Supplemental Information

Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3

Zheng Zhou, Hanqiao Feng, Bing-Rui Zhou, Rodolfo Ghirlando, Kaifeng Hu, Adam Zwolak, Lisa M. Miller Jenkins, Hua Xiao, Nico Tjandra, Carl Wu, Yawen Bai

The file contains Supplementary Methods, Supplementary Figures 1-21 with legends, Supplementary Tables 1-3 and additional references.
Methods

Protein sample preparation

All proteins were expressed in *E. coli* (BL21-codonPlus(DE3)-RIL) with pET vectors (Stratagene). N-terminal His6-tagged fragments of Cse4 and Scm3 and their mutants were first purified via Ni-NTA (Qiagen) whereas H4 and non-His-tagged fragments of Cse4 and Scm3 were first purified via SP sepharose (GE Healthcare). They were next subjected to reverse phase HPLC purification using acetonitrile and water as solvents. Purified proteins were lyophilized. Isotope-labeled proteins for NMR studies were produced by growing *E. coli* cells in M9 media with $^{15}$NH$_4$Cl, U-$^{13}$C$_6$-Glucose, and D$_2$O as the sole source for nitrogen, carbon, and deuterium, respectively. For the measurement of side chain NOEs, specific methyl labeling ($^{13}$CH$_3$) for Ile, Leu, and Val residues was also made following the protocol of Kay and coworkers$^1$.

To prepare the Cse4, H4 and Scm3 complexes, lyophilized proteins were first dissolved in H$_2$O. Their concentrations were determined by measuring the absorbance at 280 nm. Equal amounts of each species were mixed together and dialyzed against 10 mM Tris-HCl and 2 M NaCl at pH 7.4 and 4°C. After centrifugation, the soluble fractions were subjected to gel filtration on Superdex 75 10/300 GL column (GE Healthcare, USA). The eluted complexes were concentrated with an Amicon with Ultra Ultracel-10 membrane (Millipore, USA) and exchanged to a final buffer of 50 mM MES at pH 5.6. The Cse4-H4 complexes were made in the same way. Scm3$^{56-187}$ samples for NMR study were prepared by dissolving them in 8M urea and dialysis against corresponding buffer. scSCH and all other single chain molecules derived from scSCH are purified with Ni-NTA (Qiagen) under native conditions (20 mM Tris-HCl and 0.5 M NaCl at pH 8.0), followed by gel filtration with Superdex 200 10/60 column at 4°C (GE healthcare, USA). The fractions with target protein were concentrated and exchange to its final buffer.

Analytical ultracentrifugation

Sedimentation velocity experiments were conducted in duplicate at 20.0°C
on a Beckman Coulter Proteome XL-I analytical ultracentrifuge. 400 µL of the sample of 35 µM in 50 mM Mes (pH, 5.6) was loaded in a double sector centerpiece cell and analyzed at a rotor speed of 50 krpm. 100 scans were acquired as single absorbance measurements (λ = 280 nm) at 7.1 minute intervals using a radial spacing of 0.003 cm. Data were analyzed in SEDFIT 11.71 in terms of a continuous c(s) distribution to obtain a sedimentation coefficient, s, and molecular mass M (ref 2). Solution densities ρ were measured at 20.00 °C on a Mettler Toledo DE51 density meter and solution viscosities η were measured using a Cannon-Ubbelohde viscometer and Cannon-CT 500 constant temperature bath set at 20.00 °C. The partial specific volume ν of the complex was calculated in SEDNTERP 1.09 (ref 3). c(s) analyses were carried out using an s-value range of 0.5 to 6.0 with a linear resolution of 100 and a confidence level (F-ratio) of 0.68. The analyses, implemented using time independent noise corrections, returned root mean square deviation (rmsd) values for the best fits of 0.0040 absorbance units. Sedimentation equilibrium experiments were conducted at 20.0°C on a Beckman Optima XL-A. 135 µL volumes of the complex were studied at loading concentrations of 20, 39 and 78 µM, along with the sample recovered from the sedimentation velocity experiments. Experiments were carried out using six channel centerpiece cells at rotor speeds ranging from 18 to 34 krpm. In all cases data were acquired as an average of 4 absorbance measurements at wavelengths of 280 and 250 nm using a radial spacing of 0.001 cm. Sedimentation equilibrium at each speed was achieved within 40 hours. Data were analyzed globally in terms of a single ideal species using SEDPHAT 6.21 (refs 3, 4).

