Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units

Feng Song *et al.*

Science **344**, 376 (2014);
DOI: 10.1126/science.1251413

Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units

Feng Song,[1,2]* Ping Chen,[1]* Dapeng Sun,[1,2] Mingzhu Wang,[1] Liping Dong,[1,2] Dan Liang,[1,2] Rui-Ming Xu,[1] Ping Zhu,[1]† Guohong Li[1]†

The hierarchical packaging of eukaryotic chromatin plays a central role in transcriptional regulation and other DNA-related biological processes. Here, we report the 11-angstrom–resolution cryogenic electron microscopy (cryo-EM) structures of 30-nanometer chromatin fibers reconstituted in the presence of linker histone H1 and with different nucleosome repeat lengths. The structures show a histone H1-dependent left-handed twist of the repeating tetranucleosomal structural units, within which the four nucleosomes zigzag back and forth with a straight linker DNA. The asymmetric binding and the location of histone H1 in chromatin play a role in the formation of the 30-nanometer fiber. Our results provide mechanistic insights into how nucleosomes compact into higher-order chromatin fibers.

Understanding the structure of chromatin is key to illuminating the functions of chromatin dynamics in epigenetic regulation of gene expression. The structure of the native 30-nm chromatin fiber in nuclei or isolated from nuclei is a regular helix of nucleosomes with a diameter of about 30 nm and a packing density of about 6 to 7 nucleosomes per 11 nm (*1*–*7*). Nucleosomes can be arranged either linearly in a one-start solenoid-type helix with bent linker DNA or zigzag back and forth in a two-start stack of nucleosomes connected by a relatively straight DNA linker (*7*–*9*). The latter class can be further divided into the helical ribbon model and the twisted crossed-linker model by the different orientation angles between the linker DNA and fiber axes (*8*–*10*). The manner by which nucleosome core particles (NCPs) interact with each other in a beads-on-a-string nucleosomal array to form a condensed 30-nm chromatin fiber remains unresolved (*11*). The irregular native chromatin fiber cannot readily form a homogeneous structure suitable for high-resolution structural analyses. The problem has been partially addressed by reconstituting chromatin fibers in vitro on regular tandem repeats of unique nucleosome-positioning DNA sequences with purified histone proteins (*12*, *13*). Fibers reconstituted in the presence of histone H5 with different nucleosome repeat lengths (NRLs) showed similar structures and could fit the one-start interdigitated solenoid structure (*14*). Chromatin fiber with two-start zigzag conformation was also observed on long reconstituted nucleosome arrays (*15*). The x-ray crystal structure of a tetranucleosome with a 20–base pair (bp) linker DNA and without linker histones reveals two stacks of nucleosomes connected by straight linker DNA (*16*); however, a tetranucleosomal array is too short to form a solenoid structure. Linker histones, which are present at close to one molecule per nucleosome in the majority of eukaryotic organisms, have been considered to be essential for 30-nm chromatin fiber formation (*17*–*20*), but their precise location and exact roles in the organization of the higher-order structure still remain to be determined.

Cryo-EM Reconstruction of 30-nm Chromatin Fibers

We have determined, at about 11 Å resolution, a three-dimensional (3D) cryogenic electron microscopy (cryo-EM) structure of 30-nm chromatin fibers reconstituted in vitro on the 12 tandem repeats of 187-bp and 177-bp (12 × 187 bp and 12 × 177 bp) 601 DNA sequence with recombinant *Xenopus laevis* canonical histones lacking post-translational modifications (Fig. 1). The reconstituted 30-nm fibers are in a compact form in the presence of histone H1 with stoichiometry about one H1 per nucleosome under low-salt conditions (fig. S1). The reconstituted chromatin fibers were fixed by 0.2% glutaraldehyde before cryo-EM analysis and displayed as homogeneously compacted particles in the representative field views of the cryo-EM images (Fig. 1A). About 31,000 particles of 12 × 187 bp and 25,000 particles of 12 × 177 bp 30-nm fibers were visually screened and subjected to 2D classification and 3D reconstruction, beginning with an initial model of a featureless Gaussian blob (fig. S2A). The Euler angle distribution indicates that the particles used in our 3D reconstruction have no preferred orientation (fig. S2B). Some selected unsupervised 2D classification averages (Fig. 1A, right) agree well with the raw particles, as indicated by the white box in the micrograph (Fig. 1A, left). The 3D cryo-EM map defines the spatial location of all individual nucleosomes and the path of linker DNA in the 30-nm fiber (Fig. 1, B and C, and movie S1). The overall structure of the dodecanucleosomal 30-nm fiber comprises three tetranucleosomal structural units, which are twisted against each other with linker DNA extended straight to form a two-start helix (Fig. 1, B and D). This disposition can also be deduced from the particularly oriented reference-free average, in which obvious densities connect two adjacent nucleosomes (Fig. 1A, right, white dashed brackets indicated). The four nucleosomes within the structural unit zigzag back and forth to form two stacks of two nucleosome cores (Fig. 1D), which is consistent with the previous observations that deoxyribonuclease I digestion of nuclei produces dinucleosomal periodicity patterns (*21*). According to the previously proposed zigzag two-start helix model, the diameter of the chromatin fiber could be increased accordingly with the length of DNA crossing back and forth between the two-start helix (*8*–*10*). To analyze how the variations in NRLs affect the overall structure of the resulting fibers, we compared the 3D cryo-EM structures of 30-nm fibers reconstituted with two different NRLs, that is, 12 × 187 bp and 12 × 177 bp 601 DNA sequences (Fig. 1C and fig. S2). The two 30-nm fibers with 177-bp NRL and 187-bp NRL display a very similar overall stacking mode of nucleosomes with the connected linker DNAs extended and straight. An increase of 10 bp of NRL does not affect the overall structure and organization of the reconstituted chromatin fiber, but the increase does change the fiber dimension (diameter × height) from about 27.2 × 28.7 nm for 177-bp repeats to about 29.9 × 27.0 nm for 187-bp repeats (Fig. 1C), which is consistent with a basic zigzag two-start helix model. Overall, our cryo-EM structure shows that the 30-nm chromatin fiber follows a path that is basically compatible with a zigzag two-start helix (Fig. 1D), although the fine details of the structure are distinct from the originally proposed model.

