A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C

Hidenori Kato[1], **Jiansheng Jiang**[2], **Bing-Rui Zhou**[1], **Marieke Rozendaal**[3], **Hanqiao Feng**[1], **Rodolfo Ghirlando**[4], **T. Sam Xiao**[2], **Aaron F. Straight**[3], and **Yawen Bai**[1,*]

[1]Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, NIH, Bethesda, MD 20892, USA

[2]Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892, USA

[3]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA

[4]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD 20892

Abstract

Chromosome segregation during mitosis requires assembly of the kinetochore complex at the centromere. Key to kinetochore assembly is the specific recognition of the histone variant CENP-A in the centromeric nucleosome by centromere protein C (CENP-C). We have defined the determinants of this recognition mechanism and discovered that CENP-C binds a hydrophobic region in the CENP-A tail and docks onto the acidic patch of histone H2A/H2B. We further find that the more broadly conserved CENP-C motif uses the same mechanism for CENP-A nucleosome recognition. Our findings reveal a conserved mechanism for protein recruitment to centromeres and a histone recognition mode whereby a disordered peptide binds the histone tail through nucleosome-docking-facilitated hydrophobic interactions.

The accurate segregation of chromosomes during mitosis is essential for reproduction and development of all organisms. Chromosome segregation is facilitated by the attachment of mitotic spindle microtubules to the kinetochore, which is assembled on the chromosomal centromere (1). Human CENP-A chromatin is constitutively associated with a complex of sixteen centromere proteins (2, 3). Among these sixteen proteins, only CENP-C has been identified in all model organisms (4). CENP-C recognizes the carboxyl tail of CENP-A in the centromeric nucleosome (5, 6). Human CENP-C consists of four functional regions (fig. S1A). The N-terminal region interacts with the Mis12 complex (7). The central region and the CENP-C motif are required for targeting CENP-C to the centromere (5, 8-11). The central region directly binds to the CENP-A nucleosome while the targeting mechanism for the CENP-C motif is unknown. The C-terminal region is responsible for CENP-C dimerization (12).

To explore how CENP-C recognizes the CENP-A nucleosome, we first investigated the central region (CENP-C$_{426-537}$) alone by nuclear magnetic resonance spectroscopy (NMR) and found that it is disordered (fig. S1B, C). We then examined CENP-C$_{426-537}$ in complex with reconstituted *Drosophila* nucleosomes containing the H3$_{1-132}$-LEEGLG chimera. The residues in the CENP-A nucleosome that are exposed for binding of CENP-C are essentially conserved in the *Drosophila* H3 nucleosome (5, 13) (fig. S2A, B), and CENP-C$_{426-537}$ binds

[*]Correspondence should be addressed to: yawen@helix.nih.gov.

to the CENP-A and *Drosophila* H3$_{1-132}$-LEEGLG nucleosomes with the same stoichiometry and similar affinity (figs. S2C-E, S3). We found that the residues 426-481 of CENP-C$_{426-537}$ remain disordered while residues 482-537 fold into the chimeric nucleosome core (fig. S1B).

