Biophysical Journal, Volume *118*

# Supplemental Information

# Minimal Cylinder Analysis Reveals the Mechanical Properties of Oncogenic Nucleosomes

**Mary Pitman, Yamini Dalal, and Garegin A. Papoian**

# Supporting Material

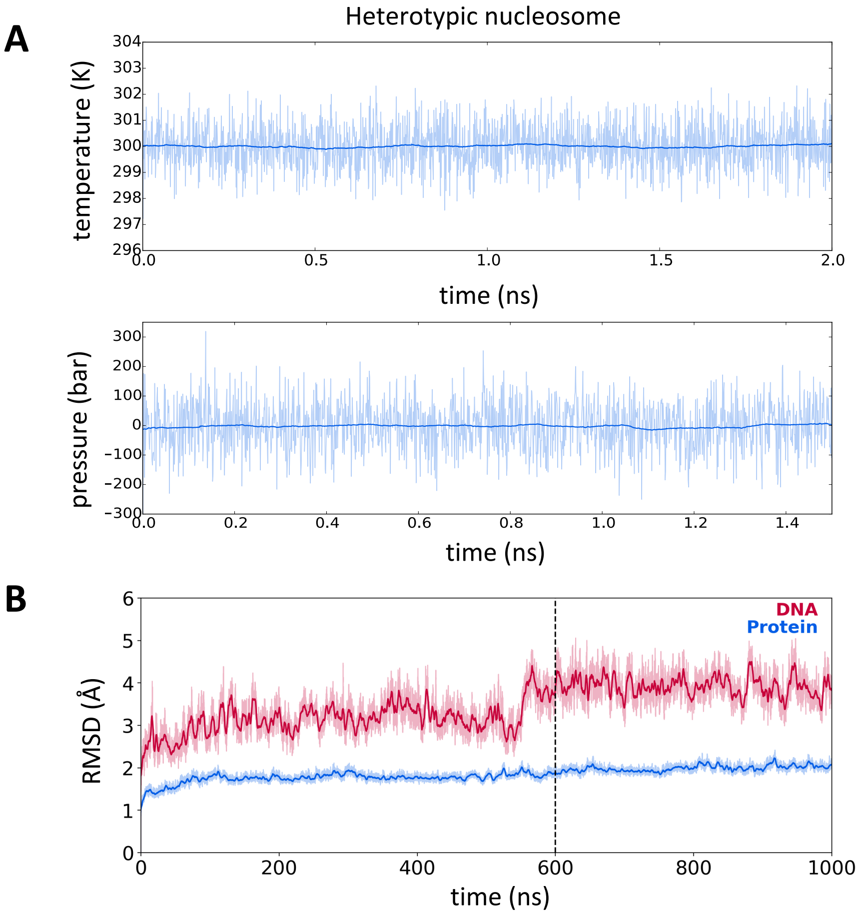

**Figure S1:** (A) The second thermal equilibration to 300 K with weak position restrains, $K_{het}$, of the Heterotypic nucleosome is shown, top. Next, the system pressure was equilibrated to 1 bar, shown on the bottom. (B) The RMSD of the Heterotypic DNA and protein compared to the initial production run configuration. We performed analysis from 600 to 1000 ns.