NMR experiments

NMR experiments were performed on Bruker 500, 600, 800, and 900 MHz and Varian 600 and 800 MHz spectrometers at 35°C. The following experiments were recorded. 2D: ['H, 'H]-NOESY, ['H, 15N]-TROSY, ['H, 13C]-HMOC, 15N-{1H} NOE; TROSY version 3D: HNCACB, HNCCAC, HNCA, HNCOCA, HNCO, HNCA; 3D HBBACONH, HCH-TOCSY, CCH-TOCSY, CCC(CO)NH,
[\textsuperscript{1}H,\textsuperscript{15}N]-NOESY-HSQC, [\textsuperscript{1}H,\textsuperscript{15}N] NOESY-HSQC (\textsuperscript{13}C methyl-labeled sample), [\textsuperscript{1}H,\textsuperscript{13}C]-NOESY-HSQC, [\textsuperscript{1}H,\textsuperscript{13}C]-NOESY-HSQC (\textsuperscript{13}C methyl-labeled sample). The spectra were processed using NMRPipe\textsuperscript{5} and analyzed with NMRView\textsuperscript{6}.

**Structure calculation**

Structure calculation was done using the program Xplor-NIH\textsuperscript{7}. The NOE-derived restraints were subdivided into four classes, strong, medium, weak, and very weak by comparison with NOEs of protons separated by known distances as described in Zhou et al.\textsuperscript{8}. Backbone dihedral angle restraints ($\phi$ and $\psi$ angles) were obtained from analysis of $\textsuperscript{1}H_\alpha$, HN, $\textsuperscript{13}C_\alpha$, $\textsuperscript{13}C_\beta$, $\textsuperscript{13}CO$, and $\textsuperscript{15}N$ chemical shifts by using the program TALOS\textsuperscript{9}. Two constraints per hydrogen bond ($d_{NH-O} \leq 2.2$ Å and $d_{N-O} \leq 3.2$ Å) were added in the final structure calculation after initial NOE-derived structures were obtained. The program PROCHECK_NMR\textsuperscript{10} was used to evaluate the quality of the calculated structures.

**Isothermal titration calorimetric experiments**

The ITC experiments were performed on a MicroCal VP-ITC by injecting Scm\textsubscript{383-169} solution (250 µM) to a solution of single chain Cse4-H4 or their mutants (His6-KK-Cse4\textsubscript{151-207}-LVPRGS-H4\textsubscript{45-103}) solution (20 µM) in a chamber of 1.4 ml at 25 °C in 50 mM MES (pH 5.4) and 0.1 M NaCl. Twenty-nine injections (each of 10 µl) were made and the heat released was analyzed. The data was analyzed as described by Houtman et al.\textsuperscript{11}.

**Pull-down experiments**

Pull-down experiments were carried out in 50 mM sodium phosphate, 25 mM imidazole, 2 M NaCl, pH 8.0 at room temperature. Ni-NTA (Qiagen) beads were mixed with His6-Scm\textsubscript{366-169} with a final concentration of 6 µM. Approximately 10 fold excess of (Cse4-H4)$_2$ or (H3-H4)$_2$ or their mutants was mixed with beads and incubated. The incubation was at 25 °C for 30 minutes. The beads were washed with the same buffer three times. The complex formed on the beads was eluted with 250 mM imidazole and analyzed by SDS-PAGE.
Beads without His6-Scm3 were also incubated with corresponding (Cse4-H4)\textsubscript{2} under identical conditions to assess background binding and the integrity of the tetramer. No non-specific binding was identified in 2 M NaCl. For molecules derived from thrombin digested single chain proteins, the complex was incubated with Ni-NTA (Qiagen) beads at 25 °C for 30 minutes and then washed 3 times. The final complex formed on the beads was eluted with either 8 M Urea or 250 mM imidazole. The eluted molecules were analyzed by SDS-PAGE.

**Identification of the Scm3-Cse4-H4 complex for structural determination by NMR**

To find an Scm3-Cse4-H4 complex for structure determination, we explored the complexes formed with different fragments of Scm3, Cse4 and full length H4 in 1:1:1 ratio and checked their solubility and stability at room temperature. For Scm3, we chose the fragments that are available from the earlier work\textsuperscript{12}. For Cse4, the fragments (Supplementary Fig. 3) were chosen based on the earlier experimental results that show that the N-terminal region (1-129) of Cse4 is not needed for Cse4 to interact with Scm3\textsuperscript{13, 14} and the residues 136-150, which correspond to the region of the α\textsubscript{N} helix of H3, are disordered in the Asf1-H3-H4 complex\textsuperscript{15, 16}.

**The structure of the folded core in the scSCH is the same as that in the three-chain Scm3-Cse4-H4 complex**

To test whether the scSCH has the same folded core as the three-chain complex, we first examined the two-chain complexes, M-His6-KK-Cse4\textsubscript{150-227}-LVPR complexed with GS-Scm3\textsubscript{93-169}-GDK-H4\textsubscript{42-103} by cutting the first linker with thrombin (Supplementary Fig. 7a). The \textsuperscript{1}H-\textsuperscript{15}N TROSY spectrum of the two-chain complexes was overlaid with that of scSCH (Supplementary Fig. 7b). The residues that show large chemical shift changes are the neighbor residues of the cutting site in the sequence. The residues that show small chemical shift changes are close to the cutting site in the three-dimensional structure when considering the fact that the chain after the cut becomes more flexible.
(Supplementary Fig. 7a). The two-chain complex is stable enough to be studied by 3D NMR. We found that the residues in the C-terminal region of the α3 helix of H4 remains helical when their Cα chemical shifts are compared with the random coil values (Supplementary Fig. 7c).

