ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Denoising Autoencoder Trained on Simulation-Derived Structures for Noise Reduction in Chromatin Scanning Transmission Electron Microscopy

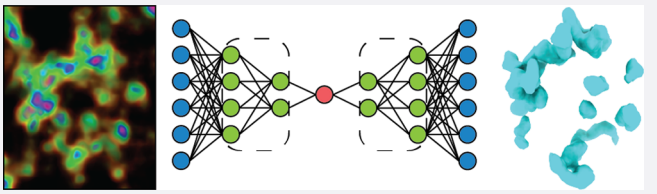
Walter Alvarado, Vasundhara Agrawal, Wing Shun Li, Vinayak P. Dravid, Vadim Backman,* Juan J. de Pablo,* and Andrew L. Ferguson*

Cite This: *ACS Cent. Sci.* 2023, 9, 1200−1212

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Scanning transmission electron microscopy tomography with ChromEM staining (ChromSTEM), has allowed for the three-dimensional study of genome organization. By leveraging convolutional neural networks and molecular dynamics simulations, we have developed a denoising autoencoder (DAE) capable of postprocessing experimental ChromSTEM images to provide nucleosome-level resolution. Our DAE is trained on synthetic images generated from simulations of the chromatin fiber using the 1-cylinder per nucleosome (1CPN) model of chromatin. We find that our DAE is capable of removing noise commonly found in high-angle annular dark field (HAADF) STEM experiments and is able to learn structural features driven by the physics of chromatin folding. The DAE outperforms other well-known denoising algorithms without degradation of structural features and permits the resolution of α-tetrahedron tetranucleosome motifs that induce local chromatin compaction and mediate DNA accessibility. Notably, we find no evidence for the 30 nm fiber, which has been suggested to serve as the higher-order structure of the chromatin fiber. This approach provides high-resolution STEM images that allow for the resolution of single nucleosomes and organized domains within chromatin dense regions comprising of folding motifs that modulate the accessibility of DNA to external biological machinery.

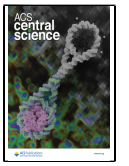


INTRODUCTION

Chromatin is the highly organized complex of DNA, RNA, and proteins that packages DNA within the cell nucleus, prevents DNA damage, and controls replication and gene expression.[1] The main organizational unit of chromatin is the nucleosome core particle constituting a complex of DNA wrapped around a histone octomer.[2] Structurally, the nucleosome is approximately 146 base pairs (bps) of DNA wrapped in 1.67 left-handed superhelical turns around two copies of the H2A, H2B, H3, and H4 proteins. Chromosomes can contain hundreds of thousands of nucleosomes linked by short strands of DNA, which give it the appearance of beads on a string. The structure of these 11 nm wide nucleosomal disks is nearly conserved across all eukaryotic cells and serves as the repeating building block of chromatin.[3] Beyond this basic structural unit, chromatin is believed to have several hierarchical levels of DNA packaging, beginning with a 10 nm fiber that further compacts into a 30 nm fiber, the latter of which has been considered to be a key intermediate level of chromatin organization and compaction within the eukaryotic nucleus.[4] The structure of the 30 nm fiber is characterized as a nucleosomal chain folding into a solenoid or a "one-start" helical structure. Each nucleosome in this configuration interacts with its fifth and sixth surrounding nucleosomes as the nucleosomes coil around a central cavity at a rate of about six nucleosomes per turn.[5] Though first observed under an electron microscope *in vitro*, the relevance of the 30 nm fiber *in vivo* remains an open question.[4,6,7] More recently, studies have suggested nucleosomes can arrange themselves into stable secondary structural arrays comprised of four nucleosomes that play an important regulatory function by controlling the accessibility of DNA to external biological machinery.[8−11] While these tetranucleosomes have been observed in reconstituted chromatin fibers *in vitro* and suggested by modeling studies *in silico*, current imaging techniques remain insufficient to resolve their existence *in situ*.[12]

Recently, chromatin staining coupled with electron and scanning transmission electron microscopy (ChromEM and ChromSTEM, respectively) have resolved the 3D organization of chromatin and observed distinct, anisotropic packing domains.[13,14] The size and variability of these domains across different cell type have been suggested to regulate gene activity

