Supporting Information

Cho and Harrison 10.1073/pnas.1106389108

SI Methods.

Cloning, Expression, and Purification of *Saccharomyces cerevisiae* (*Sc*) Scm3, Cse4, and H4. Clones of *Sc* Cse4, H4, and Scm3 were obtained by PCR amplification from a *S. cerevisiae* genomic library and subcloned into pET vectors modified for ligation-independent cloning (LIC). Scm3 was cloned into an LIC vector encoding an N-terminal His_6 tag and a tobacco etch virus (TEV) cleavage site. Cse4 and H4 were expressed without affinity tags. Constructs were transformed into Rosetta (DE3) pLysS *Escherichia coli* competent cells (Novagen) and used for protein expression. For *Sc* Cse4 and H4 purification, cells were grown at 37 °C in LB medium and induced at OD = 0.6 with 0.2 mM isopropyl β-D-1-thiogalactopyranoside. Harvested cells were resuspended in 30 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 3 mM β-merchaptoethanol with protease inhibitor cocktail tablets. Cells were sonicated and insoluble pellets were harvested from centrifugation at 20,000 × *g* for 1 h. Pellets were then resuspended in inclusion-body washing buffer (30 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1% triton X-100, and 3 mM β-merchaptoethanol), incubated at 4 °C for 1 h shaking, and centrifuged at 20,000 × *g* for 30 min. This washing procedure was repeated three times. Isolated inclusion bodies of *Sc* Cse4 and H4 were solubilized in 30 mM Tris-HCl, pH 8.0, 7 M guanidinium hydrochloride (GuHCl), 10 mM DTT, and insoluble debris was removed by centrifugation at 20,000 × *g* for 30 min. Refolding of 7 M GuHCl solubilized *Sc* Cse4 and H4 proceeded by two-step dialysis in the presence or absence of *Sc* Scm3. Step 1 was dialysis against 30 mM Tris-HCl, pH 8.0, 2 M NaCl, 10 mM DTT; step 2 was dialysis against 30 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 1 mM DTT. *Sc* Scm3 was expressed and purified as described for *Kluyveromyces lactis* Scm3 (see *Methods*). Misfolded proteins were removed by centrifugation at 20,000 × *g* for 30 min. Soluble fractions were then concentrated and applied to a Superdex 200 size exclusion column (Prep grade 16/60: GE healthcare) preequilibrated with 30 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM TCEP [tris(2-carboxyethyl)phosphine].