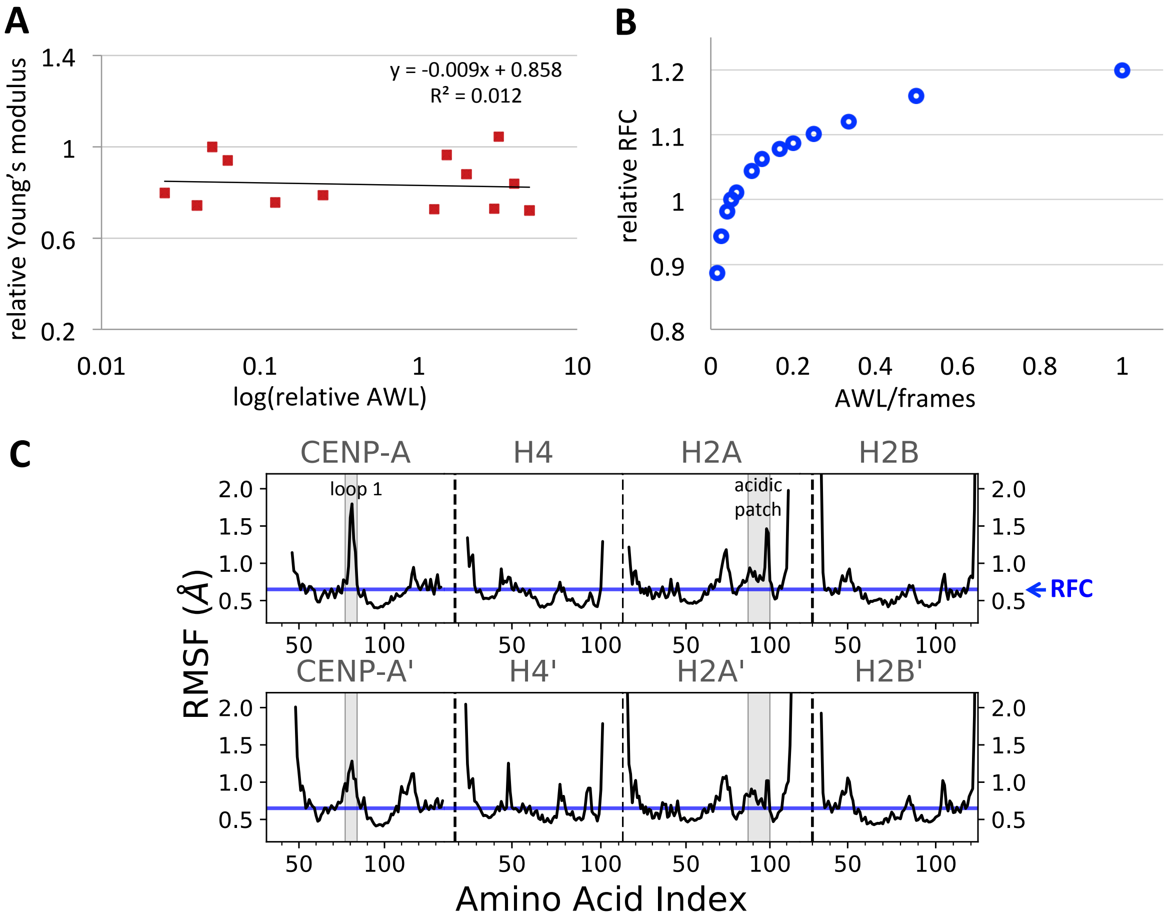

**Figure S2:** (A) The log of the relative AWL input to the relative output, Young's modulus. The parameter AWL specifies the length of trajectory divisions for MCA. AWL values are restricted to factors of the analyzed trajectory frame number, and so the log is shown due to clustering of factors at low numbers. (B) The ratio of the AWL parameter to the number of analyzed frames, where 1 equals the whole trajectory, is shown on the x-axis. The y-axis shows the 50th percentile RMSF value calculated during MCA with varied AWL input. The 'Residue Flexibility Cutoff' (RFC) specifies the maximum RMSF value counted as rigid and is calculated during MCA as the 50th percentile value of the RMSF dataset. (C) Example dataset of CENP-A C-$\alpha$ RMSF data at AWL/frames equal to 1 with the RFC value shown as a blue line. Below the blue line residues are counted as rigid.