A three-chain complex with the cutting of both linkers was further investigated. To do so, the six residues GDKGGS in the second linker were mutated to LVPRGS to create another thrombin cleavage site. The $^{1}$H-$^{15}$N TROSY spectra of the three-chain complex and the two-chain complex were overlaid. In this case, only the mutated residues and its nearest neighbors in sequence show significant chemical shift changes. Few chemical shift changes for other residues occurred, indicating the second linker has essentially no constraints on the folded region of the molecule (Supplementary Fig. 7d and Fig. 7e).

This conclusion is further supported by the result that Scm3 still can bind to Cse4-H4 with the deletion of the 20 residues in the C-terminal region of Cse4 in the pull-down experiment (Supplementary Fig. 19).
Supplementary Table 1. Statistical parameters for the scSCH structure.

|                        | Folded core | scSCH (Full length) |
|------------------------|-------------|---------------------|
|                        | (Cse4₁₅⁷-2₀¹, Scm3₉₇-₁₃₅, H₄₅₀-₉₉) |                     |
| **NMR distance and dihedral constraints** |             |                     |
| Total NOE              | 2881        | 3871                |
| Intra residues NOE     | 897         | 1198                |
| Inter residues NOE     | 1984        | 2673                |
| Sequential NOE (|i-j| = 1) | 767         | 1067                |
| Medium range NOE ([i-j] < 5) | 659     | 909                  |
| Long range NOE ([i-j] > 4) | 558      | 686                  |
| Inter-protein NOE      | 305         | 360                 |
| H-bonds                | 152         | 168                 |
| Dihedral angles        |             |                     |
| $\phi$ (TALOS)         | 115         | 142                 |
| $\psi$ (TALOS)         | 115         | 142                 |
| **Structure statistics** |             |                     |
| Violations (mean ± s.d.) |             |                     |
| Distance constraints (Å) | 0.029 ± 0.002 |                     |
| Dihedral angle constraints (°) | 0.419 ± 0.022 |                     |
| Max. dihedral angle violations (°) | 3.0 |                     |
| Max. distance constraint violations (Å) | 0.3 |                     |
| Deviation from idealized geometry |             |                     |
| Bonds (Å)              | 0.0036 ± 0.0005 |                     |
| Angles (°)             | 0.535 ± 0.025 |                     |
| Impropers (°)          | 0.389 ± 0.020 |                     |
| Average pairwise r.m.s. deviation (Å) |             |                     |
| Backbone atoms         | 0.49        | 1.65                |
| All heavy atoms        | 1.06        | 2.32                |
| **Procheck statistics** |             |                     |
| Residues in            |             |                     |
| most favored regions   | 83.9 ± 1.0  | 76.9 ± 1.0          |
| allowed regions        | 14.5 ± 0.5  | 18.9 ± 0.6          |
| generously allowed regions | 1.6 ± 0.1   | 3.5 ± 0.3           |
| disallowed regions     | 0.0 ± 0.0   | 0.7 ± 0.1           |
**Supplementary Table 2. Binding parameters between Scm3 and Cse4-H4 and their mutants.**

|       | N  | $K_a$ (M$^{-1}$) | $K_a$(wt)/$K_a$(mutant) |
|-------|----|-----------------|-------------------------|
| WT    | 0.94 | $1.7 \times 10^6$ | 1.0                     |
| II    | 1.08 | $2.0 \times 10^4$ | 85                      |
| VL--I | 0.94 | $1.4 \times 10^5$ | 12                      |
| MM    | 1.05 | $3.1 \times 10^6$ | 5.5                     |
| AS    | 0.90 | $1.8 \times 10^5$ | 9.4                     |
| LY    | 0.99 | $1.0 \times 10^6$ | 1.7                     |
| T-KDQ | 1.04 | $1.5 \times 10^6$ | 1.1                     |
| VTDE  | 0.88 | $1.2 \times 10^6$ | 1.4                     |

In the titration experiments, Scm3 (Scm3$^{83-169}$) or its mutant was titrated into the solution of a single-chain Cse4-H4 (His6-KK-Cse4$^{151-207}$-LVPRGS-H4$^{45-103}$) or its mutants (Supplementary Methods). Here N is the binding stoichiometry between Scm3 and Cse4-H4. $K_a$ is the association equilibrium constant. The standard deviations for N and $K_a$ from the fitting are typically less than 15%. Mutants are:

- II, Scm3(Ile110Asp/Ile117Asn);
- VL--I, Scm3(Val157Gly/Leu158Gly/Ile161/Gly);
- MM, Cse4(Met181Ser/Met184Gly);
- AS, Cse4(Ala189Ser/Ser190Val);
- T-KDQ, Cse4($\Delta$(172-174)/Thr170Lys);
- VTDE, Cse4(Val165Ile/Thr166Ala/Asp167Gln/Glu168Asp);
- LY, H4(Leu98Gly/Tyr99Gly).
## Supplementary Table 3. Protein samples used in the experiments.

| Sample name | Scm3/HJURP | Histones |
|-------------|------------|----------|
| ^a^Wild type | Scm384-169 | His6-KK-Cse4151-207-LVPRGS-H445-103 |
| ^a^II | Scm384-169 (Ile110Asp/Ile117Asn) | His6-KK-Cse4151-207-LVPRGS-H445-103 |
| ^a^VL-I | Scm384-169 (Val157Gly/Leu158Gly/Ile161Gly) | His6-KK-Cse4151-207-LVPRGS-H445-103 |
| ^a^MM | Scm384-169 | His6-KK-Cse4151-207 (Met181Ser/Met184Gly)-LVPRGS-H445-103 |
| ^a^AS | Scm384-169 | His6-KK-Cse4151-207 (Ala189Ser/Ser190Val)-LVPRGS-H445-103 |
| ^a^T-KDQ | Scm384-169 | His6-KK-Cse4151-207 (Δ(172-174)/Thr170Lys)-LVPRGS-H445-103 |
| ^a^VTDE | Scm384-169 | His6-KK-Cse4151-207 (Val165Ile/Thr166Ala/Asp167Gln/Glu168Asp)-LVPRGS-H445-103 |
| ^a^LY | Scm384-169 | His6-KK-Cse4151-207-LVPRGS-H445-103 (Leu98Gly/Tyr99Gly) |
| ^b^Cse4-H4 | MGSS-His6-SSGLVPKGS-H430-103 | |
| ^b^H3CATD-H4 | MGSS-His6-SSGLVPKGS-H430-103 | |
| ^b^H3LTT-H4 | MGSS-His6-SSGLVPKGS-H430-103 | |
| ^b^H3Mut4-H4 | MGSS-His6-SSGLVPKGS-H430-103 | |
| ^b^H3Mut5-H4 | MGSS-His6-SSGLVPKGS-H430-103 | |
| ^b^GS-Scm393-169-LVPR | His6-KK-Cse4152-207-LVPR, GS-H445-103 |
| ^b^GS-HJURP2-81-LVPR | His6-Cse4152-228-LVPR, GS-humanH430-103 |
| ^c^His6-KK-Cse4150-227-LVPRGS-Scm393-169-GDK-H442-103 (scSCH) | |
| ^c^His6-KK-Cse4150-227-LVPR, GS-Scm393-169-GDK-H442-103 | |
| ^c^His6-KK-Cse4150-227-LVPR, GS-Scm393-169-LVPR, GS-H442-103 | |
| ^c^His6-KK-Cse4152-228-LVPRGS-GGGS-Scm383-169-GDK-H442-103 | |
| ^c^Cse4150-229, H41-102 | |
| ^d^Scm384-169, His6-Cse4152-228, H41-102 | |

Note: samples with superscripts a, b, c, and d are for ITC, pull-down, NMR, and sedimentation experiments, respectively.
Supplementary Figure 1. The Cse4-H4-binding domain of Scm3 alone has a disordered structure. $^1$H-$^{15}$N HSQC spectrum of Scm3$_{56-187}$ at pH 6.0 and 22°C (100 mM NaCl and 25 mM MES). The dashed boxes include the backbone (left) and side chain (right) amides. The lack of dispersion of the chemical shifts in the $^1$H dimension is indicative of disordered proteins. Proteins with folded structures would have peaks outside the dashed boxes.
Supplementary Figure 2. The Scm3-Cse4-H4 complex is a hexamer in 2M NaCl but becomes a trimer in 0.5 NaCl. 

a, Gel filtration elution profiles for Scm3_56-187-Cse4_129-228 (WT)-H4 (major black peak) and Scm3_56-187-Cse4_129-228 (L204E)-H4 (major red peak) in the presence of 2.0 M NaCl. 