Using methyl-based chemical shift perturbation and paramagnetic relaxation enhancement (14, 15), we mapped the binding sites between CENP-C$_{426-537}$ and the chimeric nucleosome. Upon binding of CENP-C$_{426-537}$, the cross-peaks of methyl groups of residues Leu$_{62}$ in H2A, Leu$_{103}$ in H2B, Leu$_{100}$, Leu$_{103}$, Ile$_{124}$, Ile$_{126}$, and Ile$_{130}$ in H3, and Ile$_{50}$ and Leu$_{58}$ in H4 displayed large chemical shift changes, while those of Leu$_{64}$ in H2A and Val$_{57}$, Val$_{60}$, Ile$_{62}$, and Val$_{65}$ in H4 disappeared (Fig. 1A, C and fig. S4), suggesting close contact of these residues with CENP-C$_{426-537}$. We next individually mutated each of the residues Phe$_{500}$, Val$_{509}$, Val$_{517}$, Ile$_{523}$, and Ser$_{535}$ in CENP-C$_{484-537}$ to Cys and linked them to a paramagnetic spin label. Binding of paramagnetic spin-labeled CENP-C$_{484-537}$ reduced the peak intensities of the methyl groups in the nucleosome in accordance with their distance from the paramagnetic center. While spin labels of residues Val$_{509}$ and Val$_{517}$ mainly affected the methyl groups near the C-terminal region of the histone fold of H2B, the spin label of residue Ile$_{523}$ affected methyl groups at an acidic residue-rich region (acidic patch) (Fig. 1B, C and fig. S5). In contrast, the spin label of residue Ser$_{535}$ affected the methyl groups of H3$_{1-132}$-LEEGLG near the C-terminal tail (Fig. 1B, C). These results indicate that CENP-C residues 509-535 bind to the histone surface, using hydrophobic residues $_{530}$WWVV$_{533}$ to recognize the LEEGLG tail and docking the positively charged region (Arg$_{521}$, Arg$_{522}$, Arg$_{525}$ and Arg$_{526}$) on the acidic patch region (fig. S5E, S6, and SOM text). The CENP-C residues 482-508, which include many positively charged residues, likely bind to DNA (fig. S6). The CENP-C binding region in the core histones of the *Drosophila* H3 nucleosome has the same structure as the corresponding region of the human CENP-A nucleosome (16) (fig. S7), further validating the use of the chimeric nucleosome.

Deletion and isothermal titration calorimetry studies showed that CENP-C 520-537 region contains residues critical for binding to the chimeric nucleosome (table S1). Ala-scanning revealed that substitution of any one of the Arg$_{522}$, Trp$_{530}$, or Trp$_{531}$ residues with Ala in CENP-C$_{444-537}$ abolished binding (Fig. 2A, fig. S8A and table S2). Conversely, the single charge-reversal mutations Glu60Lys, Glu63Lys, and the neutralizing mutation Glu90Thr in H2A each essentially abolished the binding (fig. S8A and table S2). Individual substitution of the two Leu residues in the C-terminus of H3$_{1-132}$-LEEGLG with Ala reduced the dissociation equilibrium constant (K_d) by a factor of ~3 and ~4, respectively (fig. S8A and table S2). In contrast, Ala substitution of the two Glu residues increased the K_d by a factor of ~3 and ~10 (table S2), respectively, indicating that the CENP-A C-terminal tails have not evolved for high affinity binding with CENP-C. Mutation of CENP-C central region residues Arg$_{522}$, Trp$_{530}$ or Trp$_{531}$ to Ala also substantially decreased the localization of GFP-fused CENP-C$_{1-537}$ to centromeres in human cells, whereas the S520A or S524A mutation had smaller effects (Fig. 2B and fig. S8B). Mutation of the two acidic patch residues Glu$_{61}$ and Glu$_{64}$ residues in H2A to either Ala or positively charged Lys inhibited the recruitment of CENP-C from the *X. laevis* egg extract to the reconstituted CENP-A nucleosome array (Fig. 2C and fig. S8C) (6). Neutralization of the acidic patch by the mutations did not disrupt the formation of condensed CENP-A nucleosome arrays (fig. S9).

The central region of CENP-C is conserved in most mammals, whereas the CENP-C motif is conserved from budding yeast to human (fig. S10A). Residues Arg$_{522}$, Trp$_{530}$, and Trp$_{531}$, which are most important for the binding of the CENP-C central region to the chimeric nucleosome, are essentially conserved in the CENP-C motif and correspond to residues Arg$_{742}$, Tyr$_{750}$ and Trp$_{751}$, suggesting that the human CENP-C motif may also bind to

