Supplementary information for

Structural basis of the crosstalk between histone H2B monoubiquitination and H3 lysine 79 methylation on nucleosome

Authors: Tonghui Yao[1], Wei Jing[1], Zhiguo Hu[1], Ming Tan[2], Mi Cao[2], Qianmin Wang[2], Yan Li[3], Guiyong Yuan[2], Ming Lei[2], Jing Huang[2]*.

Correspondence to: huangjing@shsmu.edu.cn

Materials and Methods

Protein Expression and Purification

Human DOT1L (residues 1-416) and its point mutations were expressed in *E. coli* BL21 (DE3) using a modified pET28b vector with a His_6-SUMO tag fused at the N-terminus. After induction for 16 hours with 0.1 mM IPTG at 20 ºC, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 5 mM 2-mercaptoethanol, 1 mM PMSF, 5 mM benzamidine, 1 μg/ml leupeptin, and 1 μg/ml pepstatin). The cells were then lysed by sonication and the cell debris was removed by centrifugation. The supernatant was mixed with Ni-NTA agarose beads (QIAGEN) and rocked for 2 hours at 4 ºC before elution with 250 mM imidazole. Then Ulp1 protease was added to remove the His_6-SUMO tag. After Ulp1 digestion, the DOT1L proteins were further purified through MonoS ion-exchange chromatography and HiLoad Superdex 200 gel-filtration chromatography (GE Healthcare), sequentially. Buffer A (25 mM Tris-HCl pH 7.5 and 150 mM NaCl) and Buffer B (25 mM Tris-HCl pH 7.5 and 1 M NaCl) were used for the gradient elution of proteins during the MonoS ion-exchange chromatography, and the buffer A was also used for

the HiLoad Superdex 200 gel-filtration chromatography. The purified proteins were concentrated to 10 mg/ml and stored at -80 ºC.

Human DOT1L (residues 1-351) was cloned into the GST-fusion expression vector pGEX6p-1 (GE healthcare), and was expressed in *E. coli* BL21 (DE3) as described above. After an affinity purification with the Glutathione Sepharose 4B beads (GE Healthcare) in lysis buffer, the GST tags were removed through the Protease 3C digestion, except the one that was used for the GST pull-down assays. The proteins were further purified through a HiLoad Superdex 200 gel-filtration chromatography in buffer A, and then were concentrated to 10 mg/ml and stored at -80 ºC.

Xenopus laevis histones and the related mutants were expressed in *E. coli* BL21 (DE3) and were purified as previously described[1] with some modifications. Briefly, the histones were expressed and isolated as inclusion bodies, and were further purified through sequential Q Sepharose HP (GE Healthcare) and SP Sepharose HP (GE Healthcare) chromatography in denatured buffer. The purified histones were dialyzed thoroughly against distilled water containing 2 mM 2-mercaptoethanol, and were then lyophilized and stored at -80 ºC.

Human ubiquitin protein with the G76C mutation was expressed in a modified pET28b vector (Novagen) with a His_6 tag and a Protease 3C recognition site fused at its N-terminus (His3C-Ub-G76C). The protein was purified on a Ni-NTA affinity column and on a HiLoad Superdex 75 gel-filtration column, sequentially. The purified His3C-Ub-G76C protein was then dialyzed thoroughly against 50 mM ammonium acetate (pH 4.5), and was lyophilized and stored at -80 ºC.

Chemical ubiquitination of H2B

A K120C mutation was deposited in H2B (H2B-K120C) to mediate the cross-linking between the residue K120 of H2B and the C-terminus of His3C-Ub-G76C. The H2B-K120C was chemically ubiquitinated as described before with some modifications[2]. In brief, 100 μM H2B-K120C and 100 μM His3C-Ub-G76C were incubated in 50 mM Borate (pH 8.1) and 1 mM Tris(2-carboxyethyl)phosphine (TCEP) at 50 ºC for 1 hour and then cooled down on ice for 1 hour. Dichloroacetone (DCA), the crosslinking reagent, was dissolved in dimethyl formamide (DMF), added to the above protein mixture to a final concentration of 100 μM, and incubated on ice for 1 hour. The reaction was quenched by adding 2-mercaptoethanol to a final concentration of 50 mM and was then lyophilized. The H2B-ub conjugate was separated from the reaction mixture through a denatured Ni-NTA affinity purification and a semi-preparative C4 reversed phase HPLC chromatography, sequentially; and the His3C tag was removed through the Protease 3C digestion. The purified ubiquitinated H2B was lyophilized and stored at -80 ºC.

Nucleosome reconstitution

The 147-bp nucleosome core particles (NCP) were reconstituted from four core histones (with or without ubiquitination) and the Wisdom 601 DNA as described previously[1]. In brief, histone octamers were first assembled and purified with HiLoad Superdex 200 gel filtration in the refolding buffer (20 mM Tris-HCl pH 7.5, 2 M NaCl, 1 mM EDTA, and 5 mM 2-mercaptoethanol). Then, the purified histone octamer was mixed with the Wisdom 601 DNA in a molar ratio of 0.9:1.0, and was dialyzed against the reconstitution buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT, and 0.25-2 M KCl). After nucleosome reconstitution, NCP was further purified through HiTrap DEAE Sepharose FF anion exchange and HiLoad Superdex 200 gel filtration, and was stored in small aliquots at -80 ºC.

Sample preparation and cryo-EM data collection

DOT1L was mixed with ubNCP (or NCP) and SAH in a molar ratio of 4:1:8, and was incubated for 1 hour at 4 ºC. Then the DOT1L-ubNCP complex (or the DOT1L-NCP complex) was purified through Superose 6 or HiLoad Superdex 200 gel filtration, and was further stabilized using the method of GraFix[3]. In brief, a continuous 10-30% gradient of glycerol with a 0-0.1% glutaraldehyde gradient was generated on a BioComp gradient master. Then, the peak fractions of the gel filtration purifications were loaded on top of the glycerol gradient, and were ultracentrifuged at 4 ºC for 14 h at 35,000 rpm using a SW 41Ti rotor (Beckman). The peak fraction of GraFix was dialyzed against 25 mM Tris-HCl pH7.5 and 150 mM NaCl, and was concentrated to 1 mg/ml. 2 µl aliquot of the samples were applied to glow-discharged holey carbon grids (Quantifoil R1.2/1.3, 300 mesh), and the grids were blotted for 3.0 s and then plunged into liquid ethane cooled by liquid nitrogen, using a Vitrobot Mark IV (FEI).

The samples were observed under a Titan Krios transmission electron microscope (FEI) operated at 300 kV. The images were collected on a Gatan K2 summit direct electron detector operated in counting mode, with a nominal magnification of 18,000x and a pixel size of 1.3 Å. The dataset of DOT1L-ubNCP has a defocus range of -0.6 to -2.2 µm and each micrograph was dose-fractioned to 38 frames with 0.2 s exposure time for each frame. The total accumulated dose of each micrograph is 35.7 $e^-/Å^2$. The dataset of the DOT1L-NCP complex has a defocus range of -0.6 to -3.0 µm and the total accumulated dose of each micrograph is 37.1 $e^-/Å^2$. The imaging conditions were also listed in Supplementary information, Table S1.