b, Gel filtration elution profiles of Scm3_56-187-Cse4_129-228 (WT)-H4 (black) and Scm3_56-187-Cse4_129-228 (L204E)-H4 (red) in the presence of 0.5 M NaCl. The experiments were performed at pH 7.4 and 4 °C. Here, the L204E mutation in Cse4_129-228 prevents the Cse4 (WT)-H4 dimer from forming the (Cse4_129-228 (WT)-H4)_2 tetramer. Scm3_56-187-Cse4_129-228 (L204E)-H4 serves as an internal marker in the gel filtration experiment at 0.5 M NaCl. The choice of the mutation is based on the earlier result that shows that the corresponding mutation in H3 prevents the H3-H4 dimer from forming the (H3-H4)_2 tetramer.\(^\text{17}\).
Supplementary Figure 3. Fragments of Scm3 and Cse4 that are explored for reconstituting Scm3-Cse4-H4 complexes. a, Scm3. NES: nuclear export signal; BR: basic residues; NLS: nuclear localization sequences; HR: heptad repeat units with hydrophobic residues occupying the fourth position and polar residues in the first positions. Note that Scm3.3 and Scm3.4 have additional residues (MGSSHHHHHHSSGLVPKGS) at the N-terminal region\textsuperscript{12}. b, Cse4. The histone fold domain is at the C-terminal region and binds to Scm3. The N-terminus to the histone fold domain (HFD) is not involved in the binding to Scm3.\textsuperscript{14} No stable complexes could be made using Scm3.5 and Scm3.6.
Supplementary Figure 4. The Scm3\textsubscript{84-169}-Cse4\textsubscript{150-229}-H4\textsubscript{1-103} complex forms a heterotrimer at pH 5.6 (50 mM MES) and 20.0°C. a, Sedimentation velocity c(s) profile for the Scm3\textsubscript{84-169}-Cse4\textsubscript{150-229}-H4\textsubscript{1-103} complex. The major species at 2.23 ± 0.01 S has an estimated molecular mass of 33.1 ± 0.2 kDa consistent with a 1:1:1 heterotrimer (M\textsubscript{calc} = 31.3842 kDa). b, Sedimentation equilibrium profiles with the sample recovered from sedimentation velocity experiments. Data were collected at 18 (orange), 22 (yellow), 26 (green), 30 (cyan) and 34 (brown) krpm and 20.0°C. These data, along with others collected at various concentrations, were analyzed globally in terms of a single ideal solute to obtain a molecular mass of 30.9 ± 0.7 kDa, corresponding to a complex having a 1:1:1 stoichiometry. The best fits are shown as black lines through the experimental points. c, The corresponding residuals of the best fits. The complex, however, is unstable beyond several hours at ≥ 30 °C, precluding structure determination by multi-dimensional NMR.
Supplementary Figure 5. The conformation of the Cse4_{150-229}-H4 dimer as characterized by NMR at 22 °C and design of scSCH. $\Delta\delta$ chemical shift differences between measured values and the corresponding random coil values$^{18}$ for the Cse4_{150-229}-H4 dimer. **a**, Cse4_{150-229}. **b**, H4. The dashed black lines indicate values of 1 and -1. A group of three or more consecutive residues with values larger than 1 indicate they form $\alpha$ helical structures. Residues with absolute values smaller than 1 indicate they have an irregular or disordered structure. The C-terminal region (~20 residues) of Cse4 and the N-terminal region (~50 residues) of H4 in the Cse4-H4 dimer are likely disordered since their NMR signals are very strong. The dashed magenta line illustrates the design of scSCH (M-His$_6$-KK-Cse4$_{150-227}$-LVPRGS-Scm3$_{93-169}$-GDK-H4$_{42-103}$), which involves the insertion of Scm3 between Cse4 and H4. scSCH is stable in solution for weeks at a pH of 5.4 and at 35 °C. An alternative single chain in the order of Cse4-H4-Scm3 did not lead to a stable and soluble protein. The length of the linkers is estimated based on the two considerations: (1) long enough not to alter the structures of the folded region; (2) not too long to cause significant overlap of the NMR peaks in the spectra.
Supplementary Figure 6. Verification of the structure of scSCH by analysis of mutation effects on chemical shifts. Single mutations were made and their $^1$H-$^{15}$N TROSY spectra were taken and compared with that of the wild type protein. The mutated residues are shown in balls in red. Residues with large chemical shift changes are mapped on the structure in red. The criteria for large chemical shift changes are 0.03 ppm for $^1$H or 0.3 ppm for $^{15}$N. For mutations that have very small perturbation, the criteria are 0.02 for $^1$H or 0.2 for $^{15}$N. Mutations from Val and Leu to Ala were chosen to avoid disruption of the structure. Cys mutants were made because the mutations are also used for other experiments. The fact that all of the residues with large chemical shift changes are the neighbors of the mutation sites indicates that the structure of scSCH is consistent with the results of mutations.
Supplementary Figure 7. Linkers in scSCH do not affect the structure of the folded region. a, Illustration of the cleavage of the first linker between Cse4 and Scm3 (arrow) and mapping of the residues that show significant chemical shift changes (red) on scSCH structure upon its cleavage (see b). b, Overlay of the $^1$H-$^{15}$N TROSY spectra of scSCH and the two-chain molecule (M-His6-KK-Cse4$_{150-227}$-LVPR, GS-Scm3$_{93-169}$-GDK-H4$_{42-103}$) with the first linker between Cse4 and Scm3 cut. Residues with recognizable changes in chemical shift outside the linker are labeled with their identities. c, The C-terminal region of the $\alpha$3 helix of H4 still keeps its helical conformation after the linker between Cse4 and Scm3 is cut by thrombin, as illustrated by the differences between measured C$\alpha$ chemical shifts and the corresponding random coil values for the residues at the C-terminus of the $\alpha$3 helix of H4. d, Illustration for the cutting of both linkers (arrows) and mapping of the residues with large chemical shift changes on the
structure of scSCH (see e). e, Overlay of the spectra between the two-chain molecules with the first linker between Cse4 and Scm3 cut and the three-chain molecules (M-His6-KK-Cse4_{150-227}-LVPR, GS-Scm3_{93-169}-LVPR, and GS-H4_{42-103}) with both linkers cut. Note that the second cut involves the engineering of a new thrombin-cleavage site by substituting the residues in the second linker region between Scm3 and H4 with the sequence for thrombin digestion (GDKGGV to LVPRGS). There are very small chemical shift changes for residues outside the second linker region.
Supplementary Figure 8. The structure of the three-chain complex obtained by cutting the two linkers in the scSCH remains the same after denaturation with 6M GdmCl and refolding. Overlay of the $^1$H-$^{15}$N TROSY spectra of the three-chain complex (M-His6-KK-Cse4$_{150-227}$-LVPR, GS-Scm3$_{93-169}$-LVPR, and GS-H4$_{42-103}$) obtained by cutting the two linkers with thrombin (black) and the one after denaturation with 6M GdmCl and refolding (red). The two spectra are identical, indicating that they have the same structure. Linkers do not control the fold of the protein complex. However, the linkers will increase the thermodynamic stability of the single chain molecule and prevent Scm3 from leaving Cse4 and H4. Disassociation of Scm3 from Cse4-H4 causes sample aggregation under our experimental conditions (pH 5.4 and 35 °C).
Supplementary Figure 9. Structure and dynamics of scSCH. a, Overlay of 20 calculated low energy structures of full-length scSCH. The structures are aligned on the folded core (Fig. 1e). The statistical parameters for the structures are listed in supplementary Table 1. b, Backbone amide $^{15}\text{N}-^{1}\text{H}$ NOEs. The cylinders represent helices and the lines represent loops. The residual helical structures in Cse4 that have large dynamic motions are named with primes. c-e, Front, bottom, and back views of the scSCH structure. Cse4 (cyan) and H4 (green) are presented as surface model and Scm3 (magenta) is in ribbon. The linkers are omitted.
Supplementary Figure 10. The region 145-154 in Scm3 is disordered. The measured $C_{\alpha}$ chemical shift values for the residues 145-154 are very close to the corresponding random coil values\textsuperscript{18}. In addition, this region has low $^{15}N$-$^1H$ NOE values (Fig. 9b).
| Residue | Chemical Shift |
|---------|----------------|
| Leu87   | -0.96          |
| Thr88   | -1.09          |
| Asp89   | 0.39           |
| Asp90   | 0.85           |
| Glu91   | 0.02           |
| Val92   | -0.08          |
| Met93   | -0.14          |
| Glu94   | -0.79          |