Tetranucleosomal Unit with a Two-Start Zigzag Conformation

Within the tetranucleosomal structural unit of the 30-nm fiber, two stacks of two nucleosome cores are connected by straight linker DNA (Fig. 2A and fig. S3C). The two nucleosomes in each stack directly contact head to head through their octamer surfaces. For the structural unit with a 187-bp NRL, the two stacked nucleosome cores are separated center to center by 53.6 Å with each superhelical axis and dyad axis angled at 11.8° and 16.5°, respectively (Fig. 2A, left). In the interface between the cores, a bulk density is present at the junction of the adjacent H2A-H2B dimer, indicating a strong interaction between the H2B-helix α1/αC and the adjacent H2A-helix α2 (Fig. 2B). As described in the x-ray structure of the tetranucleosome, this strong interacting interface does not allow the internucleosomal interaction between the positive N terminus of histone H4

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China.

*These authors contributed equally to this work.
†Corresponding author. E-mail: liguohong@ibp.ac.cn (G.L.); zhup@ibp.ac.cn (P.Z.)

(amino acids 16 to 26) and the acidic patch of the H2A-H2B dimer observed in the nucleosome core-particle structure (*16*). The H2B-α3/αC internucleosomal four-helix bundle might be involved in the compaction and stabilization of chromatin fiber (*22*), which was not observed in our cryo-EM map of reconstituted chromatin fiber, presumably because of the interaction of DNA involved in chromatin reconstitutions.

The two dinucleosomal stacks are separated by 196 Å center to center and are rotated 54.5° left-handed with respect to their stack axes (the axis passing through the centers of the two nucleosomes in each stack) for the 187-bp NRL (Fig. 2A, right). We docked the x-ray atomic structure of the 167-bp NRL in the absence of linker histone (*16*) into the EM density map of our tetranucleosomal structural units with 187- and 177-bp NRLs, respectively (Fig. 2B and fig. S3C). The structures for each stack of two nucleosomes fit very well with our cryo-EM map in both cases. However, the distance and rotation between the two stacks are quite different, which may be caused by the presence of histone H1 or the different NRLs used. The two stacks are separated by 146.1 Å for the tetranucleosome with the 167-bp NRL used for x-ray investigation, compared with 167 Å for the 177-bp NRL (fig. S3C) and 196 Å for the 187-bp NRL (Fig. 2A) in this study, confirming that the fiber diameters change accordingly to the lengths of the NRLs. In addition, the rotation between the two stacks in the x-ray structure is left-handed by 71.3°, compared with 63.7° for the 177-bp NRL and 54.5° for the 187-bp NRL in our cryo-EM structure (Fig. 2 and fig. S3). If we only consider the intrinsic property of the difference in NRLs, each 10-bp increment in linker DNA should increase the length of DNA by 3.4 nm and change the twist between adjacent nucleosomes by −17 based on the DNA twist of 10.5 bp per turn, suggesting that the presence of H1 may impose additional effects on the specific distance and the rotation of the two stacks in our cryo-EM structure (Fig. 3A). The defined location of H1 in each nucleosome, as indicated in Fig. 3B, reveals that the direct interaction of H1 with both the dyad and the entering and exiting DNA in the NCPs may alter the angle of the entry-exit DNA and constrain the linker length and rotation angle between the stacks.