CENP-A and H3$_{1-132}$-LEEGLG nucleosomes. Indeed, chemical shift perturbation studies indicate that the CENP-C motif binds to the similar sites on the core histones in the H3$_{1-132}$-LEEGLG nucleosome as the CENP-C central region (Fig. 3A and fig. S11). Furthermore, binding of the CENP-C motifs from human, rat, fission yeast, and budding yeast to the H3 chimeric nucleosomes containing the corresponding C-terminal residues of the CENP-A homolog displayed reasonably strong affinities with K$_d$ of 1.4 µM, 0.1 µM, 1.4 µM, and 0.3 µM, respectively (fig. S12 and table S3). Charge reversal mutation of a single acidic residue in the acidic patch (Glu60Lys or Glu63Lys) abolished the binding of the CENP-C motif (table S3). Furthermore, mutation of the conserved sites within the rat CENP-C motif (Arg717Ala, Tyr725Ala or Trp726Ala) substantially decreased the binding affinity (fig. S12 and table S4). Mutation of the residues corresponding to Arg$_{717}$ in frog, fly and fission yeast prevented targeting of CENP-C to the centromere (9-11).

We determined the crystal structure of the rat CENP-C motif in complex with the corresponding H3$_{1-132}$-IEGGLG nucleosome at 3.5 Å resolution (Fig. 3B, fig. S13 and table S5). In the structure, the conserved CENP-C motif residue Tyr$_{725}$ recognizes the residues Ile$_{133}$ and Leu$_{137}$ in the IEGGLG tail through hydrophobic interactions, inducing the IEGGLG tail to fold to a turn structure and promoting intramolecular hydrophobic interactions between Ile$_{133}$ and Leu$_{137}$. CENP-C residue Trp$_{726}$ forms intra-molecular hydrophobic interactions with CENP-C residue Tyr$_{725}$, in addition to interacting with the hydrophobic region of the side chain of residue Arg$_{129}$ in H3$_{1-132}$. CENP-C residues Arg$_{717}$ and Arg$_{719}$ form electrostatic interactions with the acidic patch residues Glu$_{60}$, Glu$_{63}$, Asp$_{89}$, and Glu$_{91}$ of H2A and Glu$_{102}$ of H2B. Binding to the acidic patch was similarly found in structures of LANA (17), RCC1 (18), and the Sir3 BAH domain (19) in complex with the nucleosome.

Whereas CENP-C motifs are well conserved among all species, the C-terminal tails of CENP-As are not (fig. S10B). We observed a common feature in the C-terminal tails of CENP-As: they are more hydrophobic than those of corresponding H3. In particular, the C-terminal tails of CENP-As have at least one bulky hydrophobic residue such as Ile, Leu, Val or aromatic residues (with the exception of fission yeast), which are absent in the C-terminal tails of H3. In addition, nucleosomes containing chimeric H3$_{1-132}$-QFI (C-terminal tail of CENP-A in budding yeast) bound to the central region and the CENP-C motif of human CENP-C with similar affinity as nucleosomes containing H3$_{1-132}$-LEEGLG (Fig. 4 and table S3). Thus, the higher hydrophobicity of the C-terminal tail of CENP-A rather than a specific amino acid sequence is the key determinant for specific recognition by CENP-C.

Our findings can explain diverse experimental results in the literature. In human cells, the CENP-C$_{584-943}$ fragment, which contains the CENP-C motif and the dimerization domain, is capable of being targeted to the centromere when over-expressed (20). In *Xenopus*, chromatin in which H2A/H2B is replaced by protamines does not support CENP-C localization to the centromere of sperm (9, 21). However, CENP-C co-localizes with CENP-A upon addition of egg extract, as the histone chaperone nucleoplasmin catalyzes the eviction of protamines and the incorporation of H2A/H2B dimers into the sperm chromatin (21). The function of human CENP-A can be complemented by the budding yeast *Saccharomyces cerevisiae* homolog Cse4 (22) (Fig. 4). Our study also suggests that CENP-C can recognize CENP-A chromatin through multivalent interactions, allowing it to associate more strongly and selectively (figs. S14, S15) (20, 23). In addition, our results have implications for various structural models of the centromeric nucleosome, a widely discussed and controversial topic (24). The finding that the broadly conserved CENP-C motif in a dimeric CENP-C could bind to two H2A/H2B and two CENP-A/H4 molecules constraints these models since CENP-C is constitutively associated with the centromeric nucleosome. Our study provides the structural basis for the recognition of the CENP-A