**Supplementary Figure 11. Extension of the N-terminal region of Scm3 does not lead to formation of additional helical structure of Scm3.** The theoretically predicted helical structure at the N-terminal region of Cse4-binding domain of Scm3 is from residue 92 to 114 (ref 19), which is longer than the one observed in scSCH (97-116). To investigate this issue, a new single chain construct was made that extends the N-terminal region of Scm3 to residue 83 (M-His6-KK-Cse4<sub>152-228</sub>-LVPRGSGGG-Scm3<sub>83-169</sub>-GDK-H42-103). We then cut the linker between Cse4 and Scm3 with thrombin (cutting site is between R and G in the linker) and assigned the Cα chemical shifts for the added residues. The measured Cα chemical shifts indicate that the added residues do not form helical structure. The left column shows the identity of the residues of Scm3. The right column shows the chemical shift deviations from random coil values. All of the values are smaller than 1, indicating that they do not form stable helical structures under our experimental conditions (pH 5.4 and 30 °C).
Supplementary Figure 12. Distribution of the residues that are specific to Cse4 on the structure of scSCH. a, Highlight of the residues specific to Cse4 (red) in the sequence of Cse4. b, Highlight of the residues specific to Cse4 (red) on the three-dimensional structure of scSCH. c, The side chain of Trp178 in loop 1 of Cse4 does not make direct contacts with Scm3.
Cse4 and H4 in scSCH. **a**, Interactions of the αN of Scm3 with the α3 of H4. **b**, Interactions of the loop of Scm3 with L1 loop of Cse4. **c**, Electrostatic interactions between the loop (118-132) of Scm3 to the histones (loop 1 of Cse4 and loop 2 of H4). Charged residues in the loop of Scm3 are shown as spheres. Positively charged residues are in blue and negatively charged residues are in red. Electrostatic potential on the histones surfaces calculated in vacuum are shown in blue for positive and red for negative, in the range from 85 to -85. **d**, The bulge in the loop of Scm3 sits on top of the L2 loop of H4. **e**, Interactions of loop 2 of Scm3 with the C-terminal region of the α2 helix of H4. **f**, Interactions of the αC helix of Scm3 with the N-terminal region of H4. A triple mutation (Val157Gly/Leu158Gly/Ile161Gly) in the αC helix of Scm3 decreased the binding affinity by a factor of 12 (Fig. 1a, d and Supplementary Table 2). Note that the hydrophobic interactions described here are confirmed by the observed ^1^H^-^1^H NOEs.
Supplementary Figure 14. Typical curves of isothermal titration calorimetry experiments. Titrations of Scm3\textsubscript{83-169} to His6-KK-Cse4\textsubscript{151-207}-LVPRGS-H\textsubscript{445-103} (solid black square) and their mutants: Scm3\textsubscript{83-169}(Ile110Asp/Ile117Asn) to His6-KK-Cse4\textsubscript{151-207}-LVPRGS-H\textsubscript{445-103} (magenta open circle), Scm3\textsubscript{83-169} to His6-KK-Cse4\textsubscript{151-207}(Ala189Ser/Ser190Val)-LVPRGS-H\textsubscript{445-103} (cyan open triangle), Scm3\textsubscript{83-169} to His6-KK-Cse4\textsubscript{151-207} LVPRGS-H\textsubscript{445-103}(Leu98Gly/Tyr99Gly) (green solid circle). The solid lines are the fitting curves (Supplementary methods).
Supplementary Figure 15. The CATD of Cse4 is sufficient for recognition by Scm3. SDS-PAGE gels that show that the CATD of Cse4 can be recognized by Scm3 in a pull-down experiment. Upper panel shows the input. Lower panel shows the components that were pulled down by His6-Scm3_{84-197} that is bound to the Ni^{2+}-coated beads. Lane 1: the loop 1 of H3 is replaced with the loop 1 of Cse4. Lane 2: the loop 1 and α2 helix of H3 is replaced with the CATD of Cse4. Lane 3: wld type Cse4. Here, wild type Cse4 is H3_{1-38}Cse4_{129-228}. H3 is H3_{1-135}. H4 is H4_{1-102}. L1 represents the residues \textsuperscript{167}TDEFTTKDQDLRWQ\textsuperscript{179} in the loop 1 of Cse4, which substitutes the corresponding residues of H3 in H3^L1.