Twist and Interactions Between Tetranucleosomal Units

For the dodecanucleosomal 30-nm chromatin fiber, we define three structural units as unit 1 for nucleosomes N1 to N4, unit 2 for N5 to N8, and unit 3 for N9 to N12. The rotation angles and the separation distances between units 1 and 2 and between units 2 and 3 are slightly different, which may be a result of the end effect of the dodecanucleosomal chromatin fiber. In the cryo-EM structure of the 30-nm fiber with the 187-bp NRL, units 1 and 2 are related by a 48.9° rotation around the fiber axis and a 72.2 Å translation along the axis, whereas a 52.9° rotation and a 68.8 Å translation were observed for units 2 and 3. Here, the fiber axis is defined as the axis bisecting the angle between the two stack axes and orthogonally intersecting the axis connecting the centers of the two stacks in each structural unit. The internucleosomal interactions between the structural units are different from that observed within the tetranucleosomal units. The internucleosomal interface between the structural units displays relatively strong density where the N terminus of H4 meets the adjacent H2A-H2B dimer (Fig. 3C), which indicates that the internucleosomal interactions between the positively charged residues of the H4 N-terminal tail (residues 16 to 23) and the acidic patch of the H2A-H2B dimer may account for the twist between the tetranucleosomal units. The interactions within this region were not observed in the previous x-ray tetranucleosome structure (*16*), but these interactions have been reported to be crucial for the formation of the 30-nm chromatin fibers (*15*, *23*). We generated a series of single mutants in the H4 N-terminal tail, including H4 Lys^{16}→Ala^{16} (K16A), H4R17A (R, Arg), H4R19A, H4K20A, and H4R23A, to examine the functions of these residues in chromatin folding in the presence of histone H1 (fig. S4). Sedimentation velocity in conjunction with van Holde–Weischet analysis was used to identify the structural changes in the chromatin fibers. The 30-nm chromatin fibers containing H4K16A, H4R17A, H4R19A, or H4K20A showed very similar distributions of sedimentation coefficients in comparison with those of the wild-type fibers with a S_{ave} of 46.6S ± 1.7S (where S_{ave} is defined as the sedimentation coefficient at the boundary fraction equal to 50%). However, the chromatin

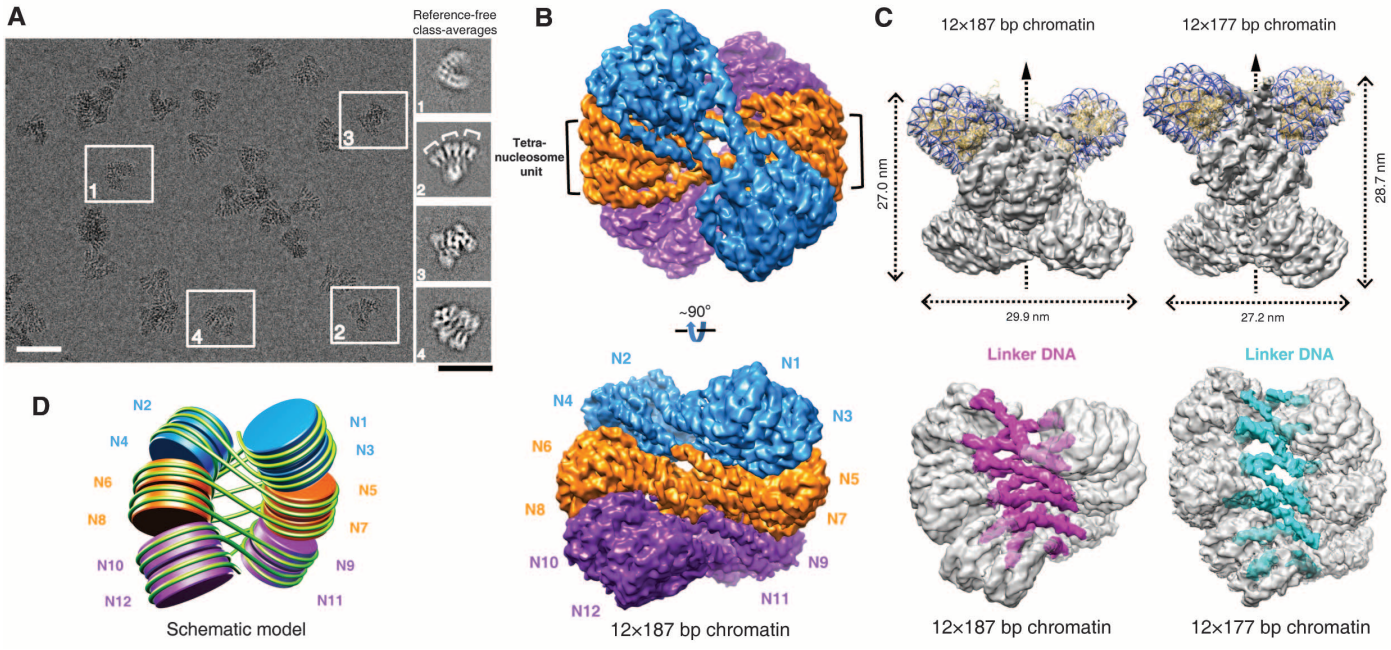

Fig. 1. Cryo-EM reconstruction of 30-nm chromatin fibers. (**A**) A representative cryo-EM micrograph of 30-nm chromatin fibers reconstituted on 12 × 187 bp DNA. Scale bars indicate 50 nm. Four selected unsupervised classification generated averages are shown in the right images, which are in good agreement with the raw particles indicated by white boxes in the micrograph. (**B**) The overall 3D cryo-EM map of the 30-nm chromatin fibers reconstituted on 12 × 187 bp DNA with the three tetranucleosomal structural units highlighted by different colors and viewed from two angles. (**C**) A comparison of the overall structure of 30-nm chromatin fibers reconstituted on 12 × 187 bp and 12 × 177 bp DNA, viewed from two angles with the fiber dimensions directly labeled and their straight linker DNA highlighted. (**D**) A schematic representation of the cryo-EM structure of 30-nm chromatin fibers as shown in (B).