Received: February 9, 2023
Published: June 5, 2023



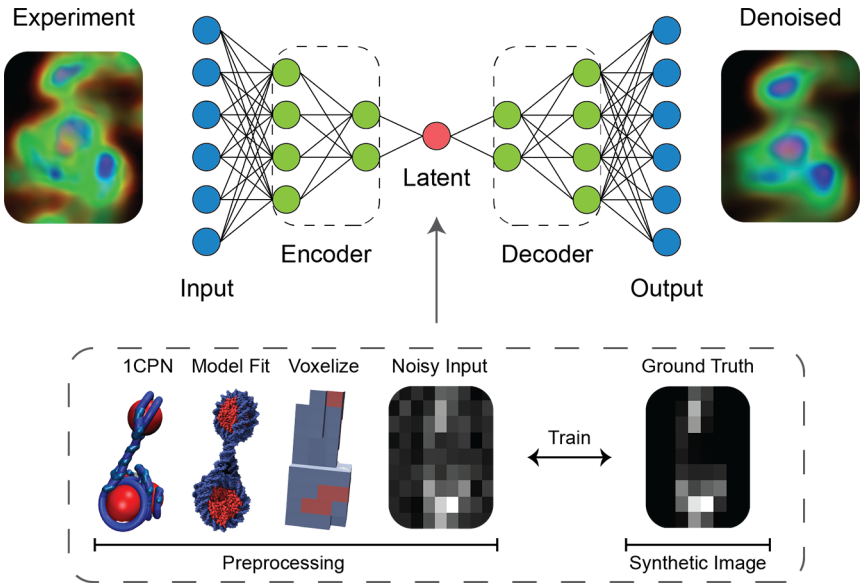


Figure 1. A denoising autoencoder (DAE) is constructed and trained on simulations of the chromatin fiber. We simulate nucleosome arrangements using the 1CPN model of chromatin and use the resulting trajectories to generate synthethic STEM images by superimposing crystal structures of the nucleosome (PDB: 1KX5) and DNA snippets. Noise commonly found in angle annular dark field (HAADF) STEM experiments is applied to the images and the DAE trained to remove this noise and preserve the underlying signal.

by controlling the size of macromolecular complexes that can access DNA within these clusters, thereby affecting processes such as DNA transcription, replication, and repair. In addition, variability in statistical and morphological properties of packing domains may potentially play an important role in the construction of higher-order chromatin structures such as euchromatin and heterochromatin.[15] While these experimental imaging techniques have provided key insights into the chromatin structure, nucleosome-level packing remains obscured by statistical noise inherent to STEM imaging.[12,16] In particular, the spatial organization of nucleosomes within dense chromatin regions suffers from low signal-to-noise ratios at these smaller length scales. Denoising STEM images provides a means to identify folding motifs and advance understanding of the details of chromatin structure, nucleosome packing, and the structure−function relation.

By combining the advances made in STEM imaging for chromatin, molecular dynamics simulations, and machine learning, we designed a deep convolutional denoising autoencoder (DAE) for STEM image denoising. Since noiseless experimental images upon which to train our denoising models are not available, we instead generate noise-free training data using by molecular dynamics (MD) simulations. This strategy is similar to the approach employed by Ziatdinov et al. in studying the surface of molecular structures.[17] We conduct simulations of the chromatin fiber using the 1-cylinder per nucleosome (1CPN) model that has been shown to accurately reflect the possible conformations of oligonucleosomal structures.[11,18,19] Snapshots from these MD trajectories are then converted to synthetic ChromSTEM image data sets which are used to train the DAE to remove noise artificially added to the training images and produce images with enhanced structural resolution that enable the identification and analysis of folding motifs within dense DNA regions. The DAE outperforms other well-known denoising algorithms and, as we demonstrate in applications of the trained model to experimental ChromSTEM images, resolves specific tetranucleosome motifs that induce local chromatin compaction and are known to mediate DNA accessibility. Notably, we find no evidence for the 30 nm fiber, which has been suggested to serve as the higher-order structure of the chromatin fiber.[20,21] Our machine-learning-enabled DAE presents a means to bridge experimental ChromSTEM imaging and physics-based molecular dynamics simulations to realize high-resolution, denoised images capable of resolving previously unidentifiable tetranucleosome motifs to advance the understanding of the small-scale organization of chromatin and the relationship of structure to function.