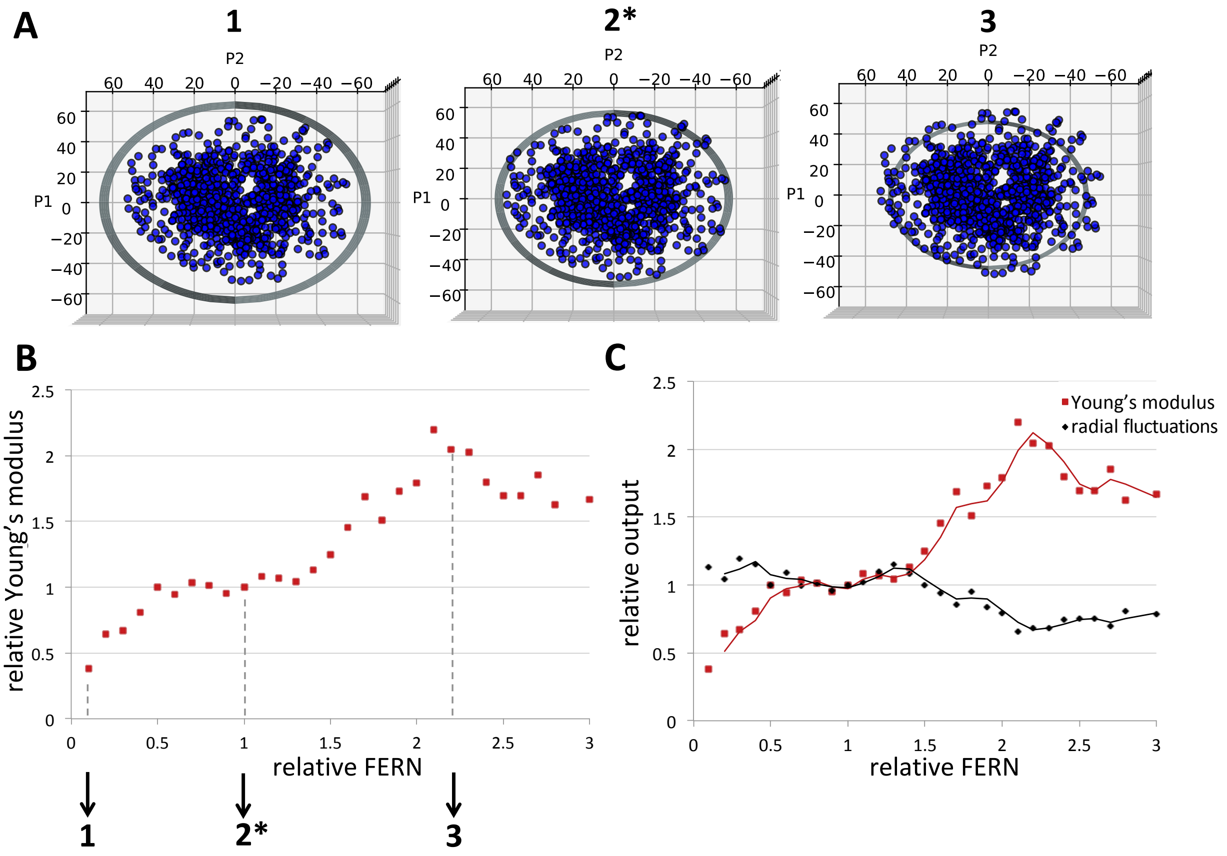

**Figure S3:** (A) The NCP C-α and phosphorus atoms are shown in blue on the plane of the first two principal axes. The cylinder radius is measured from the origin to the grey circular boundary shown. Graph 1 shows the nucleosome and the calculated cylinder dimensions for a FERN value of 1, graph 2* for a FERN value of 10, and graph 3 for a value of 22. (B) The relative Young's modulus output dependent on the relative FERN value is shown. The data points graphically shown in panel A are labeled. Point 2* was used for analysis. (C) The relative Young's modulus (red) and the relative radial fluctuation, Δr (black), dependent on the relative FERN value.

| System | z (Å) | Δz (Å) | r (Å) | Δr (Å) |
|---|---|---|---|---|
| H3 | 55.20 | 0.62 | 54.50 | 0.29 |
| CENP-A | 57.26 | 0.70 | 53.66 | 0.41 |
| +CENP-C | 64.41 | 0.72 | 53.36 | 0.31 |
| Heterotypic | 63.14 | 0.61 | 53.36 | 0.35 |

**Table S1:** Sample trial MCA data for the average dimensions of the height and radius, $z_{avg}$ and $r_{avg}$, and the standard deviation of the height and radius, Δz and Δr. The flexibility of the structure depends on a nonlinear combination of the ratio of the fluctuation to the dimension size of both r and z.