fibers containing H4R23A shifted the sedimentation coefficient distribution to a S_{ave} of 50.8*S* (fig. S4, B and C), which indicated that the chromatin containing H4R23A could still fold into a compact structure but that the mutant H4R23A might alter the chromatin folding mode. Cryo-EM images and the corresponding reference-free class averages of the H4R23A 30-nm fibers showed that the particles appeared much more frequently as two parallel stacks of nucleosomes than the wild-type fibers and with more blurred averages (Fig. 3, D and E). The results suggest that the mutant H4R23A changes the internucleosomal interactions between the structural units and makes it difficult for the tetranucleosomal units to twist stably against each other to form a helical structure.

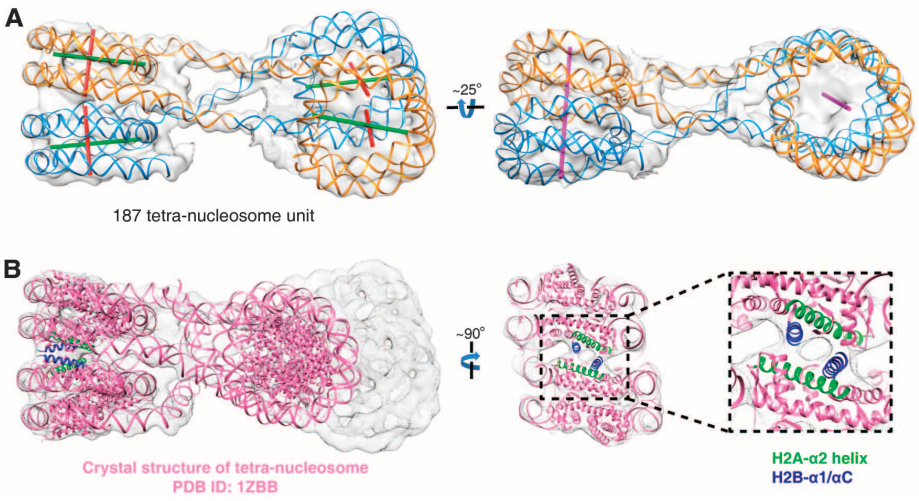

Fig. 2. The structure of a tetranucleosomal unit with the 187-bp NRL. (**A**) The segmented density map for the tetranucleosomal unit in the 30-nm chromatin fibers reconstituted on 12 × 187 bp DNA, shown with the atomic structure of DNA from a docked mononucleosome crystal structure (PDB 1AOI) and modeled presumptive linker DNA and viewed from two angles. Different axes are highlighted by colors, including nucleosome core dyad (green), nucleosome superhelix axes (red), and stack axes (pink). (**B**) A comparison of the 3D cryo-EM map (gray) with the x-ray structure (PDB 1ZBB, pink) of the tetranucleosome (*16*). The strong density where the adjacent H2A-H2B dimer meets is magnified and highlighted in the interface between the nucleosome cores within each stack.

The Asymmetric Location and Self-Association of Histone H1 in Chromatin Fiber

The incorporation of histone H1 plays a fundamental role in determining the higher-order structure of the 30-nm chromatin fiber. As shown in Fig. 3A and figs. S1 and S5A, histone H1 in the 30-nm fibers exhibit a proper 1:1 stoichiometric association with the nucleosome cores. The well-defined locations of the 12 histone H1 molecules in the dodecanucleosomal 30-nm fiber can be visualized in our cryo-EM structure (Fig. 3A, fig. S5A, and movie S2). In each nucleosome core, histone H1 directly interacts with both the dyad and the entering and exiting linker DNA (Fig. 3B), which determines the trajectory of the entry or exit linker DNA segments in nucleosomes and stabilizes the fiber. This interaction mode of histone H1 with the nucleosome core has been proposed previously (*20*, *24*–*26*). In addition, an