Image processing and model building

A total of 2,336 cryo-EM images of the DOT1L-ubNCP complex were collected, and motion correction was performed on the dose-fractioned image stacks using MotionCor2 with dose weighting[4,5]. The CTF parameters of each image were determined with Gctf[6], and automatic

particle picking was carried out using Gautomatch based on a 2D-class template generated from 3,000 manually picked particles. 2D classification, 3D initial model, 3D classification, and 3D auto-refine were performed with RELION 2.1[7]. An overview of the data processing procedure was shown in Supplementary information, Fig. S4. Two rounds of 2D classification and one round of 3D classification with exhaustive angular searches on the DOT1L-ubNCP dataset yielded a class of 151,545 particles that primarily correspond to the DOT1L-ubNCP complex in a molar ratio of 1:1 and another class of 189,786 particles that could be further classified to identify the DOT1L-ubNCP complex with a molar ratio of 2:1 (a class of 55,778 particles). After 3D auto-refine and solvent masked post-processing, the 151,545 particles of DOT1L-ubNCP with a molar ratio of 1:1 generated a map with an overall resolution of 4.1 Å. To improve the map densities of the N-terminus of DOT1L as well as ubiquitin, the 151,545-particle dataset was further processed through masked 3D classifications with partial signal subtraction[8]. In brief, soft masks were applied to the regions of interest and the signals outside the masked regions were subtracted from the experimental images. Masked 3D classifications without angular searches were perform on the extracted regions and good classes were selected for further refinement on the original particle images. Finally, maps with good densities on $DOT1L^{N}$ or ubiquitin were calculated at resolutions of 4.5 Å and 4.3 Å, respectively. For the 55,778 particles of the DOT1L-ubNCP complex with a molar ratio of 2:1, they were processed with 3D auto-refine and generated a density map of 6.0 Å. Solvent masked post-processing was not carried out because it blurred the densities of DOT1L and ubiquitin. The resolution estimation was based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion[9] and the local resolution was estimated with ResMap[10]. For the DOT1L-NCP complex, a total of 1,829 images were collected and processed similarly as above. After 2D classification, 3D classification, masked 3D

classifications with partial signal subtraction and refinement, a class of 131,329 particles generated a final map of DOT1L-NCP at an overall resolution of 5.0 Å.

Model building was carried out by fitting the available structures of NCP, ubiquitin and DOT1L (PDB codes: 5KGF and 1NW3) in the EM density map of DOT1L-ubNCP with a molar ratio of 1:1 using UCSF Chimera[11]. The model was then manually built in Coot[12] and real-space refined with secondary structure restraints in Phenix[13].

Histone methyltransferase (HMT) assay

For a 50 µl HMT reaction, 0.5 µg DOT1L, 20 µM SAM and 2 µg NCP or ubNCP were mixed in the buffer of 20 mM HEPES-Na pH7.8, 5 mM DTT, 0.5 mM EDTA, and 2% glycerol, and were incubated for 1 hour at 30 ºC. The reaction mixtures were then separated on a 15% SDS-PAGE gel and the mono-, di- and tri-methylation of H3K79 were detected by western blot using the corresponding antibodies (H3K79me1 antibody: #39145, ActiveMotif; H3K79me2 antibody: #39143, ActiveMotif; H3K79me3 antibody: #ab2621, Abcam).

GST pull-down assay

100 µg of GST, GST-DOT1L(1-416), or GST-DOT1L(1-351) were incubated with 100 µg of NCP, ubNCP or ub, and 50 µl of Glutathion Sepharose 4B beads in the buffer of 25 mM Tris-HCl pH7.5 and 100 mM NaCl at 4 ºC overnight. Then the beads were extensively washed for four times and the bound proteins were eluted with 10 mM reduced glutathione and separated on a 15 % SDS-PAGE gel.

Data availability

The EM density maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-9783, and EMD-9784. The final model was submitted to the RCSB Protein Data

Bank under accession code 6J99. All other data are available from the corresponding author upon reasonable request.

Supplementary References

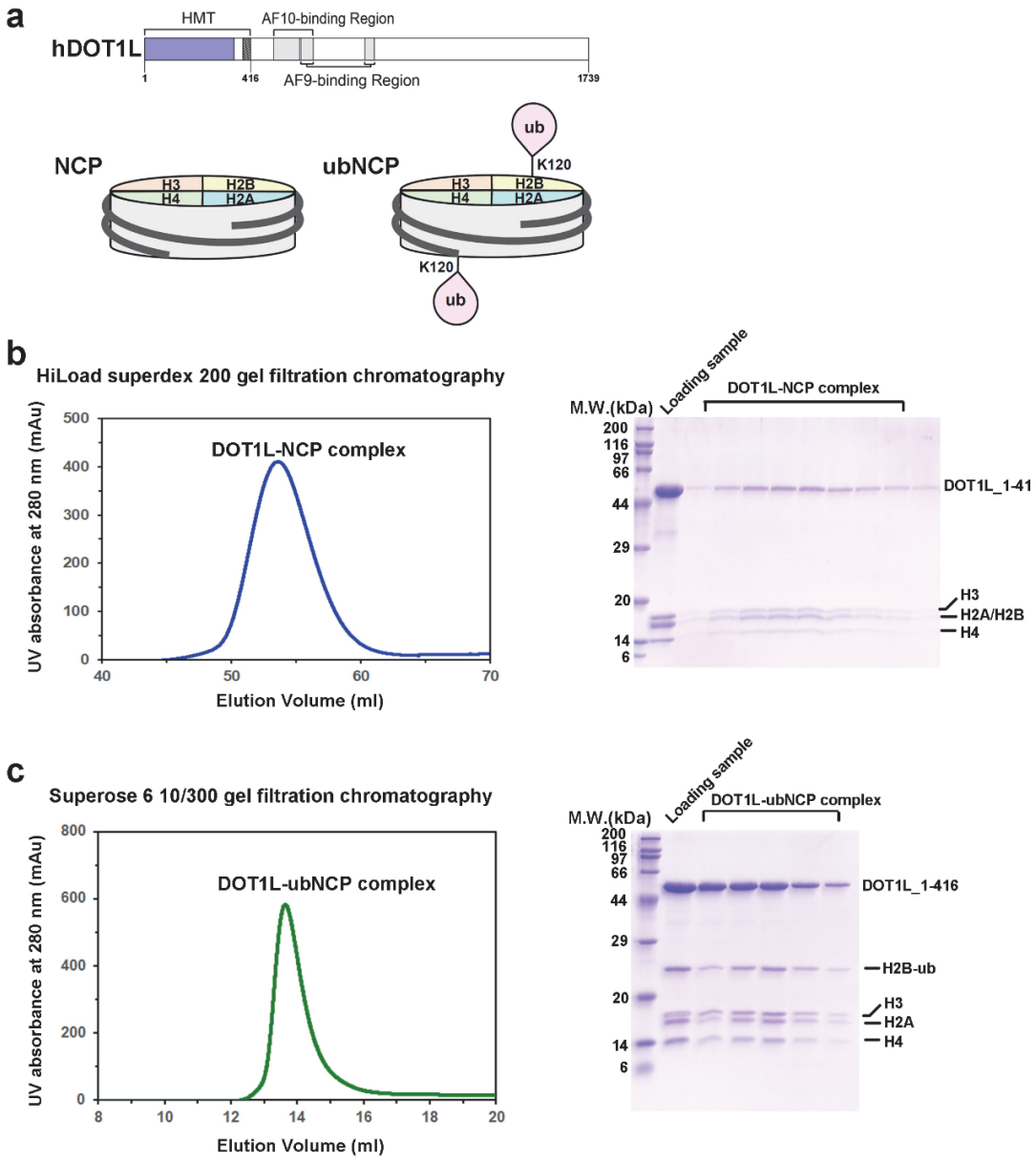
1. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375,** 23–44 (2004).
2. Morgan, M. T. *et al.* Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351,** 725–728 (2016).
3. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5,** 53–55 (2008).
4. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14,** 331–332 (2017).
5. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4,** 1–19 (2015).
6. Zhang, K. Gctf : real-time CTF Determination and Correction. *J. Struct. Biol.* **193,** 1–12 (2016).
7. Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5,** 1–21 (2016).
8. Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the conformational space of the catalytic subunit of human g-secretase. *Elife* **4,** 1–19 (2015).
9. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9,** 853–854 (2012).

10. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).

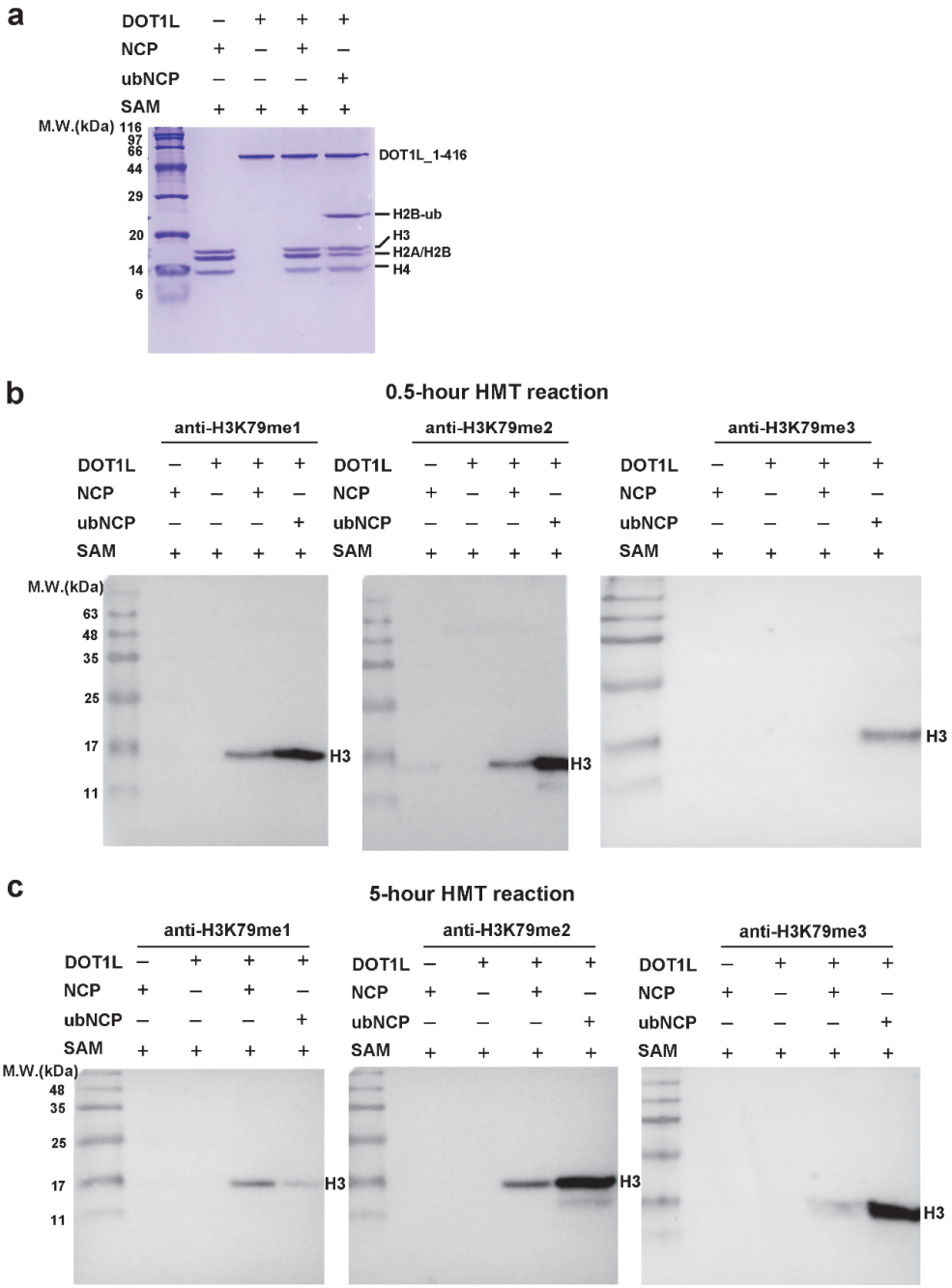
11. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

12. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

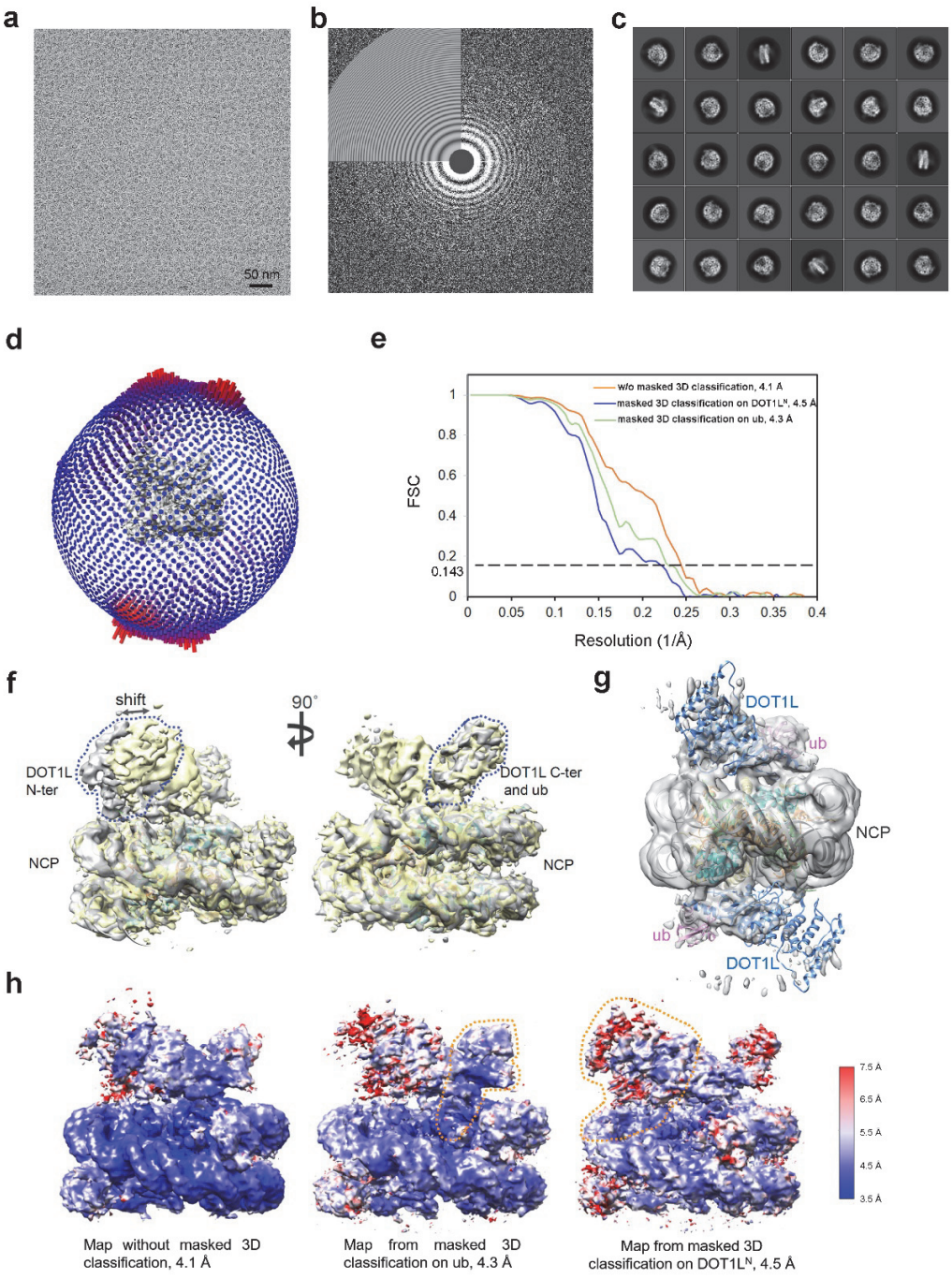
13. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).



Supplementary information, Fig. S1. ***In vitro*** **reconstitution of the DOT1L-NCP and DOT1L-ubNCP complexes. a,** Schematic drawings of the domain organization of human DOT1L as well as the structural organizations of NCP and ubNCP. **b,** Gel filtration and SDS-PAGE analysis of the *in vitro* reconstituted DOT1L-NCP complex. **c,** Gel filtration and SDS-PAGE analysis of the *in vitro* reconstituted DOT1L-ubNCP complex.

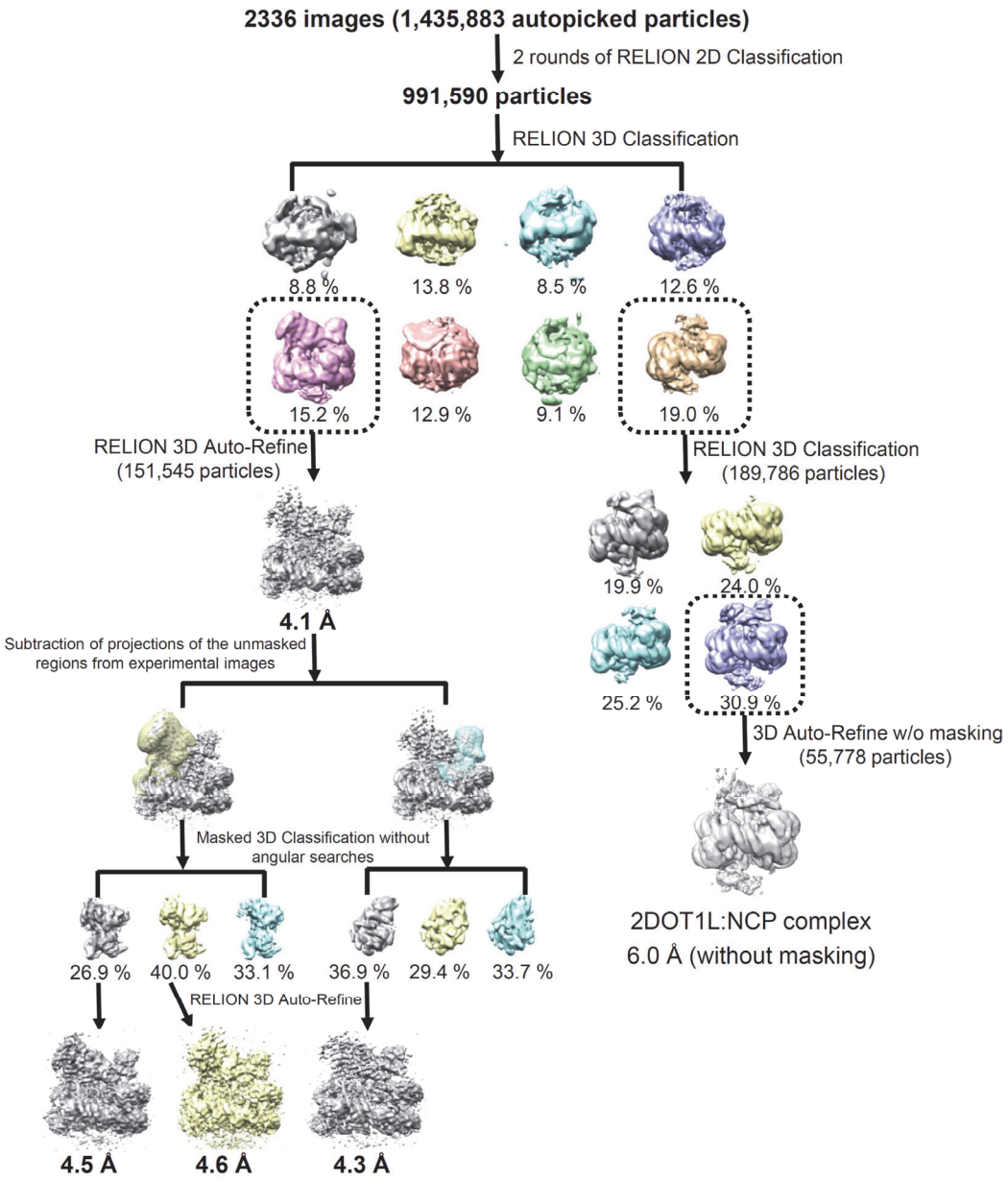


Supplementary information, Fig. S2. ***In vitro*** **HMT (Histone Methyltransferase) assays of DOT1L on NCP and ubNCP. a,** Input of the HMT reaction. **b,** The HMT reaction incubated for 30 min. **c,** The HMT reaction incubated for 5 hours. Comparison of Supplementary information, Fig. S2b and S2c indicates that longer reaction time could lead to the accumulation of di- and tri-methylation marks and decrease of mono-methylation on both types of substrates. This observation agrees with previous reports that DOT1L appears to methylate histone H3K79 in a distributive manner.