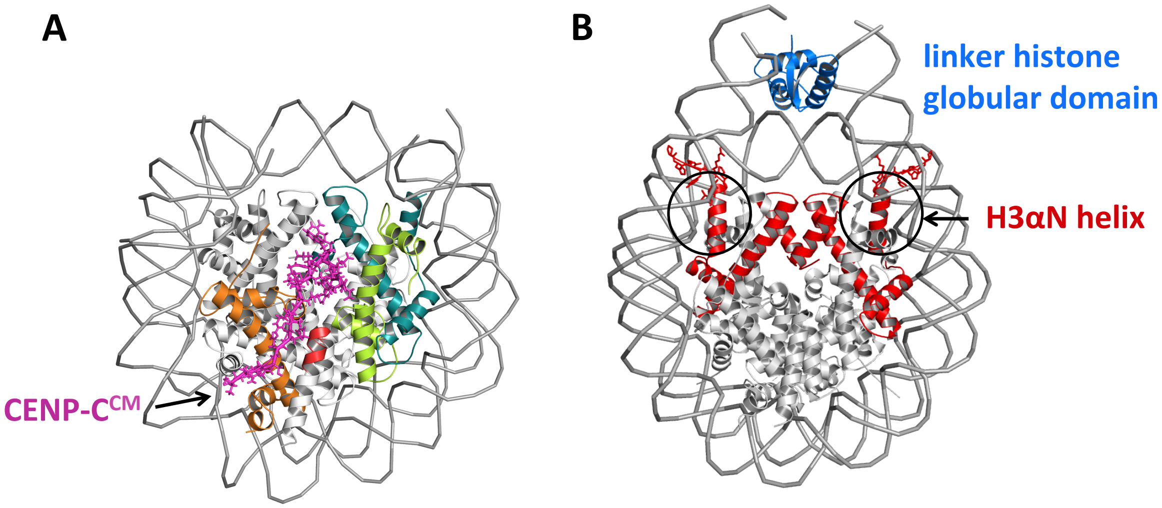

**Figure S4:** (A) Structural reference of the central motif (CM) binding domain of CENP-C, CENP-C$^{CM}$, bound to the CENP-A NCP. Depicted is the first frame of the CENP-A + CENP-C production run of our prior work (1). CENP-C is shown in magenta, CENP-A in teal, H4 in light green, H2A in orange and the fragment of H2B bound to CENP-C is shown in red. (B) Structural reference of the linker histone globular binding domain bound to the canonical nucleosome to form the chromatosome, PDB ID: 4QLC (2). H3 histones are shown in red and the globular domain of the LH is shown in blue.