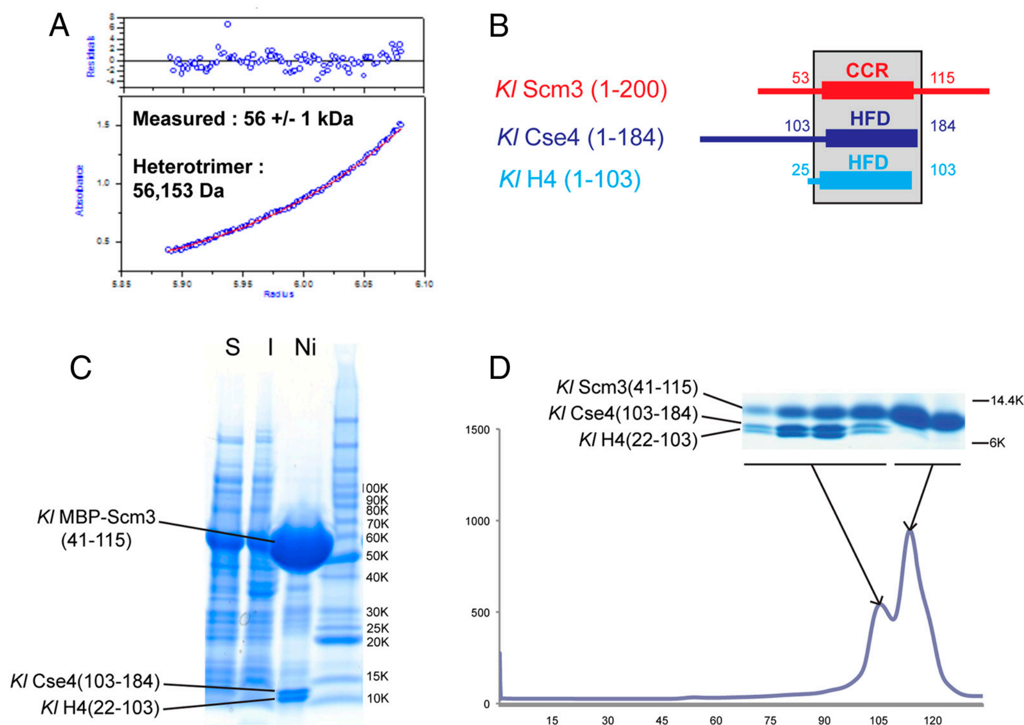

Fig. S1. Properties of *K. lactis* (*Kl*) Scm3:Cse4:H4 and its minimum stable complex. (*A*) Sedimentation equilibrium analytical ultracentrifugation of *Kl* Scm3:Cse4:H4 complex; the measured molecular mass corresponds to a heterotrimer. (*B*) Minimum required fragments of *Kl* Scm3, Cse4, and H4 for stable complex formation. Purified Scm3:Cse4:H4 complex was treated with either trypsin or subtilisin; protease-resistant fragments (gray box) were determined by mass spectrometry. CCR, central conserved region of Scm3; HFD, histone-fold domain. (*C*) Expression of the minimum stable complex of *Kl* Scm3(41–115):Cse4(103–184):H4(22–103). Scm3 (41–115) is sufficient for proper folding of Cse4 HFD and H4 HFD. S, soluble fraction; I, insoluble fraction; Ni, Ni-NTA eluate. (*D*) Purification of *Kl* Scm3(41–115):Cse4(103–184):H4(22–103) by size exclusion chromatography. SDS-PAGE gel shows that the three proteins coelute as a stable complex.

Fig. S2. *S.cerevisiae* (*Sc*) Scm3 as a chaperone for *Sc* Cse4 and H4. (*A*) Refolding of *Sc* Cse4 and H4 in the absence of *Sc* Scm3 was carried out as described in *SI Methods*. The dialyzed sample was centrifuged to remove misfolded proteins after dialysis (lane labeled "Soluble"). As shown here, *Sc* Cse4 and H4 failed to refold in the absence of *Sc* Scm3. (*B*) Refolding of *Sc* Cse4 and H4 in the presence of *Sc* Scm3. Refolding *Sc* Cse4 and H4 together with *Sc* Scm3 yields a stable complex as shown in the column profile from Superdex 200 size exclusion chromatography. SDS-PAGE gel of column fractions shows that the three proteins coelute as a complex. (*C*) Sedimentation equilibrium analytical ultracentrifugation of the *Sc* Scm3:Cse4:H4 complex. As with the *K. lactis* Scm3:Cse4:H4 complex, the molecular mass of the *Sc* Scm3:Cse4:H4 complex corresponds to a heterotrimer. (*D*) Minimum required fragments of *Sc* Scm3, Cse4, and H4 for stable complex formation. Purified Scm3:Cse4:H4 was treated with subtilisin; protease-resistant fragments (gray box) were determined by mass spectrometry. CCR, central conserved region of Scm3; HFD, histone-fold domain.