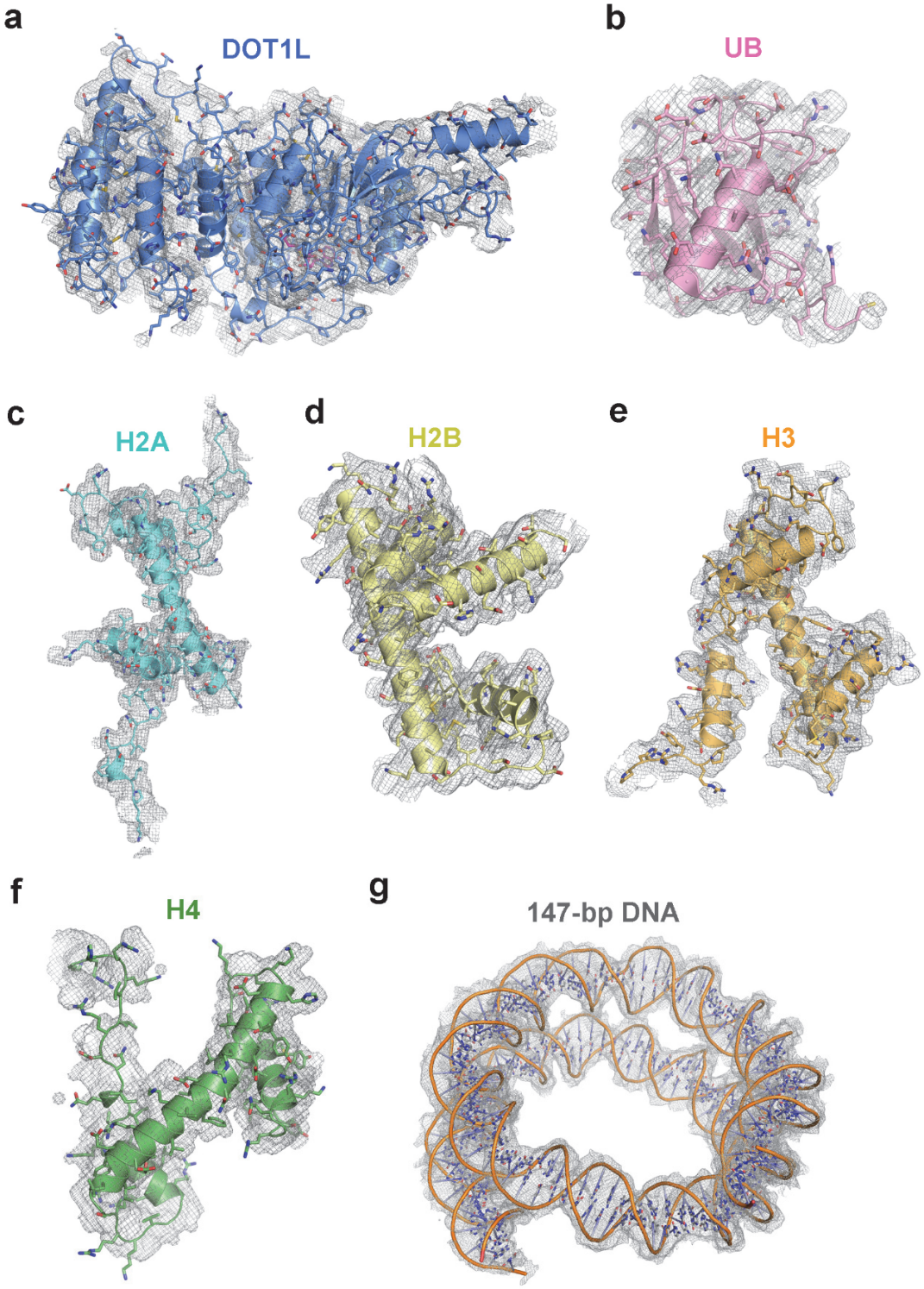


Supplementary information, Fig. S3. Cryo-EM analysis of the DOT1L-ubNCP complex. a, Representative micrograph of the cryo-EM dataset of the DOT1L-ubNCP complex. **b,** Fast Fourier transform of the image in **a**, with the Thon rings extending to ~3.0 Å. **c,** Representative two-dimensional (2D) class averages obtained from cryo-EM particles of the DOT1L-ubNCP complex with RELION. **d,** Angular distribution of particle projections of the DOT1L-ubNCP

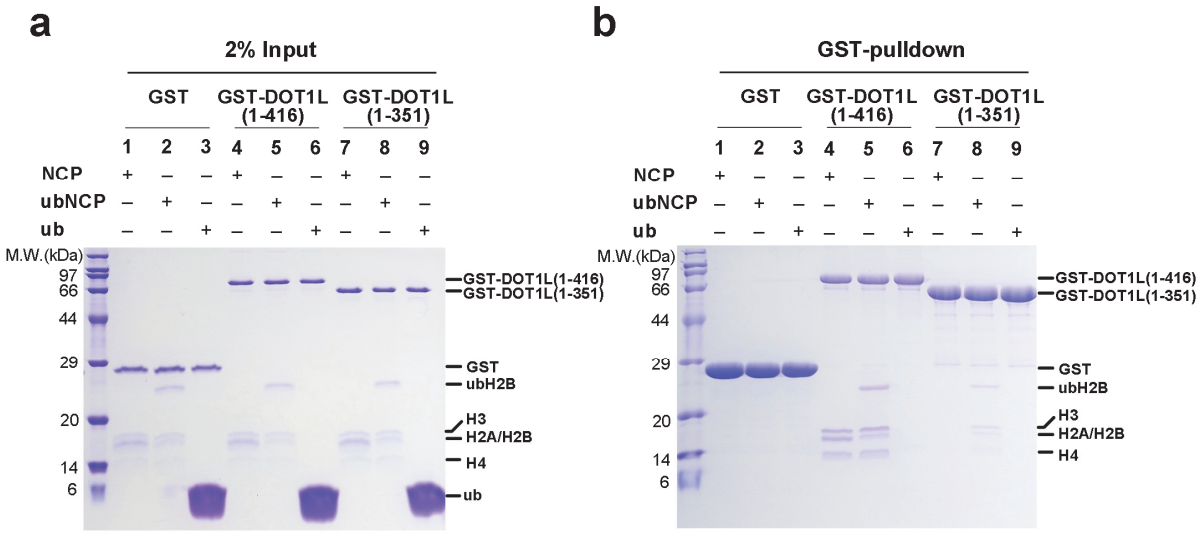
reconstruction. **e,** The 'gold-standard' FSC curves calculated between two halves of datasets for the 4.1-Å map (orange), the 4.5-Å map (blue), and the 4.3-Å map (green). **f,** Masked classifications on $DOT1L^N$ (left) or ubiquitin (right) with subtraction of the signal from the rest of the complex generated three-dimensional (3D) reconstructions of different classes. Structural comparisons of the 3D classes revealed small movement of $DOT1L^N$ on the disk-face of nucleosome. **g,** Cryo-EM map of the 2:1 DOT1L-ubNCP complex that's fit in with the 1:1 DOT1L-ubNCP model. **h,** Local resolution estimates of the 4.1-Å, 4.3-Å and 4.5-Å maps of the 1:1 DOT1L-ubNCP complex. The map densities that were improved after the respective masked 3D classification with subtraction of the signal from the rest of the complex were circled with orange dash lines.