octameric nucleosome by CENP-C. Further understanding of how it engages chromatin should provide important insight into favored compositions of CENP-A nucleosomes.

In a broader perspective, our finding that the widely conserved CENP-C motif binds to the centromeric nucleosome, along with the recent identification of the budding yeast homologues of the vertebrate centromere proteins CENP-T and CENP-W (25, 26), supports a conserved mechanism of centromere targeting by the kinetochore. Furthermore, our study has revealed a histone recognition mode whereby an intrinsically disordered peptide binds to the histone tail through nucleosome-docking-facilitated hydrophobic interactions, which broadens the repertoire of the cell to "read" histone variations (27, 28), not only at the centromere but possibly also in other contexts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References and notes

1. Henikoff S, Ahmad K, Malik HS. Science. 2001; 293:1098. [PubMed: 11498581]

2. Verdaasdonk JS, Bloom K. Nat. Rev. Mol. Cell Biol. 2011; 12:320. [PubMed: 21508988]

3. Hori T, et al. Cell. 2008; 135:1039. [PubMed: 19070575]

4. Przewloka MR, et al. Curr. Biol. 2011; 21:399. [PubMed: 21353555]

5. Carroll CW, Milks KJ, Straight AF, Cell Biol J. 2010; 189:1143.

6. Guse A, Carroll CW, Moree B, Fuller CJ, Straight AF. Nature. 2011; 477:354. [PubMed: 21874020]

7. Screpanti E, et al. Curr. Biol. 2011; 21:391. [PubMed: 21353556]

8. Song K, Gronemeyer B, Lu W, Eugster E, Tomkiel JE. Exp. Cell Res. 2002; 275:81. [PubMed: 11925107]

9. Milks KJ, Moree B, Straight AF. Mol. Biol. Cell. 2009; 20:4246. [PubMed: 19641019]

10. Tanaka K, Chang HL, Kagami A, Watanabe Y. Dev. Cell. 2009; 17:334. [PubMed: 19758558]

11. Heeger S, et al. Genes Dev. 2005; 19:2041. [PubMed: 16140985]

12. Cohen RL, et al. Mol. Biol. Cell. 2008; 19:4480. [PubMed: 18701705]

13. Carroll CW, Silva MC, Godek KM, Jansen LE, Straight AF. Nat. Cell Biol. 2009; 11:896. [PubMed: 19543270]

14. Tugarinov V, Kanelis V, Kay LE. Nat. Protoc. 2006; 1:749. [PubMed: 17406304]

15. Kato H, et al. Proc. Natl. Acad. Sci. U.S.A. 2011; 108:12283. [PubMed: 21730181]

16. Tachiwana H, et al. Nature. 2011; 476:232. [PubMed: 21743476]

17. Barbera AJ, et al. Science. 2006; 311:856. [PubMed: 16469929]

18. Makde RD, England JR, Yennawar HP, Tan S. Nature. 2010; 467:562. [PubMed: 20739938]

19. Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE. Science. 2011; 334:977. [PubMed: 22096199]