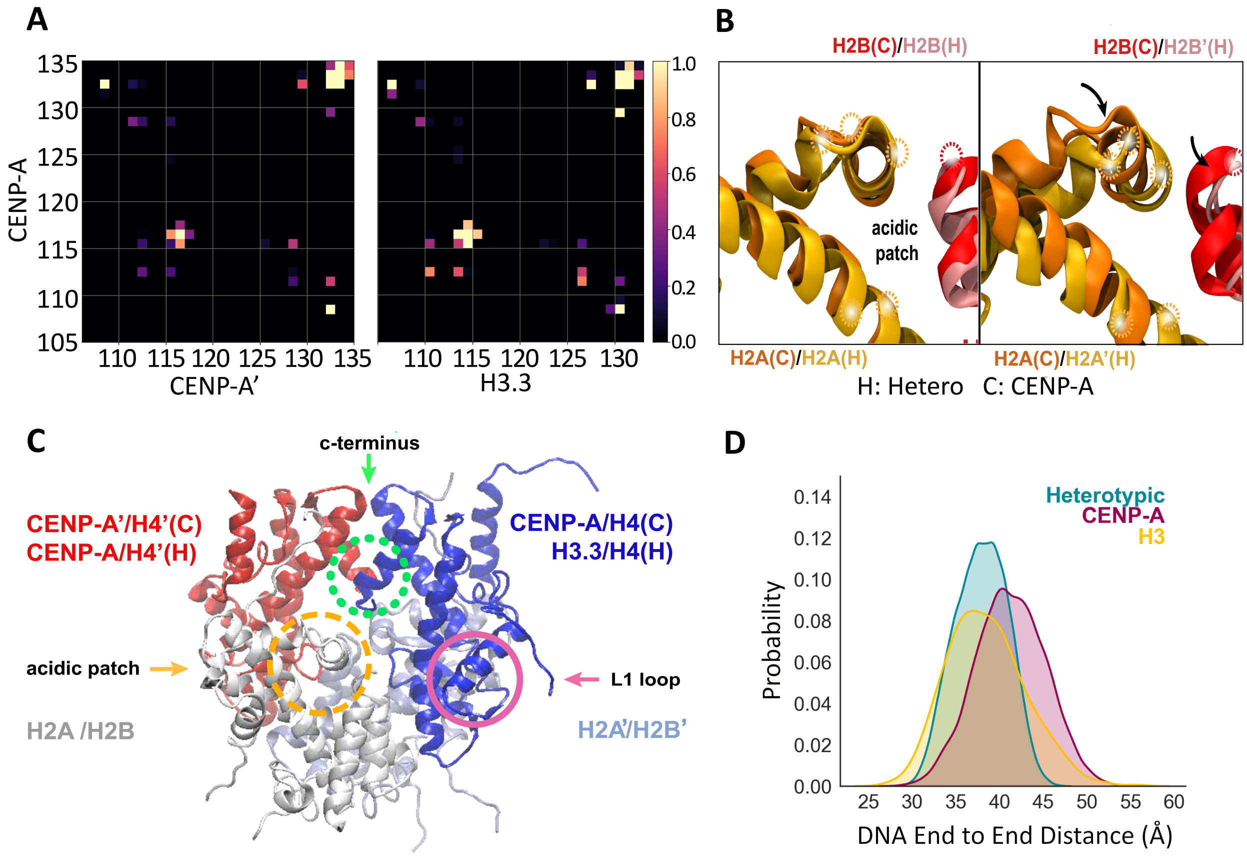

**Figure S5:** (A) Contact analysis at the 4-helix bundle interface of CENP-A:CENP-A in the context of the CENP-A nucleosome on the left in comparison with CENP-A:H3.3 in the heterotypic nucleosome. Increased brightness of color shows the propensity of C-α contact within 8 Å. Black means that contact does not occur, and pale yellow indicates contact at all time-steps. (B) The overlay of the acidic patches from the representative structure of the first principal component. System "C" denotes histones from the CENP-A nucleosome and "H" from the CENP-A:H3.3 heterotypic nucleosome. On the left, the CENP-A region from system C is shown with minimum RMSD alignment to the CENP-A region of the heterotypic nucleosome. For comparison, on the right, the H3.3 domain of H is compared to CENP-A in C. CENP-C binding residues are highlighted. (C) Representative image from Movie S1, which depicts histone core motions of the first principal component. Regions of interest in the PCA movie are highlighted such as the histones (where H is the Heterotypic nucleosome and C is the CENP-A nucleosome), the acidic patch in the dashed orange circle, the L1 loop in the pink circle, and the CENP-A or H3.3 c-terminal region in the dotted green circle. (D) The histogram of DNA end to end distances for the Heterotypic nucleosome in teal, the CENP-A nucleosome in magenta, the H3 nucleosome in yellow.

## S1. *Nucleosome Orientation*

The axes of symmetry of the nucleosomes align to the three principal axes. We confirmed that the axes were orthogonal and performed a rotational transformation of the atomic coordinates so that the principal axes aligned to the Cartesian coordinate system with the center-of-mass at the origin.

## S2. *Parameter Sensitivity*

To calculate the Young's modulus, we rationally selected both input parameters 'Averaging Window Length' (AWL) and the 'Flexible Exterior Residue Number' (FERN). The AWL parameter specifies the time length of trajectory divisions for MCA. For each temporal segment we calculate local flexibility through the time-averaged parameter, residue RMSF. If the RMSF is less than the 'Residue Flexibility Cutoff' (RFC), a residue is counted as rigid. The RFC value is calculated in MCA based on the RMSF dataset and is dependent on AWL (Fig. S2B), described below. The second input parameter, FERN, defines the number of flexible residues cleaved from the exterior surfaces to determine the minimal cylinder dimensions.

We will first address the parameter sensitivity of our algorithm to AWL and its selection. Since AWL divides the trajectory into collections of frames over which RMSF is time averaged, AWL must be large enough to produce RMSF data with a spread. If the value of the AWL is too small, there is unclear separation of rigid and flexible residues. Excessively long sampling intervals correspond with a statistically insignificant number of generated cylinders. We assessed AWL values ranging from 10 to 400 frames while holding the other input parameter, FERN, constant (Fig. S2A). When we fit the Young's modulus output data by linear regression, we found a slightly negative slope (-0.009) but with a low $R^2$ value of 0.012. The standard deviation of the Young's modulus output when varying AWL from 10 to 400 is ± 0.82 MPa. We selected an AWL value of 20 since it falls on the lower end of this range and therefore more cylinders are calculated. The lack of sensitivity to AWL over this range indicates that our trajectories are long enough to be insensitive to being partitioned into longer sampling intervals over this range. Shorter trajectories may be sensitive to AWL over the range of tested AWL values and should be tested based on the trajectory length.

The RFC specifies the maximum RMSF value counted as rigid and is calculated by MCA as the 50th percentile value of the RMSF dataset. We plotted the relative change in the RFC value based on the number of times the trajectory is divided, the AWL input parameter (Fig. S2B). We found a monotonic increase in the RFC value and dependence on the AWL parameter. At point AWL/frames is equal to 1, a singular temporal segment is generated which is equal to the trajectory length and the RFC value is calculated for the whole trajectory (Fig. S2B). We show a snapshot of the RMSF distributions for the CENP-A NCP to illustrate how RFC intersects the RMSF data (Fig. S2C).

The FERN value was selected so that exterior higher fluctuation residues are cleaved from dimension calculations, and boundary motions are probed versus internal motions. FERN is applied to each cylinder surface. We visualized the resulting cylinder dimensions with varying

FERN values (Fig. S3A). For parameter sensitivity analysis, we tested a range of FERN values from 1 to 30 (3 to 90 residues for the whole cylinder). For a FERN value of 1-5, we saw a continuous increase in Young's modulus values (Fig. S3B, C). We visualized this region, labeled 1, and found that the cylinder boundaries were visibly larger than the nucleosome (Fig. S3A). From a FERN value of 5-13 ($0.5 \leq$ relative FERN $\leq 1.3$) we saw a plateau region in Young's modulus output (Fig. S3B, C). By visualization of the point labeled 2* in the plateau region, we saw the calculated cylinder appeared to be near the surface of the nucleosome (Fig. S3A) Beyond a FERN value of 13, we found that the Young's modulus increases to a maxima of nearly double and then drops (Fig. S3B, C). At the top of this peak (Fig. S3B) labeled region 3, we output the calculated cylinder dimensions and found the spike in rigidity was caused by the cylinder boundaries crossing through the center of the DNA double helix (Fig. S3A). At the peak, labeled 3, the cylinder was too small, intersected through the middle of the DNA, and the cylinder no longer approximated the boundaries of the nucleosome (Fig. S3A). Therefore, to measure the dimensional boundaries of the nucleosome we selected a value of 10 residues, which lies within the plateau region at 2*, where we found that the nucleosome fits more snuggly within the cylinder (Fig. S3A). As a control, the parameters FERN and AWL were held constant between NCP systems during analysis.