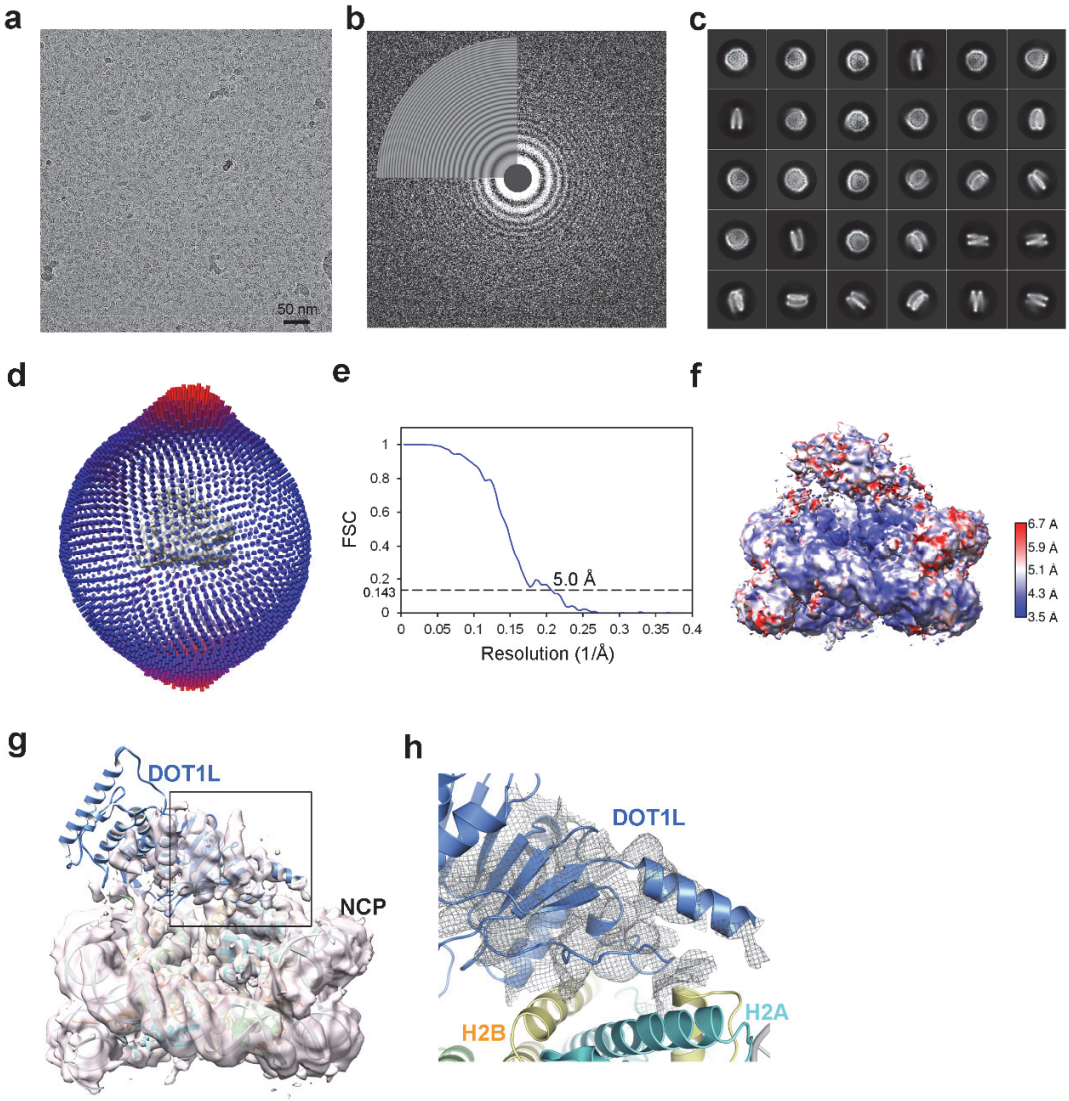
Supplementary information, Fig. S4. Flow chart of cryo-EM data processing of the DOT1L-ubNCP dataset. After the initial 3D classification, particles for the 1:1 and 2:1 ratio of DOT1L-ubNCP complexes were further processed, respectively. Solvent masked post-processing was not carried out on the 2DOT1L-NCP complex because it severely blurred the densities of DOT1L and ubiquitin.