20. Sugimoto K, Kuriyama K, Shibata A, Himeno M. Chromosome Res. 1997; 5:132. [PubMed: 9146917]

21. Philpott A, Leno GH. Cell. 1992; 69:759. [PubMed: 1591776]

22. Wieland G, Orthaus S, Ohndorf S, Diekmann S, Hemmerich P. Mol. Cell. Biol. 2004; 24:6620. [PubMed: 15254229]

23. Ando S, Yang H, Nozaki N, Okazaki T, Yoda K. Mol. Cell. Biol. 2002; 22:2229. [PubMed: 11884609]

24. Black BE, Cleveland DW. Cell. 2011; 144:471. [PubMed: 21335232]

25. Bock LJ, et al. Nat. Cell Biol. 2012; 14:614. [PubMed: 22561345]

26. Schleiffer A, et al. Nat. Cell Biol. 2012; 14:604. [PubMed: 22561346]

27. Khorasanizadeh S. Curr. Opin. Struct. Biol. 2011; 21:744. [PubMed: 22056613]

28. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. Nat. Struct. Mol. Biol. 2012; 19:1218. [PubMed: 23211769]
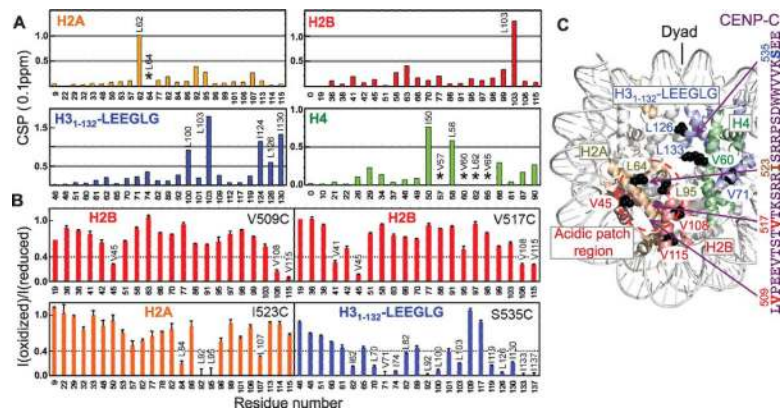
Figure 1.
Binding sites between the CENP-C central region on the chimeric nucleosome mapped by methyl-based NMR. (**A**) Histone methyl chemical shift perturbation (CSP) upon CENP-$C_{426\text{-}537}$ binding. (**B**) Effects on methyl groups by paramagnetic spin labels in CENP-$C_{484\text{-}537}$. (**C**) Nucleosome structure showing representative side chains (black balls) whose methyl groups display large changes in (**A**) or (**B**), and approximate locations of CENP-C residues (arrows). E133 in H3 is used to represent L133 in $H3_{1\text{-}132}$-LEEGLG.
