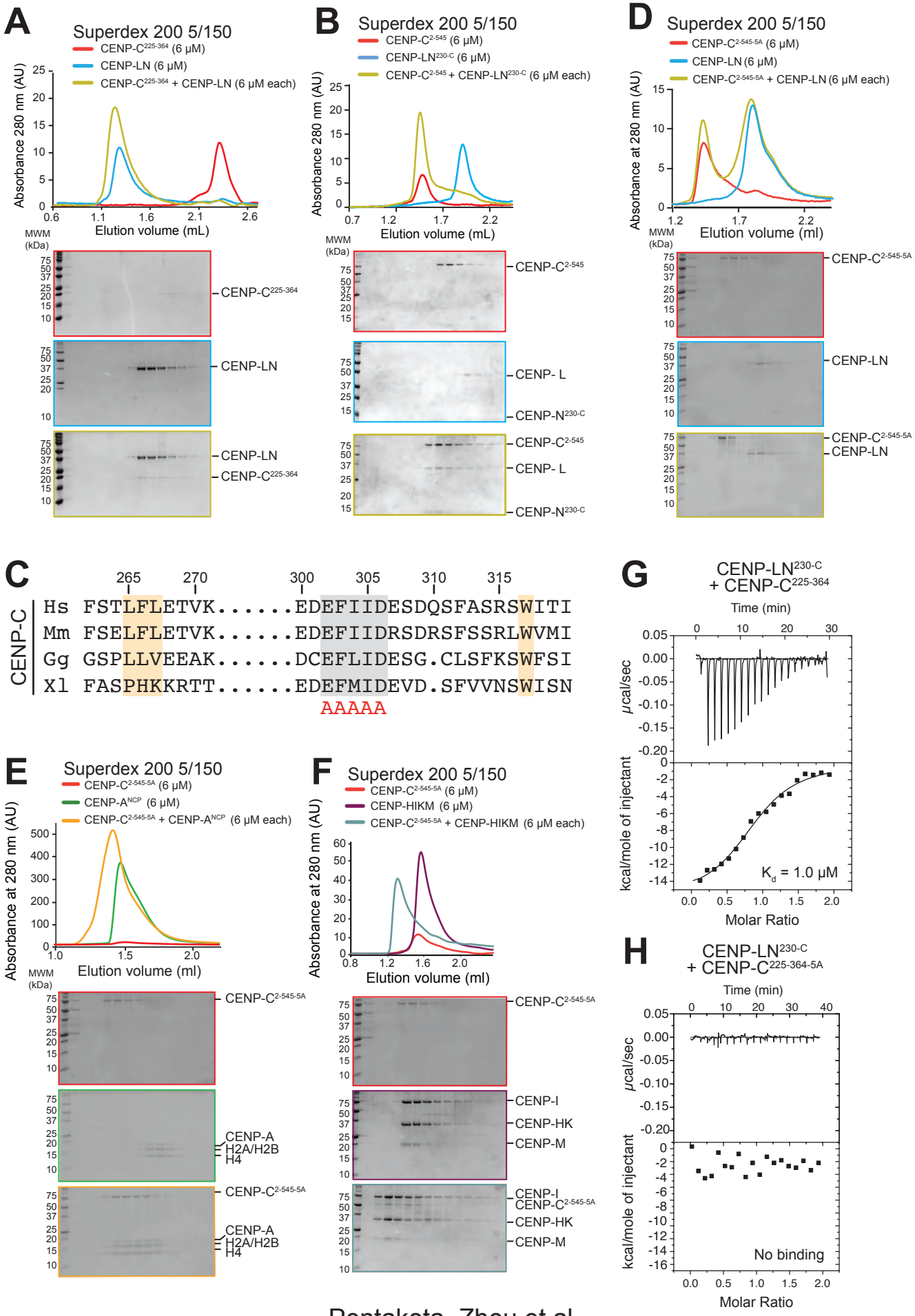
Pentakota, Zhou et al.
Figure 4



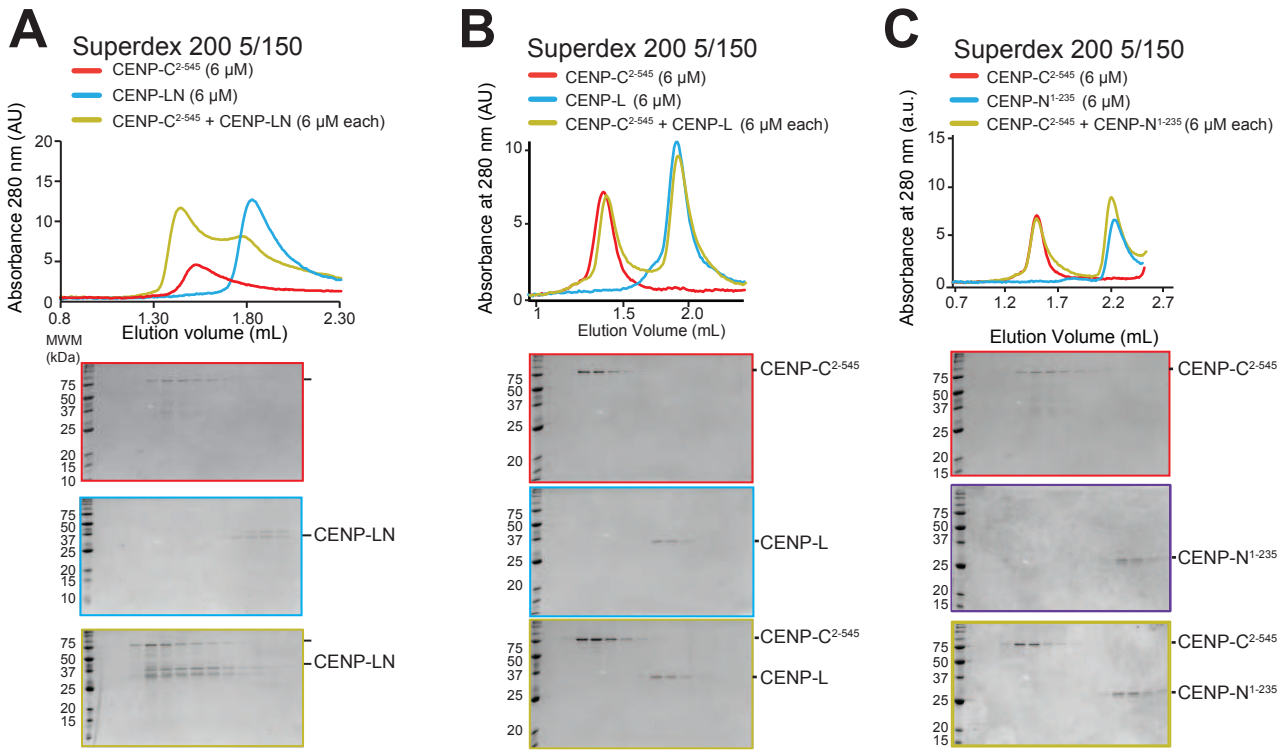
Pentakota, Zhou et al.
Figure 4 Supplement 1



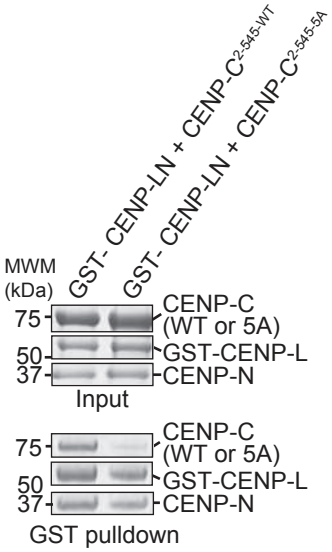
Pentakota, Zhou et al.
Figure 4 Supplement 2