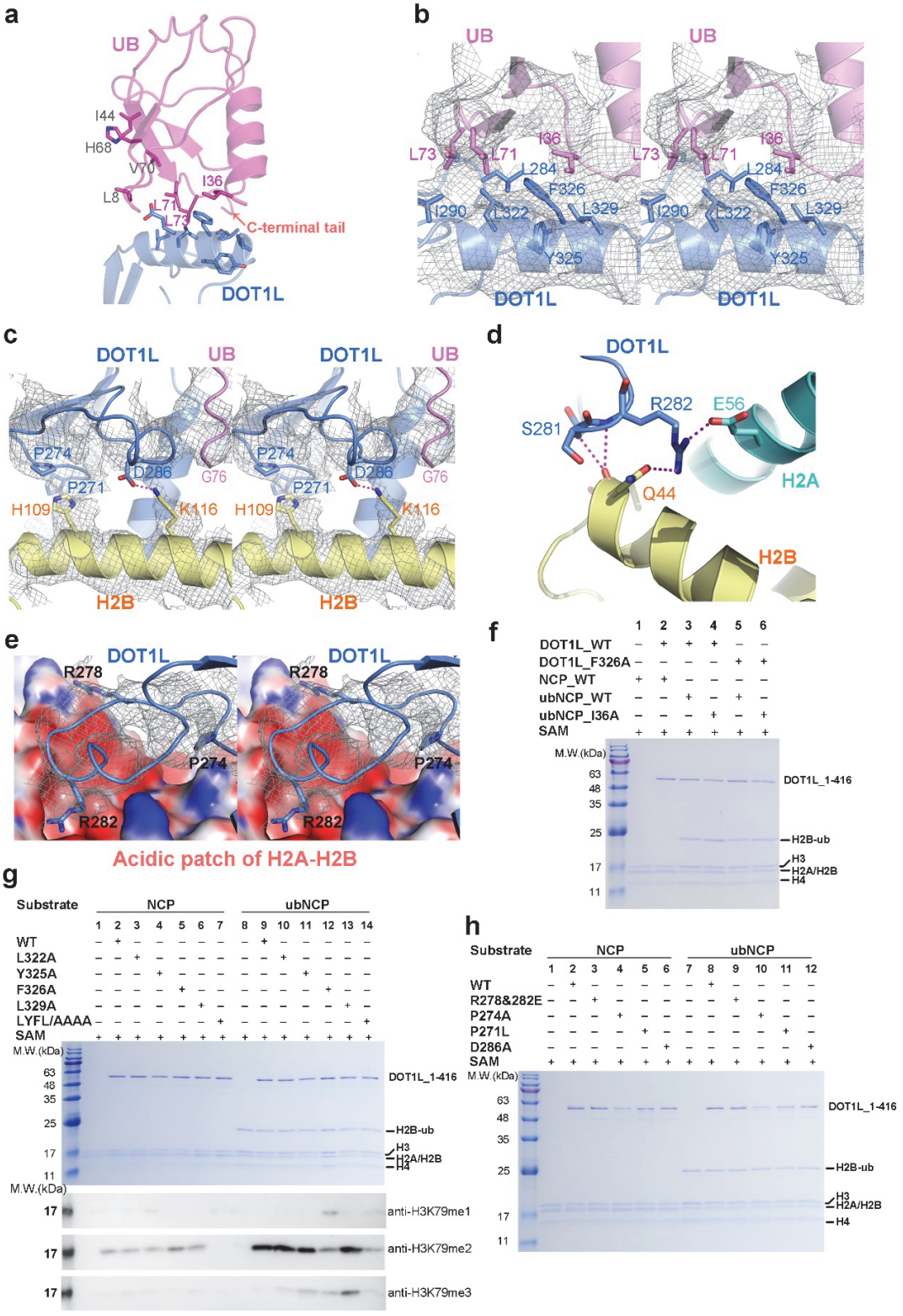
Supplementary information, Fig. S5. Cryo-EM data quality of the DOT1L-ubNCP complex. Segmented map of each component of DOT1L-ubNCP is fit in with the respective structural model.



Supplementary information, Fig. S6. GST pull-down assay of DOT1L with NCP, ubNCP, and ub. a, 2% Input. **b,** DOT1L (residues 1-351) that lacks a DNA-binding positively charged region (residues 351-416) could only interact with the ubiquitinated NCP but not with the unmodified NCP or free ubiquitin, as shown in the GST pull-down assay.

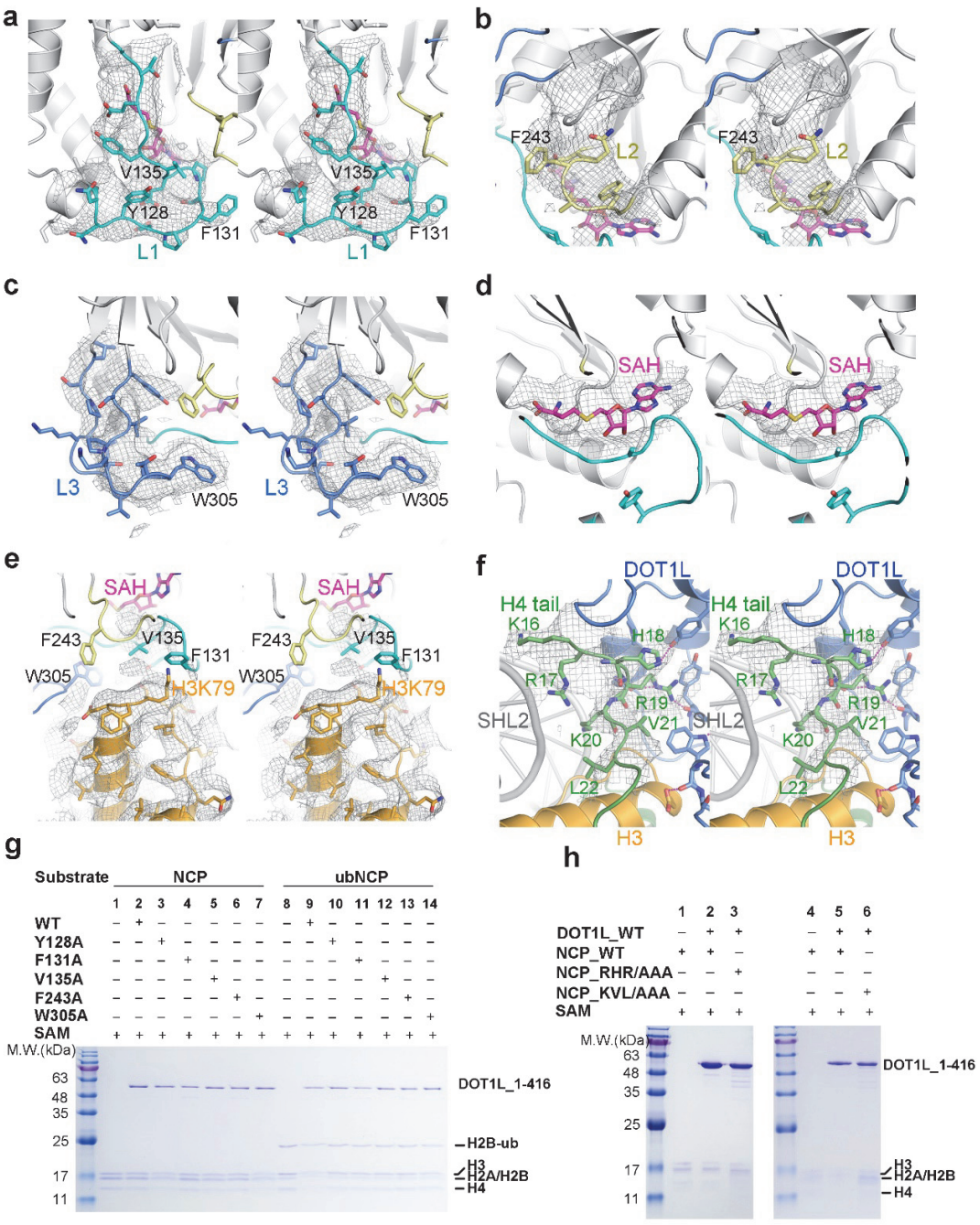


Supplementary information, Fig. S7. Cryo-EM analysis of DOT1L in complex with unmodified NCP. a, Representative micrograph of the cryo-EM dataset of the DOT1L-NCP complex. **b,** Fast Fourier transform of the image in **a**, with the Thon rings extending to ~3 Å. **c,** Representative 2D class averages obtained from cryo-EM particles of the DOT1L-NCP complex with RELION. **d,** Angular distribution of particle projections of the DOT1L-NCP reconstruction. **e,** The 'gold-standard' FSC curve calculated between two halves of datasets for the DOT1L-NCP reconstruction. **f,** Local resolution estimate of the density map of DOT1L-NCP. **g,** Docking of the DOT1L-ubNCP structural model into the density map of DOT1L-NCP shows a good fit of the DOT1L model from the ubiquitinated complex with the DOT1L density of the unmodified one. **h,** An enlarged view of the boxed region in **g**.