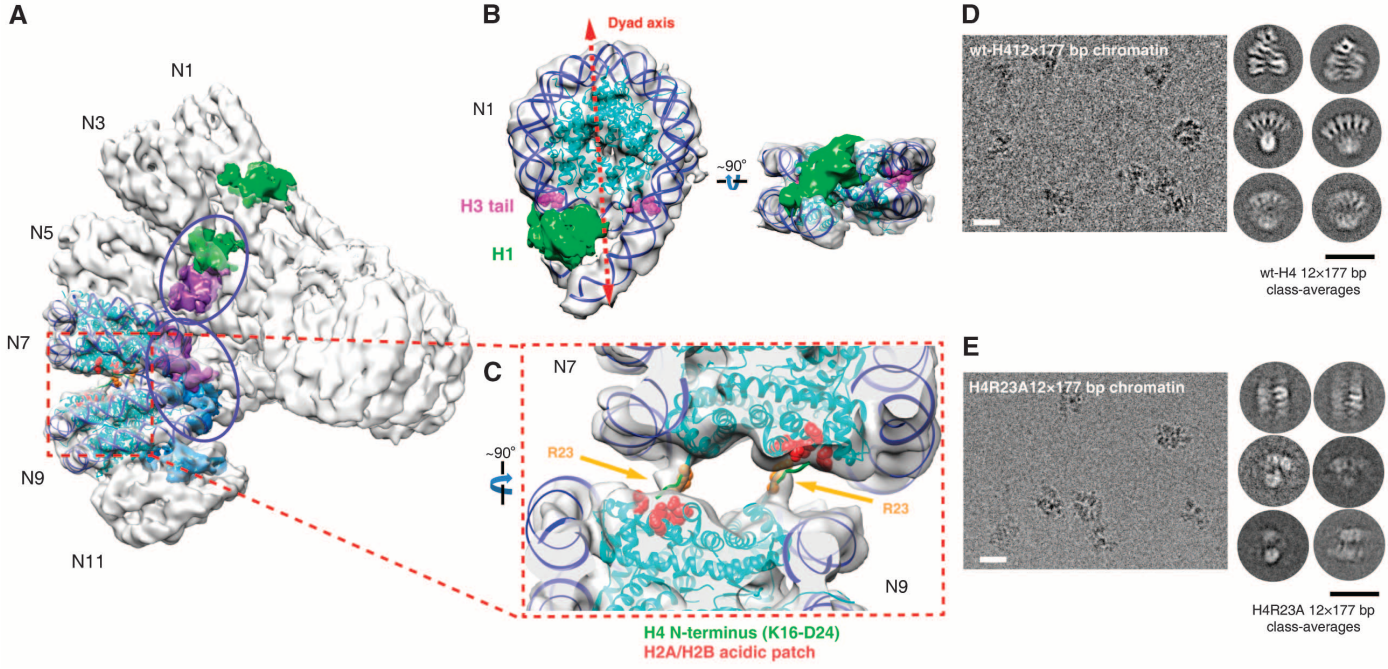

Fig. 3. The interactions between the tetranucleosomal units within the 30-nm fibers. (**A**) The tetranucleosomal units are twisted against each other to form 30-nm fibers. The locations of H1 are highlighted by colors. (**B**) The asymmetric location of H1 in the nucleosome core as viewed from two angles. The mononucleosome N1 segmented from the 12 × 187 bp map (A) with linker DNA density at both entry and exit regions is fitted with an atomic structure of the mononucleosome with an arbitrary length of linker DNA extracted from the tetranucleosome structure (PDB 1ZBB). The presumptive histone H1 is highlighted in green. The H3 tails in the x-ray structure are colored in magenta to locate their relative positions to H1. (**C**) Detail of the strong density where the N terminus of H4 meets the adjacent H2A-H2B dimer. (**D** and **E**) A comparison of the cryo-EM images of the particles and their related unsupervised classification generated averages for 30-nm chromatin fibers reconstituted with wild-type (wt)–H4 (D) and H4R23A (E), respectively. A subset of the wt-H4 12 × 177 bp data with the same particle number, 2664, as that of the H4R23A 12 × 177 data set were randomly picked from the entire data set and subjected to an independent unsupervised 2D classification for comparison. Scale bars, 30 nm.

apparent off-axis location of the globular domain of histone H1 not only in a mononucleosome (Fig. 3B) but also in a tetranucleosomal unit (Fig. 3A, fig. S5A, and movie S2) can be observed in our cryo-EM structure, and this domain localization plays a critical role in the formation of a twist between tetranucleosomal units. The asymmetric location of H1 in mononucleosomes results in the discrimination between different sides of a nucleosome, similar to a coin; we define the "head" as the side of nucleosome with the small portion of H1 and the "tail" as the side with the large portion of H1. The head-to-head nucleosomal interaction in each stack within each tetranucleosomal unit only permits the tail-to-tail nucleosomal interactions between tetranucleosomal units for further stacking and twisting of the chromatin fiber (Fig. 3A). The self-association of H1 had been reported to play an important role in the organization and stabilization of the 30-nm chromatin fiber (*27*). Our cryo-EM structure also revealed a previously unknown arrangement of histone H1 molecules in the reconstituted 30-nm chromatin fiber, in which only the large portions of H1 (most likely its global domain) on the tail side of nucleosomes can interact directly with each other and impart an additional twist between each structural unit (Fig. 3A and fig. S5A). The specific asymmetric binding and location of histone H1 in both mononucleosome and tetranucleosomal units determine the formation of the double-helical 30-nm fiber with a spiral twist of tetranucleosomal units. To further explore the interactions between histone H1 and nucleosomal DNAs, we extracted the densities of individual mononucleosomes with the presumptive H1 from the reconstructed map of the dodecanucleosomal chromatin fiber and averaged them on a 3D level (for details, see materials and methods). Guided by a rigid-body fitting, we found that the crystal structure of *Gallus gallus* histone H5 globular domain (gH5) [Protein Data Bank (PDB) code 1HST] can be docked well into a region, which is a part of the presumptive H1 density, in the averaged H1-containing mononucleosomal EM density map (fig. S5B). This region, most likely the globular domain of human histone H1.4 used in this study, interacts with the nucleosomal entry, exit, and dyad DNAs in a three-contact mode (fig. S5B, red dots), in agreement with the computational analysis of gH5 (*28*) and the mapping of histone H1.5-nucleosome interactions (*24*). Extra densities that cannot be assigned to the globular domain of H1 are presumably contributed by the N- and C-termini of H1 (fig. S5B, blue lines and black stars indicated) and/or partially contributed by the N-termini of histone H3 (Fig. 3B and fig. S5B, magenta tail), which was previously shown to contribute to chromatin folding (*29*).