A

```
KLa_Scm3p/41-115   RDG------VVYIMSKENRLIPKLSDEEVMERHKKADENMKRVWSQIIQKYESIDNQGDVIDLQTGEVIEDNGHLRNLSNE
SCe_Scm3p/64-144   SRPIAERNGHVYIMSKENHIIPKLTDDEVMERHKLADENMRKVWSNIISKYESIEEQGDLVDLKTGEIVEDNGHIKTLTAN
CGl_Scm3p/62-142   YKEVKKKNGRAYIYSKENQLIPKLTDEEVMEKHKLADQRMKTVWSNIIDKYSNVVDEGDVIDLQSGEIITDNGHIRNLNTG
AGo_Scm3p/34-108   PAG------VVYVMSRENEPIPRLSDAEVMERHRRADENMKQAWTRIIEKYEALDDQGDVVDLRTGEIVEDNGHIRGLADG
SKu_Scm3p/65-145   VRPVAERNGHVYIMSKENQVIPKLTDDEVMERHKRADENMKEVWSNIISKYESIEDQGDLVDLKTGEIIEDNGHIKKLTVN
SKl_Scm3p/66-146   NDPVTQHDGVTFIMSKENQLIPKLTEEEVIERHRKADENMKQVWSKIIQKYENVEDQGDLIDLRTGEIVEDNGHLRGLSAT
KWa_Scm3p/67-147   EQGTQEHDNVVYIMSKESQLIPKLTEEQVMDRHKRADENMKRAWLQLIQKYEALEDQGDVLDLNTGEIVEDNGHVRSLSER
```

```
KLa_Cse4p/103-184   STDLLISRMPFARLVKEVTDQFTTESEPLRWQSMAIMALQEASEAYLVGLLEHTNLLALHAKRITIMRKDMQLARRIRGQFI
SCe_Cse4p/148-229   STDLLISKIPFARLVKEVTDEFTTKDQDLRWQSMAIMALQEASEAYLVGLLEHTNLLALHAKRITIMKKDMQLARRIRGQFI
SKu_Cse4p/148-229   STDLLISKIPFARLVKEVTDEFTTKDQDLRWQSMAIMALQEASEAYLVGLLEHTNLLALHAKRITIMKKDMQLARRIRGQFI
KWa_Cse4p/102-183   STDLLISKMPFARLVKEVTDQYTTEEQQLRWQSMAIMALQEASEAYLVGLLEHTNLLALHAKRVTIMRKDMQLARRIRGQFI
SKl_Cse4p/59-140    STELLISKMPFARLVKEVTEQFSTDEQQLRWQSMAILALQEASEAYLVGLLEHTNLLALHAKRITIMKKDMQLARRIRGQFI
AGo_Cse4p/123-204   STELLISRMPFARLVKEVTDQFTTVDQQMRWQSMAILALQEASEAYIVGLLEHTNLLALHAKRVTVMRKDMQLARRIRGQFI
CGl_Cse4p/159-240   STALLIQKIPFAKLVKEVTEEFAGESQDLRWQSMAILALQEASEAYLVGLLEHTNLLALHAKRITIMKKDMQLARRIRGQFI
```

```
KLa_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEEVRNVLKTFLESVIRDAVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
SCe_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
CGl_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEEVRAVLKSFLESVIRDAVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
AGo_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEDVRAVLKSFLESVIRDAVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
SKu_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
SKl_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEEVRTVLKTFLESVIRDAVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
KWa_H4/22-103   ILRDNIQGITKPAIRRLARRGGVKRISGLIYEEVRAVLKSFLESVIRDAVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
```

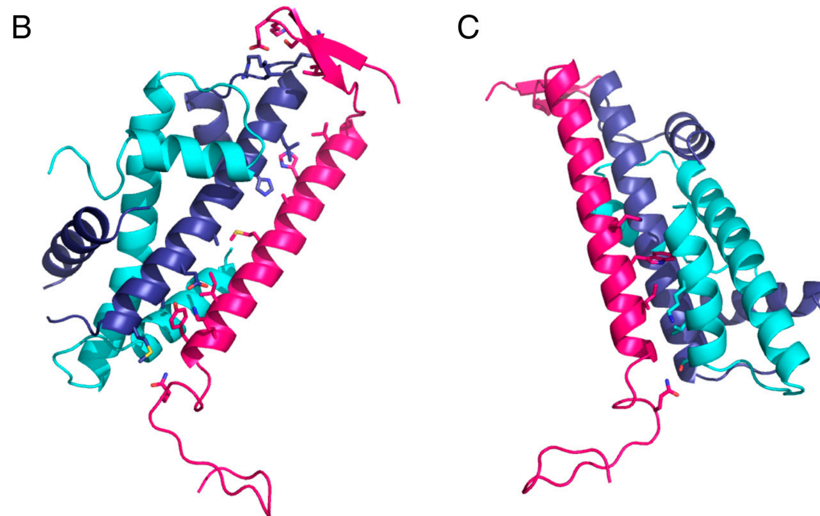

Fig. S3. Interface of Scm3:Cse4 and Scm3:H4 in the heterotrimer. (*A*) Multiple sequence alignments of Scm3 (central conserved region), Cse4 and H4 (HFDs) from point-centromere yeasts. Blue boxes represent residues at the interface of Scm3 and Cse4; red boxes, residues at the interface of Scm3 and H4. *KLa*, *K. lactis*; *SCe*, *S. cerevisiae*; *CGl*, *Candida glabrata*; *AGo*, *Ashbya gossypii*; *SKu*, *Saccharomyces kudriavzevii*; *SKl*, *Saccharomyces kluyveri*; *KWa*, *Kluyveromyces waltii*. (*B*) Interface of Scm3 and Cse4. Interacting residues are shown as stick models. (*C*) Interface of Scm3 and H4, in a similar representation.