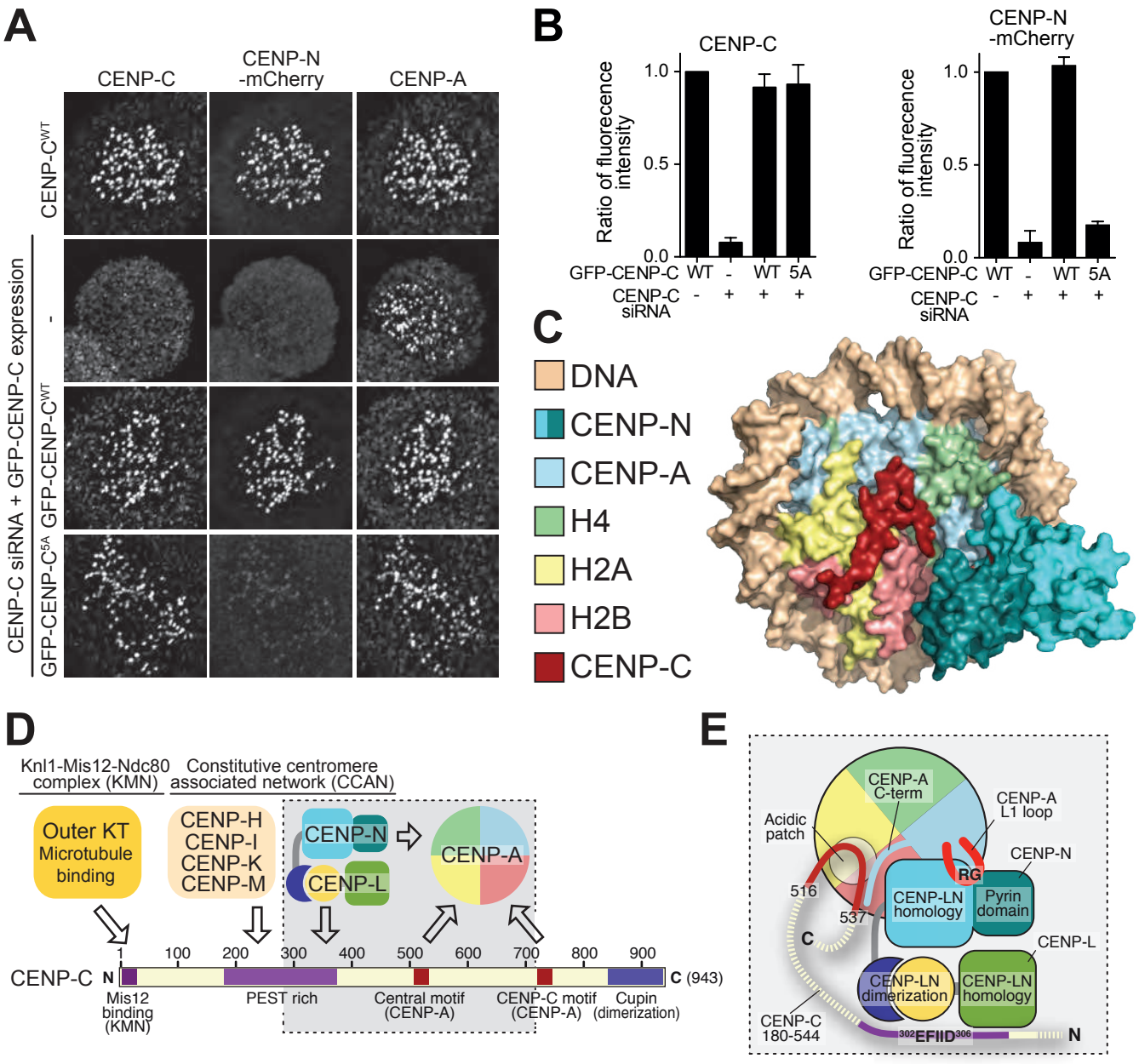
Pentakota, Zhou et al.
Figure 5



Pentakota, Zhou et al.
Figure 5 Supplement 1



Pentakota, Zhou et al.
Figure 5 Supplement 2



Pentakota, Zhou et al.
Figure 6

Table 2

EM data collection, processing, and refinement statistics

Data collection and processing	
Voltage (kV)	300
Magnification	290,000x
Defocus (μm, nominal)	-1.0 to -2.5
Pixel size (Å)	1.02
Electron dose rate (counts/pixel/s)	10
Total electron dose (e^- /Å^2)	80
Exposure time (s)	8
Number of images (collected/processed)	3900/3024
Number of frames per image	40
Initial particle number	1,843,269
Particle number for 3D classification	1,267,674
Final particle for refinement	937,118
Resolution (masked/unmasked) (Å)	4.0/4.2
Map sharpened b-factor (Å^2)	-233
Model refinement	
r.m.s. deviation (bonds)	0.005
r.m.s. deviation (angles)	0.97
All-atom clashscore	2.30
Ramachandran plot	
Outliers (%)	0.00
Allowed (%)	4.59
Favored (%)	95.81
CaBLAM analysis:	
Outliers (%)	1.92
Disfavored (%)	6.65
Ca outliers (%)	0.11
Rotamer outliers (%)	0.00

Table 1

X-ray data collection and refinement statistics

Data collection and processing				
	Native	**SeMet 1**	**SeMet 2**	**SeMet 1+2**
Space group	$P4_1$	$P4_1$	$P4_1$	$P4_1$
Wavelength	0.97793	0.9793	0.9793	0.9793
No. xtals	1	1	1	2
Source	SLS	PETRA	PETRA	PETRA
Detector	Pilatus 6M	Pilatus6M	Pilatus 6M	Pilatus 6M
Mol/AU	2	2	2	2
a,b,c (Å)	87.3 87.3 81.1	88.99 88.99 76.96	89.14 89.14 77.22	88.99 88.99 76.96
a, b, g (°)	90 90 90	90 90 90	90 90 90	90 90 90
Resolution (Å)	87.3-2.74 (2.81-2.74)*	48.7-3.3 (3.9-3.3)	48.8-3.2 (3.3-3.2)	48.7-3.3 (3.4-3.3)
R_{meas}	8.2 (155.1)	17.2 (153.4)	18.8 (173.4)	18.7(167.8)
$I/\sigma I$	17.3 (1.4)	7.5 (1.1)	7.2 (1.0)	10.4 (1.4)
Completeness (%)	99.8 (98.5)	100.0 (100.0)	99.9 (98.8)	100.0(100.0)
Redundancy	9.4 (8.7)	7.1 (7.2)	7.0 (6.3)	14.1 (14.1)
Refinement				**Phasing**
Resolution (Å)	87.3-2.7			FOM 0.39
No. reflections	17103			BAYES-CC 38.1
R_{work} / R_{free}(%)	21.6/26.1			12 Selenium-sites
No. atoms:				
Protein/ Ligands	3432/6			
Water	10			
aver. B (Å^2)	90.4			
R.m.s. deviations				
Bond lengths (Å)	0.0076	Ramachandran plot: 98.0 % favourable, 0 % outliers		
Bond angles (°)	1.27			

***** Values in parentheses are for highest resolution shell