### **S3.** ***All-atom Structural Analysis***

Simulation data was truncated to include the final 400 ns for analysis and then contact analysis was performed. A cutoff distance of 8 Å was used between histone Cα atoms to compare dimer interface formed between CENP-A:CENP-A vs. CENP-A:H3.3. A value of 1 indicates contact at all frames of the analyzed trajectory and a value of 0 indicates an absence of contact during simulation. Principle component analysis (PCA) was performed on the histone core based on previously published methods (3). The magnitude of motion is multiplied by a factor of 5 in the movies to amplify motions for visual clarity.

DNA was analyzed by residue RMSF with errors calculated over three trajectory segments from 600 ns to 1000 ns of total simulation time. DNA end distances were then calculated from the entry to exit ends of the heterotypic nucleosome and the CENP-A nucleosome, compared to H3 (3).

## S4. *Extended Derivation of Young's modulus from MCA*

The work done in the deformation of an elastic material is stored in the form of strain energy, U. The strain energy density, u, the energy stored in small volume elements, can be useful to describe variable strains along a body that sum to the total strain energy:

$$U = \iiint_R u(r,\theta,z) r dr\, d\theta\, dz \; . \text{ (1)}$$

Because the extent of cylinder fluctuations is relatively small, in the range of 1.8 to 9.1% of the average radial or lateral dimension, we rely on linear elasticity and small-deformations' theory. Under these conditions, the strain energy density in cylindrical coordinates can be calculated for low magnitude stresses from arbitrary directions (4) as

$$u = \frac{1}{2}(\sigma_{rr}\varepsilon_{rr} + \sigma_{\theta\theta}\varepsilon_{\theta\theta} + \sigma_{zz}\varepsilon_{zz}) + (\sigma_{r\theta}\varepsilon_{r\theta} + \sigma_{\theta z}\varepsilon_{\theta z} + \sigma_{zr}\varepsilon_{zr}), \text{ (2)}$$

where $\sigma_{ij}$ is the stress in the i-th direction from force applied in the j-th direction and $\varepsilon_{ij}$ is the strain in the i-j plane (Fig. 1C). In Cartesian coordinates, the strain energy density of a volume element under arbitrary stresses can be found at Eq. 8.2.18 of (4) and is given as

$$u = \frac{1}{2}(\sigma_{xx}\varepsilon_{xx} + \sigma_{yy}\varepsilon_{yy} + \sigma_{zz}\varepsilon_{zz}) + (\sigma_{xy}\varepsilon_{xy} + \sigma_{yz}\varepsilon_{yz} + \sigma_{zx}\varepsilon_{zx}), \text{ (S1)}$$

where $\sigma_{ij}$ is the stress in the i-th direction from force applied in the j-th direction and $\varepsilon_{ij}$ is the strain in the i-j plane. In the main text, Eq. 2 can be derived from and is in the same form as (S1). In the absence of internal torques, stress and strain are both second order symmetric tensors (5). This fact can then be used to understand the origin of Eq. S1. As a note, the derivation of Eq. 2 from Eq. S1 can also be done using trigonometric identities or Einstein summation notation.

First,

$$\boldsymbol{\sigma\varepsilon} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}. \text{(S2)}$$

We will take the trace of the matrix product, and so the diagonal elements are

$$(\boldsymbol{\sigma\varepsilon})_{11} = \sigma_{xx}\varepsilon_{xx} + \sigma_{xy}\varepsilon_{xy} + \sigma_{zx}\varepsilon_{zx},$$

$$(\boldsymbol{\sigma\varepsilon})_{22} = \sigma_{xy}\varepsilon_{xy} + \sigma_{yy}\varepsilon_{yy} + \sigma_{zy}\varepsilon_{zy},$$

$$(\boldsymbol{\sigma\varepsilon})_{33} = \sigma_{xz}\varepsilon_{xz} + \sigma_{yz}\varepsilon_{yz} + \sigma_{zz}\varepsilon_{zz}. \text{(S3)}$$