A

```
K.lactis_Cse4/103-184         STDLLISRMPFARLVKEVTDQFTT---ESEPLRWQSMAIMALQEASEAYLVGLLEHTNLLALHAKRITIMRKDMQLARRIRGQFI---
S.cerevisiae_Cse4/148-229     STDLLISKIPFARLVKEVTDEFTT---KDQDLRWQSMAIMALQEASEAYLVGLLEHTNLLALHAKRITIMKKDMQLARRIRGQFI---
Human_CENPA/57-140            STHLLIRKLPFSRLAREICVKFTR----GVDFNWQAQALLALQEAAEAFLVHLFEDAYLLTLHAGRVTLFPKDVQLARRIRGLEEGLG
Mouse_CENPA/51-134            STDLLFRKKPFSMVVREICEKFSR----GVDFWWQAQALLALQEAAEAFLIHLFEDAYLLSLHAGRVTLFPKDIQLTRRIRGFEGGLP
Bovine_CENPA/55-138           TTHLLLRKSPFCRLAREICVQFTR----GVDFNWQAQALLALQEAAEAFLVHLFEDAYLLSLHAGRVTLFPKDVQLARRIRGIQEGLG
S.pombe_Cnp1/38-120           STDLLIQRLPFSRIVREISSEFVANFSTDVGLRWQSTALQCLQEAAEAFLVHLFEDTNLCAIHAKRVTIMQRDMQLARRIRGA-----
```

B

```
K.lactis_Scm3/41-115          -RDGVVYIMSKENRLIPKLSDEEVMERHKKADENMKRVWSQIIQKY-ESIDNQGDVIDLQTGEVIEDNGHLRNLSNE
S.cerevisiae_Scm3/69-144      ERNGHVYIMSKENHIIPKLTDDEVMERHKLADENMRKVWSNIISKY-ESIEEQGDLVDLKTGEIVEDNGHIKTLTAN
Human_HJURP/1-69              -MLGTLRAMEGED-----VEDDQLLQKLRASRRRFQRRMQRLIEKY-NQPFEDTPVVQMATLTYETPQG-LRIWGGR
Mouse_HJURP/1-62              --------MESMG-----RQDRRLHQQLKESSSRFQTLMKRLIAKY-NQPFEDDPLVEMRTLTYETPQG-LRVWGGK
Bovine_HJURP/1-62             --------MEGEV-----LSEDELLMKLRDSRCRFQRRMQQLIEKY-NQPFEDGPVIQMSTLTYQTPQG-LRIWGGG
S.pombe_Scm3/16-91            -SCPHLLSFQSVEFNTNPTACDDVFCKRIESEKKYNDFLESLFKKYGRDTSDIADEVDLATGEIIVNNGHLEALKTK
```

Fig. S4. Multiple sequence alignments of centromeric histone H3s and its chaperone proteins. (*A*) Multiple sequence alignment of the HFDs of Cse4 and its orthologs from other eukaryotes, including human CENP-A. (*B*) Multiple sequence alignment of the central conserved region of Scm3 and the N-terminal regions of CenH3 chaperones from other eukaryotes, including human HJURP. The Lys–Tyr pair, conserved among CenH3 chaperones from yeast to man, is marked in the red box.

Fig. S5. Comparison of the Scm3/Cse4/H4 structures and its human orthologs. (*A*) The ribbon diagram of the *K. lactis* Scm3/Cse4/H4 structure reported here, that of the *S. cerevisiae* single chain construct [Zhou et al., Protein Data Bank (PDB) ID code 2L5A] (1), and its human orthologs (HJURP/CENP-A/H4, PDB ID code 3R45) (2). The overall structures of the Scm3/Cse4/H4 complex and its human orthologs are similar with rmsd of 0.7 Å for 176 overlapping residues. (*B*) Primary sequence comparison of fragments used in the Zhou et al. (NMR) experiment and in ours (X-ray). In the Zhou et al. structure, a conserved N-terminal segment of Scm3 (70–92) was removed, and an essential part of H4 HFD (helix I, 20–41) was truncated.

1 Zhou Z, et al. (2011) Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature* 472:234–237.
2 Hu H, et al. (2011) Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev* 25:901–906.