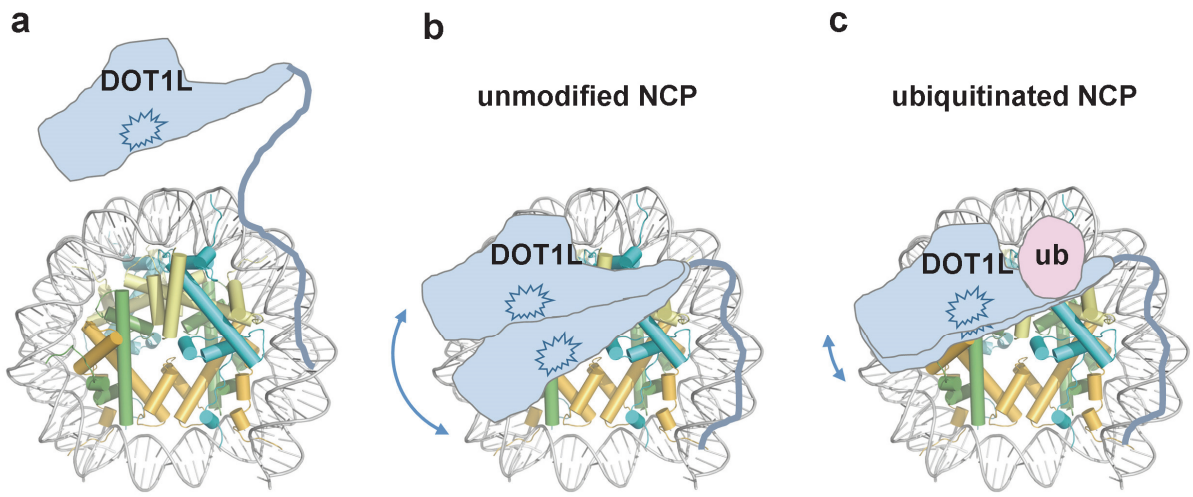


Supplementary information, Fig. S8. Interfaces among DOT1L, ubiquitin, and the histone H2A-H2B dimer. a, Key residues of the I44 patch (colored in black) and the I36 patch (colored

in magenta) of ubiquitin were denoted in the structure of DOT1L-ubNCP. **b,** Density map of the DOT1L-ubiquitin interacting region shown in stereo mode, related to Fig. 1b. **c,** Segmented cryo-EM density of the recognition interface between $DOT1L^{L}$ and the ubiquitin-conjugated C-terminal helix of histone H2B shown in stereo mode, related to Fig. 1c. **d,** Detailed interactions between DOT1L Arg282 and the acidic patch of histone H2A-H2B. **e,** Density map of $DOT1L^{L}$ showing the residues Arg278 and Arg282 interacts with the acidic patch of histone H2A-H2B (shown with electrostatic surface potential), related to Fig. 1d. **f,** Input of the HMT reactions related to Fig. 1h. **g,** HMT activities of the DOT1L mutants bearing alanine substitutions of the residues at the DOT1L-ubiquitin interface, as determined with antibodies of H3K79me1, H3K79me2, and H3K79me3. Input of the HMT reactions is shown in the upper panel. **h,** Input of the HMT reactions related to Fig. 1i.



Supplementary information, Fig. S9. Interactions between the active site of DOT1L and the histones H3 and H4. a-d, Segmented maps of the loop L1 (**a**), the loop L2 (**b**), the loop L3 (**c**), and SAH (**d**) at the active site of DOT1L shown in stereo mode, related to Fig. 1e. **e,** Segmented map of the histone H3K79 shown in stereo mode. **f,** Segmented cryo-EM density of the recognition interface between the histone H4 tail and the active site of DOT1L shown in stereo mode, related to Fig. 1g. **g,** Input of the HMT reactions related to Fig. 1j. **h,** Input of the HMT reactions related to Fig. 1k.



Supplementary information, Fig. S10. A working model for the regulation of H3K79 methylation by H2BK120ub1 on nucleosome. a, DOT1L is initially recruited to nucleosome by the electrostatic interactions between its positively charged region and the nucleosomal DNA. **b,** Then, the catalytic domain of DOT1L is attached to the disk-face of nucleosome. On the unmodified NCP, due to the weak interactions between DOT1L and the core histones, the active site of DOT1L frequently dissociates from the site of histone H3K79, resulting in the low efficiency of H3K79 methylation. **c,** On the H2BK120-ubiquitinated nucleosome, DOT1L is sandwiched between ubiquitin and the histone H2A-H2B dimer. The mobility of DOT1L on ubNCP is largely restricted, leading to enhanced catalytic efficiency and rapid accumulation of higher methylated states on the ubiquitinated nucleosome.

Table S1. Cryo-EM Data Collection and Refinement Statistics

	DOT1L-ubNCP complex	DOT1L-NCP complex
Data collection and processing		
Microscope	Titan Krios	Titan Krios
Detector	K2 summit	K2 summit
Voltage (kV)	300	300
Electron exposure (e-/ Å^2)	35.7	37.1
Defocus range (μm)	-0.6 to -2.2	-0.6 to -3.0
Pixel size (Å)	1.3	1.3
Symmetry imposed	C1	C1
Initial particle images (no.)	1,435,883	575,638
Final particle images (no.)	151,545	131,329
Map resolution (Å) FSC threshold: 0.143	4.1	5.0
Refinement		
Initial model used (PDB code)	5KGF and 1NW3	
Map sharpening B factor (Å^2)	-141	
Model-to-map fit, map correlation coefficient	0.816	
Model composition		
Non-hydrogen atoms	15,279	
Protein residues	1163	
DNA nucleotides	289	
Ligand	1	
R.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (°)	0.96	
Validation		
MolProbity score	1.83	
Clashscore	7.78	
Poor rotamers (%)	0.00	
Ramachandran plot		
Favored (%)	91.95	
Allowed (%)	8.05	
Disallowed (%)	0.00	