Therefore,

$$Tr(\boldsymbol{\sigma\varepsilon}) = \left(\sigma_{xx}\varepsilon_{xx} + \sigma_{yy}\varepsilon_{yy} + \sigma_{zz}\varepsilon_{zz}\right) + 2\left(\sigma_{xy}\varepsilon_{xy} + \sigma_{yz}\varepsilon_{yz} + \sigma_{zx}\varepsilon_{zx}\right), \text{(S4)}$$

And from Eq. S1,

$$u = \frac{1}{2} Tr(\boldsymbol{\sigma\varepsilon}). \text{(S5)}$$

The form shown in Eq. S5 becomes useful because of the cyclic property of trace. We will consider a transformation matrix, **O,** which is any orthonormal basis:

$$\frac{1}{2} Tr\left((\boldsymbol{O\sigma O^T})(\boldsymbol{O\varepsilon O^T})\right). \text{(S6)}$$

Since **O** is any orthonormal basis, $\mathbf{O}^T\mathbf{O} = \mathbf{1}$, Eq. S6 simplifies to

$$\frac{1}{2} Tr(\boldsymbol{O^T O\sigma\varepsilon}) = \frac{1}{2} Tr(\boldsymbol{\sigma\varepsilon}). \text{(S7)}$$

This means that Eq. S1 can be written in the form Eq. S5, the form of Eq. S1 will be invariant to any orthonormal basis set. Therefore, Eq. S1 in Cartesian coordinates can be written in cylindrical coordinates as shown in Eq. 2.

From Eq. 2, since our output from the MCA is in terms of strain, we solve for strains from the stress-strain relations of Hooke 's Law, which can be found from a Solid Mechanics reference at Eq. 6.1.8 (4). When solving for on diagonal stresses, the standard result is obtained that

$$\sigma_{rr} = \frac{E}{(1+\nu)(1-2\nu)}[(1-\nu)\varepsilon_{rr} + \nu(\varepsilon_{\theta\theta} + \varepsilon_{zz})]\,,$$

$$\sigma_{\theta\theta} = \frac{E}{(1+\nu)(1-2\nu)}[(1-\nu)\varepsilon_{\theta\theta} + \nu(\varepsilon_{rr} + \varepsilon_{zz})]\,,$$

$$\sigma_{zz} = \frac{E}{(1+\nu)(1-2\nu)}[(1-\nu)\varepsilon_{zz} + \nu(\varepsilon_{\theta\theta} + \varepsilon_{rr})]\,, \text{(S8)}$$

where E is Young's modulus and ν is the Poisson ratio (4). Using the relations found in Eq. S8, in the absence of shear stresses and using Hooke's law, the strain energy density in Eq. 2 can also be written in the form

$$u = \frac{\nu\mu}{1-2\nu}(\varepsilon_{rr} + \varepsilon_{\theta\theta} + \varepsilon_{zz})^2 + \mu({\varepsilon_{rr}}^2 + {\varepsilon_{\theta\theta}}^2 + {\varepsilon_{zz}}^2)\,, \text{(3)}$$

where $\mu$ is the shear modulus and is related to Young's modulus, E, by $\mu = E / 2(1+\nu)$. For displacements ($u_r$, $u_\theta$, $u_z$) in cylindrical coordinates (r, θ, z) as shown in Fig. 1C (6):

$$\varepsilon_{rr} = \frac{\partial u_r}{\partial r}\,, \qquad \varepsilon_{\theta\theta} = \frac{u_r}{r} + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta}\,, \qquad \varepsilon_{zz} = \frac{\partial u_z}{\partial z}\,. \text{(4)}$$

In the case of a homogeneous axisymmetric cylinder, the center-of-mass at the origin (Fig. 1C), $\frac{\partial u_\theta}{\partial \theta} = 0$ and at the walls of the cylinder $\varepsilon_{rr} = \frac{u_r}{r_{avg}}$ . For the case of a homogenous cylinder, we will use the notation Δr for the displacement in r, $u_r$. Therefore, in this specific case, $\varepsilon_{rr} = \varepsilon_{\theta\theta}$ in Eq. 4. For homogenous cylinders (7):

$$\varepsilon_{rr} = \frac{\partial u_r}{\partial r} = \left.\frac{\partial u_r}{\partial r}\right|_{r = r_{avg}} = \frac{\Delta r}{r_{avg}} \text{ (S9)}$$

where $\frac{\Delta r}{r_{avg}}$ is input from MCA. We will then use the relation from Eq. 4 that in our case $\varepsilon_{rr} = \varepsilon_{\theta\theta}$ and rewrite. Eq. 3 in terms of the Young's modulus, our solvable, and strains to obtain

$$u = \frac{\nu E}{2(1-2\nu)(1+\nu)}(2\varepsilon_{rr} + \varepsilon_{zz})^2 + \frac{E}{2(1+\nu)}\left(2\varepsilon_{rr}{}^2 + \varepsilon_{zz}{}^2\right) \text{,(S10)}$$