Fig. S6. Two $(Cse4 : H4)_2$ tetramers in the asymmetric unit of space group R3. Ribbon diagram of two tetramers of $(Cse4 : H4)_2$ in the asymmetric unit. Superposition of two tetramers shows the same Cse4:Cse4 tetramer interface. The contacts between tetramers, while relatively intimate, do not mimic any known nucleosomal interactions and probably represent interactions generated during crystallization.

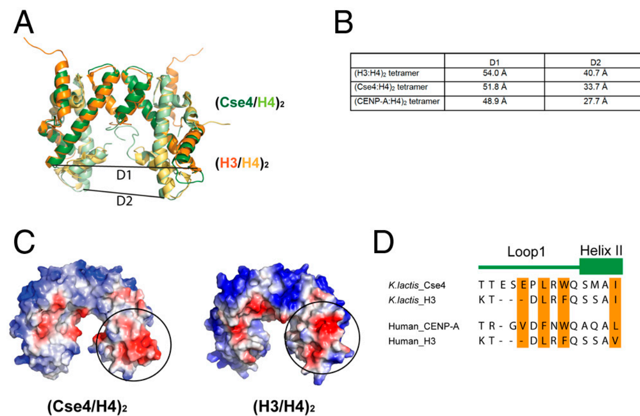

Fig. S7. Comparison of $(\mathrm{Cse4:H4})_2$ tetramer with $(\mathrm{H3/H4})_2$ tetramer. (*A*) Comparison of $(\mathrm{Cse4:H4})_2$ and $(\mathrm{H3:H4})_2$ (PDB ID code 1ID3). Two sets of distances (D1 and D2) were measured to compare the compactness of each tetramer. (*B*) Measured distances, D1 and D2, from $(\mathrm{H3:H4})_2$, $(\mathrm{Cse4:H4})_2$, and $(\mathrm{CENP\text{-}A/H4})_2$ tetramer. (*C*) Comparison of surface charge distribution in the CENP-A targeting domain (CATD) of $(\mathrm{Cse4:H4})_2$ and $(\mathrm{H3:H4})_2$. In contrast with $(\mathrm{CENP\text{-}A/H4})_2$, the surface charge distributions in $(\mathrm{Cse4:H4})_2$ and $(\mathrm{H3:H4})_2$ at CATD are similarly negative. (*D*) Sequence alignment of loop1 from CENP-A vs. H3 and from Cse4 vs. H3. Residues proposed to increase hydrophobicity in loop1 of CENP-A are marked by orange boxes. Cse4 counterparts of those hydrophobic residues are not more hydrophobic than their counterparts in H3.

Table S1. Data collection and refinement statistics for the Scm3:Cse4:H4 and $(\mathrm{Cse4:H4})_2$ complexes

Data collection	Scm3:Cse4:H4 complex	$(\mathrm{Cse4:H4})_2$ complex
Resolution (last shell), Å	50–2.3 (2.38–2.30)	50–2.6 (2.69–2.60)
Wavelength, Å	0.9795	0.9795
Space group	$P22_12_1$	R3
Unit cell dimensions (*a*, *b*, *c*), Å	32.72, 65.53, 121.10	169.48, 169.48, 81.22
Unit cell Angles (α, β, γ), °	90, 90, 90	90, 90, 120
I/σ (last shell)	17.0 (1.5)	15.0 (1.3)
R_{sym} (last shell), %*	0.085 (0.637)	0.115 (0.814)
Completeness (last shell), %	94.7 (55.3)	99.9 (98.8)
Number of reflections	47,110	100,769
unique	11,636	26,486
Redundancy	4.1	3.8
Refinement		
Resolution, Å	50–2.3	50–2.6
No. of reflections	11,073	25,302
working	10,518	23,960
free	555	1,342
R_{work}, %†	22.0	21.8
R_{free}, %†	25.7	27.6
Ramachandran plot, % (favored/additional/disallowed)	99.5/0.5/0	98.3/1.7/0
Structure/stereochemistry		
No. of atoms		
protein	1,636	4,808
iodide	2	
H_2O	58	14
rmsd bond lengths, Å	0.014	0.011
rmsd bond angles, °	1.423	1.403

*$R_{sym} = \sum \sum_j |I_j - \langle I \rangle| / \sum I_j$, where I_j is the measured intensity of reflection j, and $\langle I \rangle$ is the mean intensity for multiply recorded reflections.

†$R_{work,free} = \sum ||F_{obs}| - |F_{calc}|| / |F_{obs}|$, where the working and free R factors are calculated using the working and free reflection sets, respectively.