Structural Model for Chromatin Fibers

Our 3D cryo-EM structure of the 30-nm chromatin fiber allows construction of a fine structural model for 30-nm fibers in the presence of histone H1 (Fig. 4). To evaluate the possible end effects of the relatively short 12 tandem repeats of nucleosomes, we reconstituted the 30-nm fibers with 24 repeats of 177-bp (24 × 177 bp) 601 DNA template and acquired the 3D cryo-EM structure at ~25 Å resolution (Fig. 4A, fig. S6, and movie S3). Two copies of 12 × 177 bp chromatin density map can be docked into the map of 24 × 177 bp chromatin without further modification (Fig. 4B). In addition, the interactions between the intra- and interunits in the 24 × 177 bp chromatin are well maintained compared with the 12 × 177 bp chromatin model (fig. S6E). We then built the model of 30-nm chromatin by directly stacking the cryo-EM structure of dodecanucleosomal 30-nm fibers on top of each other to form a continuous fiber. The resulting direct model reveals that the chromatin fiber is exhibited as a left-handed double helical structure twisted by tetranucleosomal units (Fig. 4C). The 30-nm fiber with the 187-bp NRL contains 7.07 tetranucleosomal units (28.3 nucleosomes) per turn in a period of 49.9 nm, yielding a nucleosome packing density of about 6.2 nucleosomes per 11 nm; this density is very similar to the measured values for mass per unit length of the chromatin fiber (*4*, *30*). To examine the effects of the previously indicated slight difference in rotation and separation between different units in the 30-nm fiber of dodecanucleosomes in the model construction, we also built two other models with a repeat stacking of units 1 and 2 or units 2 and 3, respectively (fig. S7, A and B). These two alternative models show a high resilience to the double-helical twist model formed from the direct stacking of dodecanucleosomes, except for slight changes in the packing density of nucleosomes that were identified as 6.1 nucleosomes per 11 nm for the model built by units 1 and 2 and 6.4 nucleosomes per 11 nm for the model built by units 2 and 3 (Fig. 4C and fig. S7).

Conclusions

The existence of 30-nm fibers in nuclei still remains to be elucidated, whereas the structure of chromatin fiber must be variable in vivo because of its highly dynamic and heterogeneous property with intrinsic compositions, such as different NRLs, linker histones, histone and DNA modifications, and histone variants. Although our 3D cryo-EM structures for the reconstituted 30-nm chromatin fiber show basically a two-start zigzag configuration for nucleosome arrangement, other forms of chromatin structures may exist in different conditions, for example, the one-start solenoid structure in the presence of H5 and magnesium with longer NRLs as suggested previously (*14*, *31*). Nevertheless, the formation of the double-helical structure of the reconstituted canonical 30-nm fibers using the 177-bp and 187-bp NRLs in the presence of H1 under a low-salt condition is basically driven by their intrinsic biophysical and biochemical properties; thus, the fundamental principles may also be applicable in vivo. Histone modifications and histone variants may also play important roles in the regulation of higher-order chromatin structure via modulating the internucleosomal surface interactions between tetranucleosomal units.