Eq. S10 can then be simplified by factoring to arrive at Eq. 5:

$$u = \frac{E}{2(1+\nu)}\left[\frac{\nu(\varepsilon_{zz} + 2\varepsilon_{rr})^2}{(1-2\nu)} + \varepsilon_{zz}{}^2 + 2\varepsilon_{rr}{}^2\right] \text{.(5)}$$

Strain values in Eq. 5 are calculated from the measured quantities $r_{avg}$, $z_{avg}$, $\Delta r$, and $\Delta z$ using our minimal cylinder analysis (Methods 2.1.3).

We next focus on the acoustic cylindrical mode of motion that describes compression in the z-axis along with radial extension (and *vice versa*). The average potential energy of this mode can be estimated from the equipartition theorem, $U = \frac{1}{2}k_bT$, where $k_b$ is the Boltzmann constant and T is the simulation temperature, 300 K. We then integrate Eq. 5 over the body volume, Eq. 1, and with the above-mentioned energy from equipartition theorem we solve for Young's modulus:

$$E = \frac{k_bT(1-\nu-2\nu^2)}{V\left(\varepsilon_{zz}{}^2 - \nu\varepsilon_{zz}{}^2 + 2\varepsilon_{rr}{}^2 + 4\nu\varepsilon_{zz}\varepsilon_{rr}\right)} \text{. (6)}$$

### S5. *The heterotypic nucleosome has an exposed CENP-A RG loop*

The symmetry breaking in the heterotypic nucleosome across the pseudo-dyad, also propagates away from the CENP-A:H3.3 interface to the RG loop (R80, G81) of CENP-A, L1 (Fig. S5C). The CENP-A histone displays increased exposure of L1 to solvent in the case of the heterotypic nucleosome in comparison to the CENP-A nucleosome (Movie S1). For viable kinetochore formation at the heterotypic nucleosome, other proteins such as CENP-N must be recruited (8, 9). The high degree of solvent exposure in the heterotypic nucleosome may indicate that CENP-N is able to bind to its established binding site, CENP-A R80 and G81 (9). It is not yet understood if a single copy of CENP-C and CENP-N are sufficient for kinetochore formation, though minimally, our work on the rigidification of CENP-A when bound to its partners (Melters, Pitman, Rakshit *et al. in press 2019*), indicates that these factors would further stabilize CENP-A of heterotypic nucleosomes at ectopic sights.

## Supporting References

1. Melters, D.P., M. Pitman, T. Rakshit, E.K. Dimitriadis, M. Bui, G.A. Papoian, and Y. Dalal. 2019. Intrinsic elasticity of nucleosomes is encoded by histone variants and calibrated by their binding partners. Proc. Natl. Acad. Sci. U. S. A. 116: 24066–24074.
2. Zhou, B.R., J. Jiang, H. Feng, R. Ghirlando, T.S. Xiao, and Y. Bai. 2015. Structural Mechanisms of Nucleosome Recognition by Linker Histones. Mol. Cell. 59: 628–638.
3. Winogradoff, D., H. Zhao, Y. Dalal, and G.A. Papoian. 2015. Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome. Sci. Rep. 5.
4. Piaras, K. 2015. Solid Mechanics Part 1: An introduction to Solid Mechanics. Solid Mech. Part I An Introd. to Solid Mech. Lect. notes. : 183–191.
5. O'connor, P.D.T. 1988. Advanced strength and applied elasticity, A.C. Ugural and S.K. Fenster, Second SI Edition, Elsevier Science Publishing Company Inc., 1987.
6. Jaeger, M.T. Basic Relationships in Elasticity Theory. Available at: https://dspace.mit.edu/bitstream/handle/1721.1/39137/22-

314JSpring2004/NR/rdonlyres/Nuclear-Engineering/22-314JSpring2004/ACF104D9-8C23-4EEB-B3CD-D350871D8EDE/0/elastic_theory.pdf.

7. Williams, J.G., and C. Gamonpilas. 2008. Using the simple compression test to determine Young's modulus, Poisson's ratio and the Coulomb friction coefficient. Int. J. Solids Struct. 45: 4448–4459.
8. 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.
9. Tian, T., X. Li, Y. Liu, C. Wang, X. Liu, G. Bi, X. Zhang, X. Yao, Z.H. Zhou, and J. Zang. 2018. Molecular basis for CENP-N recognition of CENP-A nucleosome on the human kinetochore. Cell Res. 28: 374–378.