Supplementary Figure 16. The CENP-A-binding domain of HJURP binds to Cse4 and human H4. SDS-PAGE gel that illustrates that the CENP-A-binding domain of human HJURP protein can bind to Cse4 and human H4 in the pull-down experiment. Lane 1, input (His6-Cse4_{152-228} (His6-Cse4_{152-228}-LVPR), HJURP_{2-81} (GS-HJURP_{2-81}-LVPR) and human H4 (GS-hH4_{30-103}); lane 2: the complex was incubated with Ni^{2+}-coated beads. They were washed with buffer and then eluted with 8 M urea that unfolds the protein complex; lane 3: same as lane 2 except eluted with 250 mM imidazole at the final step.
Supplementary Figure 17. Comparison of loop 1 of Cse4 and loop 2 of H4 in scSCH with the corresponding regions in the (CENP-A-H4)$_2$ tetramer. Cse4-H4 in scSCH and CENP-A-H4 in the (CENP-A-H4)$_2$ tetramer were overlaid by aligning their $\alpha_2$ helix, loop 2, and $\alpha_3$ helix in H4. Loop 1 of Cse4 is pushed away from the loop 2 of H4 by the loop of Scm3. Loop 2 of H4 in the two structures shows only minor conformational changes.
Supplementary Fig. 18. Different conformations of the C-terminal region of H4 in different complexes. **a**, H3-H4 in the nucleosome. The β-strand in the C-terminal region of H2A pairs with the β-strand in the C-terminal region of H4 in the nucleosome structure. **b**, H3-H4 in the Asf1-H3-H4 complex. The β-strand in the C-terminus of H4 pairs with the β-strand at the C-terminus of Asf1.
Supplementary Figure 19. The α3 helix and the C-terminal region of the α2 helix of Cse4 are not required for Scm3 binding. The SDS-PAGE gel of the pull-down experiment. Lane 1, input. It includes His6-Cse4152-207(M-His6-KK-Cse4152-207-LVPR), Scm393-169(GS-Scm393-169-LVPR), and H445-103 (GS-H445-103) that were obtained from a single chain M-His6-KK-Cse4152-207-LVPRGS-Scm393-169-LVPRGS-H445-103 by cutting the two linkers with thrombin; Lane 2: Ni^{2+}-coated beads were incubated with the three-chain complex. They were then washed with buffer and eluted with 8 M urea; Lane 3: same as lane 2 except eluted with 250 mM imidazole at the final step.
Supplementary Figure 20. Modeling of the \((\text{Scm3}-\text{Cse4}-\text{H4})_2\) hexamer and its incompatibility with binding of DNA. a, A model of the \((\text{Scm3}-\text{Cse4}-\text{H4})_2\) hexamer. The model is made by aligning residues \(186\text{EASEAYLVGLEHT}201\) in the second half of the \(\alpha2\) helix of Cse4 in scSCH with the corresponding residues \(96\text{EAAEFLVHLFEDA}109\) of CENP-A in the \((\text{CENP-A}-\text{H4})_2\) tetramer. The conformation of the residues C-terminal to residue 109 in CENP-A and N-
terminal to residue 202 in Cse4 were kept. The DNA is modeled using the histones in the dyad region based on the canonical nucleosome. A large conformational change occurs in Cse4 and H4 when compared with CENP-A and H4 in the (CENP-A-H4)$_2$ tetramer as in b. Overall, only a small region of Cse4 at the dyad may bind to DNA. Other regions in histones either do not contact with DNA or clash with DNA. b, The (CENP-A-H4)$_2$ tetramer is capable of binding much longer DNA, as modeled based on canonical nucleosome structure. c, The dynamic C-terminal region of the CBD (143-161) in scSCH has minor structural overlaps with the $\alpha_1$ helices of CENP-A and H4 when the H4 in scSCH is aligned with the H4 in human (CENP-A-H4)$_2$ tetramer. The residues 49-91 of H4 in scSCH are aligned with the corresponding residues of H4 of CENP-A-H4 in the (CENP-A-H4)$_2$ tetramer. Shown here are human CENP-A-H4 and the C-terminal region of the CBD of Scm3 (137-161). Structural overlaps occur between the N-terminal region of $\alpha_1$ of CENP-A and loop region of Scm3 and between the $\alpha_1$ helix of H4 in CENP-A-H4 and the $\alpha_C$ helix of Scm3 in scSCH.
Supplementary Figure 21. Major structural features of known histone chaperone-histone complexes. a, Chz1-H2A.Z-H2B. b, scSCH. c, Asf1-H3-H4. Chz1 blocks DNA-binding site on H2A.Z-H2B. Scm3 blocks DNA-binding site and induces large conformation changes in both Cse4 and H4 histones. Asf1 blocks the interface for tetramerization of H3-H4 dimer and induces the β-strand at the C-terminal region of H4 to extend the β-sheet in Asf1.
Supplementary References

1. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* 1, 749-754 (2006).

2. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity untracentrifugation and Lamm equation modeling. *Biophys. J.* 78, 1606-1619 (2000).

3. Cole, J. L., Lary, J. W., Moody, T. P. & Laue, T. M. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol.* 84, 143-179 (2008).

4. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* 320, 104-124 (2003).

5. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277-293 (1995).

6. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* 4, 603-614 (1994).

7. Schwiers, C. D., Kuszewski, J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160, 65-73 (2003).

8. Zhou, Z. et al. NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nat. Struct. Mol. Biol.* 15, 868-869 (2008).

9. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13, 289-302 (1999).

10. Laskowski, R. A. et al. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477-486 (1996).

11. Houtman, J. C. et al. Binding specificity of multiprotein signaling complexes is determined by both cooperative interactions and affinity preferences. *Biochemistry* 43, 4170-4178 (2004).

12. Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M. M. & Wu, C. Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell* 129, 1153-1164 (2007).
13. Morey, L., Barnes, K., Chen, Y., Fitzgerald-Hayes, M. & Baker, R. E. The histone fold domain of Cse4 is sufficient for CEN targeting and propagation of active centromeres in budding yeast. *Eukaryot. Cell* **3**, 1533-1543 (2004).

14. Stoler, S. *et al.* Scm3, an essential *Saccharomyces cerevisiae* centromere protein required for G2/M progression and Cse4 localization. *Proc. Natl. Acad. Sci. U. S. A* **104**, 10571-10576 (2007).

15. Natsume, R. *et al.* Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature* **446**, 338-341 (2007).

16. English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E. & Tyler, J. K. Structural basis for the histone chaperone activity of Asf1. *Cell* **127**, 495-508 (2006).

17. Banks, D. D. & Gloss, L. M. Folding mechanism of the (H3-H4)2 histone tetramer of the core nucleosome. *Protein Sci.* **13**, 1304-1316 (2004).

18. Schwarzinger, S. *et al.* Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**, 2970-2978 (2001).

19. Sanchez-Pulido, L., Pidoux, A. L., Pointing, C. P. & Allshire, R. C. Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell* **137**, 1173-1174 (2009).