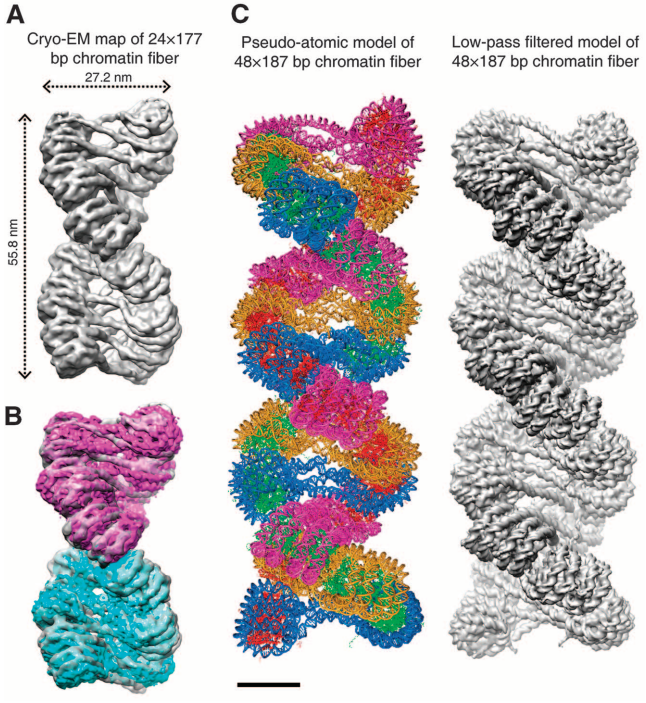

Fig. 4. The 30-nm chromatin fiber model. (A) The overall 3D cryo-EM map of the 30-nm chromatin fiber reconstituted on 24 × 177 bp 601 DNA, with the length and diameter of the fiber indicated. (**B**) The structure of 24 × 177 bp 30-nm fibers is docked by two copies of the cryo-EM structure of 12 × 177 bp 30-nm fibers. The fitting was optimized by the reported correlation value in UCSF Chimera. (**C**) A pseudo-atomic model (left, structure of H1 is not included) and its corresponding density map low-pass filtered to 11 Å (right), built by directly stacking the cryo-EM structure of the dodecanucleosomal 30-nm fiber with 187-bp NRLs on top of each other to form a continuous fiber. Scale bar, 11 nm.

References and Notes

1. R. Ghirlando, G. Felsenfeld, *J. Mol. Biol.* **376**, 1417–1425 (2008).
2. E. C. Pearson, P. J. Butler, J. O. Thomas, *EMBO J.* **2**, 1367–1372 (1983).
3. V. Graziano, S. E. Gerchman, D. K. Schneider, V. Ramakrishnan, *Nature* **368**, 351–354 (1994).
4. S. E. Gerchman, V. Ramakrishnan, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7802–7806 (1987).

5. J. Widom, A. Klug, *Cell* **43**, 207–213 (1985).
6. J. P. Langmore, J. R. Paulson, *J. Cell Biol.* **96**, 1120–1131 (1983).
7. J. T. Finch, A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1897–1901 (1976).
8. C. L. Woodcock, L. L. Frado, J. B. Rattner, *J. Cell Biol.* **99**, 42–52 (1984).
9. S. P. Williams *et al.*, *Biophys. J.* **49**, 233–248 (1986).
10. M. F. Smith, B. D. Athey, S. P. Williams, J. P. Langmore, *J. Cell Biol.* **110**, 245–254 (1990).
11. H. G. Davies, J. V. Small, *Nature* **217**, 1122–1125 (1968).
12. L. M. Carruthers, C. Tse, K. P. Walker 3rd, J. C. Hansen, *Methods Enzymol.* **304**, 19–35 (1999).
13. P. T. Lowary, J. Widom, *J. Mol. Biol.* **276**, 19–42 (1998).
14. P. J. Robinson, L. Fairall, V. A. Huynh, D. Rhodes, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6506–6511 (2006).
15. B. Dorigo *et al.*, *Science* **306**, 1571–1573 (2004).
16. T. Schalch, S. Duda, D. F. Sargent, T. J. Richmond, *Nature* **436**, 138–141 (2005).
17. F. Thoma, T. Koller, A. Klug, *J. Cell Biol.* **83**, 403–427 (1979).
18. J. O. Thomas, *Curr. Opin. Cell Biol.* **11**, 312–317 (1999).
19. D. L. Bates, J. O. Thomas, *Nucleic Acids Res.* **9**, 5883–5894 (1981).
20. J. Allan, P. G. Hartman, C. Crane-Robinson, F. X. Aviles, *Nature* **288**, 675–679 (1980).
21. D. Z. Staynov, *Bioessays* **30**, 1003–1009 (2008).
22. T. D. Frouws, H. G. Patterton, B. T. Sewell, *Biophys. J.* **96**, 3363–3371 (2009).
23. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **389**, 251–260 (1997).
24. S. H. Syed *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9620–9625 (2010).
25. D. Z. Staynov, C. Crane-Robinson, *EMBO J.* **7**, 3685–3691 (1988).
26. B. R. Zhou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19390–19395 (2013).
27. G. J. Carter, K. van Holde, *Biochemistry* **37**, 12477–12488 (1998).
28. L. Fan, V. A. Roberts, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8384–8389 (2006).
29. S. H. Leuba, C. Bustamante, K. van Holde, J. Zlatanova, *Biophys. J.* **74**, 2830–2839 (1998).
30. J. Bednar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14173–14178 (1998).
31. A. Routh, S. Sandin, D. Rhodes, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8872–8877 (2008).

Acknowledgments: This work was supported by grants from the National Basic Research Program of China (2010CB912400 to P.Z., 2011CB966300 to G.L., and 2009CB825500 to R.M.X. and P.Z.); the National Natural Science Foundation of China (91219202 to G.L., 31230018 to P.Z., 91019007 to G.L., 21261130090 to P.Z., and 31000566 to P.C.); Strategic Priority Research Program (XDA01010304 to G.L. and XDB08010100 to P.Z. and R.M.X.) and Key Research Program (KJZD-EW-L05 to P.Z., G.L., and R.M.X.) from the Chinese Academy of Sciences; and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, to P.C. All EM data were collected and processed at the Center for Bio-imaging, Institute of Biophysics, Chinese Academy of Sciences. We thank G. Ji and X. Huang for their technical help and support with electron microscopy and L. Ling for technical help and support with the data processing in the High Performance Computing Service Station. We are also indebted to the colleagues whose work could not be cited because of the limitation of space. The cryo-EM maps for the 12 × 177 bp, 12 × 187 bp and 24 × 177 bp chromatin fibers were deposited into the Electron Microscopy Data Bank with the accession codes EMD-2600, EMD-2601 and EMD-2602, respectively. The authors declare no conflicts of interest.