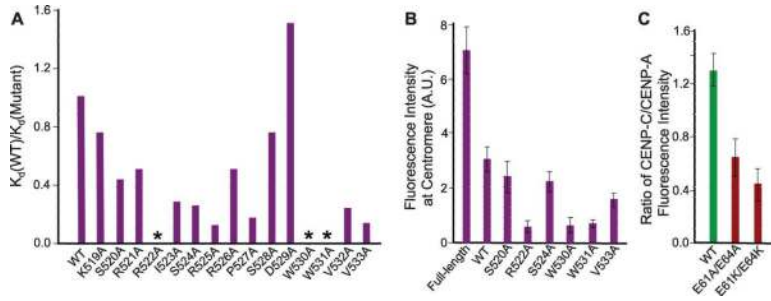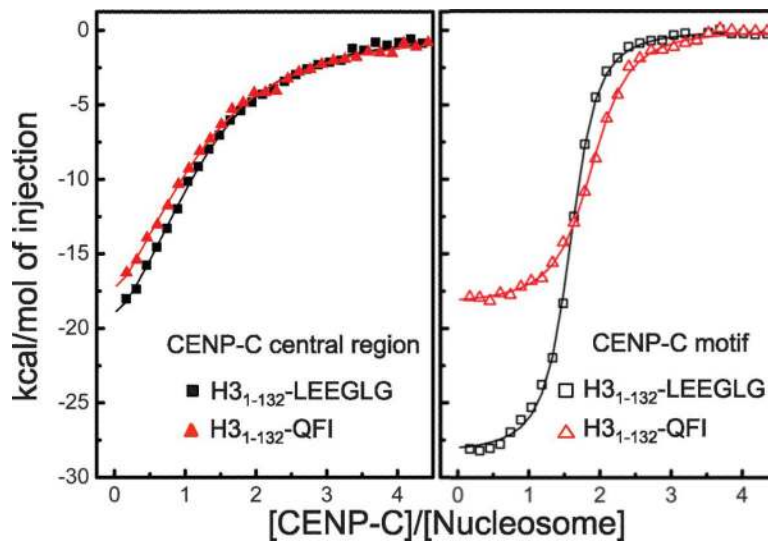
Figure 2.
Residues important for binding of the central region of CENP-C and the chimeric nucleosome. (**A**) Effects of mutations in CENP-C$_{426-537}$ on its binding to the H3$_{1-132}$-LEEGLG nucleosome. No detectable binding (*). (**B**) Effect of CENP-C$_{1-537}$ mutations on centromere targeting in human cells (fig. S8B). Error represents SEM n=4, Students t probability p ≤0.02 for all mutants. (**C**) Effect of H2A acidic patch mutations on localization of CENP-C to the CENP-A nucleosome arrays. Error represents SEM, n=4, Students t probability p ≤0.05 for both mutants.
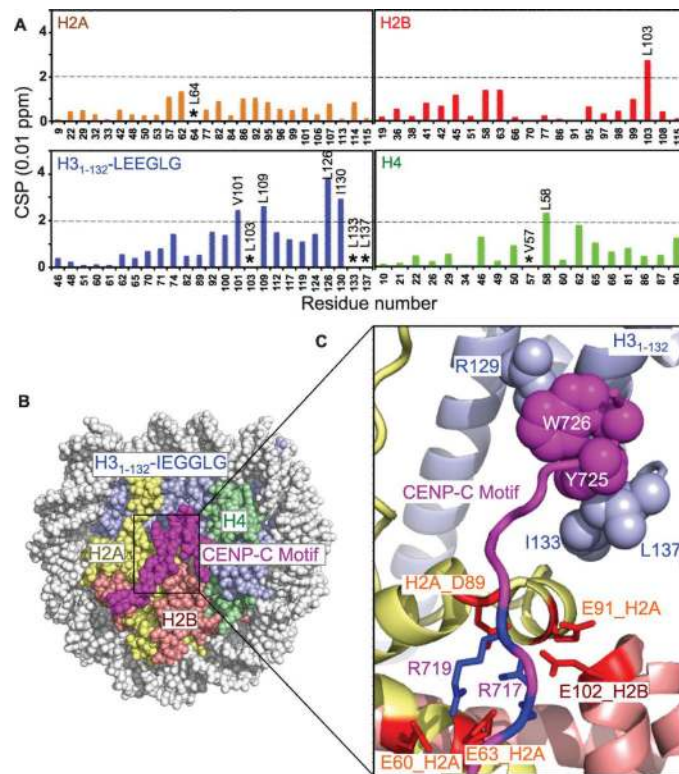
Figure 3.
Broadly conserved CENP-C motifs bind to the H3 nucleosome chimera containing CenH3 C-terminal tails. (**A**) Nucleosome methyl CSP upon binding of the CENP-C motif. (**B**) The structure of the rat CENP-C motif in complex with the $H3_{1-132}$-IEGGLG nucleosome. (**C**) Enlarged region showing hydrophobic and electrostatic interactions.

Figure 4.
Hydrophobicity of the CenH3 tail is the primary determinant for recognition of CenH3 by CENP-C. Isothermal titration calorimetric curves for binding of the human CENP-C central region (CENP-C$_{444-537}$) and the CENP-C motif (CENP-C$_{727-767}$) to nucleosomes containing H3$_{1-132}$-LEEGLG (human) and H3$_{1-132}$-QFI (budding yeast), respectively.