Supplementary Materials

www.sciencemag.org/content/344/6182/376/suppl/DC1
Materials and Methods
Figs. S1 to S8
References (*32*–*41*)
Movies S1 to S3

28 January 2014; accepted 18 March 2014
10.1126/science.1251413

Genome Sequence of the Tsetse Fly (*Glossina morsitans*): Vector of African Trypanosomiasis

International *Glossina* Genome Initiative*†

Tsetse flies are the sole vectors of human African trypanosomiasis throughout sub-Saharan Africa. Both sexes of adult tsetse feed exclusively on blood and contribute to disease transmission. Notable differences between tsetse and other disease vectors include obligate microbial symbioses, viviparous reproduction, and lactation. Here, we describe the sequence and annotation of the 366-megabase *Glossina morsitans morsitans* genome. Analysis of the genome and the 12,308 predicted protein–encoding genes led to multiple discoveries, including chromosomal integrations of bacterial (*Wolbachia*) genome sequences, a family of lactation-specific proteins, reduced complement of host pathogen recognition proteins, and reduced olfaction/chemosensory associated genes. These genome data provide a foundation for research into trypanosomiasis prevention and yield important insights with broad implications for multiple aspects of tsetse biology.

African trypanosomiasis is transmitted by the tsetse fly to humans (sleeping sickness) and livestock (nagana) throughout sub-Saharan Africa, with an estimated 70 million people at risk of infection. Rearing livestock in endemic areas is difficult to impossible and results in an economic loss in agricultural output of several billion U.S. dollars per year. Human infections are fatal if untreated, but tools for disease control are limited because it has not been possible to develop vaccines and current trypanocidal drug treatments result in undesirable side effects with growing reports of drug resistance. The reduction or elimination of tsetse populations is an effective method for disease control that could be improved with greater knowledge of their biology and genetics (*1*).

Tsetse flies are key representatives of the dipteran clade Calyptratae, which represents 12% of the known diversity within the dipteran order. Many of the calyptrate species are blood feeders of biomedical importance (*2*). In addition, members of the calyptrate family of Glossinidae and superfamily Hippoboscoidea, to which tsetse belong (fig. S1) (*3*), are defined by the ability to nourish intrauterine offspring from glandular secretions and give birth to fully developed larvae (obligate adenotrophic viviparity). Tsetse flies live considerably longer than other vector insects, which somewhat compensates for their slow rate of reproduction. Trypanosome infections in tsetse are acquired by blood feeding from an infected vertebrate host, and trypanosomes have to overcome multiple immune barriers to establish an infection within the fly. As a result, trypanosome infection prevalence is low in field populations and in experimentally infected tsetse (*4*). Tsetse have symbionts that compensate for their nutritionally restricted diet by the production of specific metabolites and influence multiple other aspects of the fly's immune and reproductive physiology (*5*).

In 2004, the International *Glossina* Genome Initiative (IGGI) was formed (*6*) to expand research capacity for *Glossina*, particularly in sub-Saharan Africa, through the generation and distribution of molecular resources, including bioinformatics training. An outcome of the effort undertaken by IGGI is the annotated *Glossina morsitans* genome presented here and further developed in satellite papers on genomic and functional biology findings that reflect the unique physiology of this disease vector (*7*–*14*).

Characteristics of the *Glossina* Genome

A combination of sequencing methods were used to obtain the *Glossina morsitans morsitans* (*Gmm*) genome, including Sanger sequencing of bacterial artificial chromosomes (BACs), small-insert plasmid and large-insert fosmid libraries, and 454 and Illumina sequencing (tables S1 and S2). The sequences were assembled into 13,807 scaffolds of up to 25.4 Mb, with a mean size of 27 kb and half the genome present in scaffolds of at least 120 kb. The 366-Mb genome is more than twice the size of the *Drosophila melanogaster* genome (fig. S2A and table S3). Clear conservation of synteny was detected between *Glossina* and *Drosophila*, but with the blocks of synteny tending to be twice as large in *Glossina* due to larger introns and an increase in the size of intergenic sequences, possibly as a result of transposon activity and/or repetitive sequence expansions. Sequences from most of the major groups of retrotransposons and DNA transposons are found in the *Glossina* genome (table S4). These sequences comprise ~14% of the assembled genome, in contrast to 3.8% of the *Drosophila*

*Members of the International *Glossina* Genome Initiative, affiliations, and individual contributions appear at the end of this paper.
†Corresponding author. E-mail: serap.aksoy@yale.edu (Serap Aksoy); geoffrey.attardo@yale.edu (G.M.A.); mb4@sanger.ac.uk (M.B.)