METHODS

Coarse-Grained Molecular Dynamics Simulations and Generation of Synthetic STEM Data. We train our DAE on tomographic images generated from MD simulations of the chromatin fiber (Figure 1). To generate a synthetic data set, coarse-grained molecular dynamics simulations were carried out using the 1-cylinder per nucleosome (1CPN) model of chromatin.[18] The 1CPN model is parametrized by explicit experimental measurements and atomistic models of DNA that preserve molecular-level nucleosome physics enabling kilobase-scale simulations of genomic DNA. The 1CPN model is an appropriate choice, since it has been extensively validated in the literature as a reliable model for capturing chromatin dynamics.[18] The model was fitted against experimental data and has demonstrated its ability to reproduce a wide range of chromatin processes that include nucleosome unwrapping, sedimentation coefficients, and interactions between nucleosomes, which is a primary mechanism that drives chromatin folding.[11,19]

We conducted the 1CPN simulations under conditions representative of those under which the ChromSTEM images were acquired. As anticipated, the 30 nm fiber was not observed within in our simulations, as the conditions that typically involve its formation are due to specific *in vitro* environmental conditions such as the inclusion of high-affinity 601 DNA repeats and a cationic environment (e.g., 1−2 mM Mg^{2+}).[22] Furthermore, cryo-EM images of the 30 nm fiber

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

have not been reported for mitotic chromosomes *in vivo*.[21] We note, however, that our pipeline is designed to be easily adaptable to new conditions and that transfer learning could be used to augment the existing model by repeating the simulations under the conditions under which the new experimental data were gathered and retraining the DAE.

After equilibration, three 30 μs replicas were conducted totaling 150 μs of simulation time of chromatin fibers varying from 150 to 200 nucleosome repeat lengths (NRLs) and comprised of 4−16 nucleosomes. The lengths and sizes were chosen to account for the natural variability in biological systems. We highlight that our simulations cover long time scales that have not been reached by previous studies. This extended simulation time allows for a more comprehensive exploration of the phase space and reduces the risk of being trapped in certain emergy minima. The 1CPN model's effectiveness in representing chromatin behavior helps to ensure that our simulation snapshots are representative of the physical system under study. The combination of long-time-scale simulations and the use of the 1CPN model provides a strong foundation for generating a diverse and representative training data set for our denoising autoencoder. We performed an internal consistency verification that the 150 μs simulations of each system were sufficiently long to comprehensively probe the relevant configurational phase space by verifying that the phase space ensemble visited by the first 75 μs and second 75 μs produced similar distributions in key structural order parameters such as radius of gyration and root-mean-square deviation in reference to the initial elongated fiber structure.

Approximately 16000 snapshots from all simulation trajectories were extracted at 28 × 28 pixel resolution. These synthetic images represent a variety of conformations of the chromatin fiber at a resolution commensurate with that of typical ChromSTEM imaging experiments.[14,15] From this data set, 12702 conformations were selected for training and 3176 held out as a validation set. An X-ray crystal structure of the nucleosome core particle at 1.9 Å resolution (PDB: 1KX5) was superimposed to the location of each nucleosome bead and linker DNA was built with repeating ATAT bases.[23] Each structure was converted to a point cloud representation and then voxelized to resemble a high-angle annular dark-field scanning transmission electron microscope (HAADF-STEM) tomogram. Each synthetic image stack contained 28 × 28 × 9 voxels with a voxel dimension of approximately 3 × 3 × 3 nm^3 corresponding to the approximate 27 nm^3 volume captured in an experimental STEM voxel. Mathematically, the voxel intensity, $I_{m,n}$, is given by the total number of atoms that are enclosed within the volume of a voxel unit, $V_{m,n}$

$$I_{m,n}(x) = \sum_{i=1}^{N} [x_i \in V_{m,n}] \tag{1}$$

where the position of a given atom is given by x_i, and m and n denote the row and column indices of a voxel within a 28 × 28 × 1 voxel 2D planar slice, I, of the 3D voxel stack and where we have used Iverson's bracket notation to denote the indicator function. Finally, the synthetic image intensity is normalized to match the distribution of voxel intensity in experimental tomograms.[24−26]

HAADF-STEM has emerged as a powerful imaging technique that provides nanoscale-level structural detail.[27,28] It is, however, sensitive to environmental and instrumental noise during image acquisition that introduces extraneous signals not associated with the scattering of the sample.[16,29,30] For example, images are acquired at different projection angles by tilting the sample stage, at high tilt angles; however, focusing becomes more difficult, which leads to image blurring.[31] In addition, limited beam penetration and focal depth coupled with the restricted tilt range results in a lower set of projections which also introduces artifacts (i.e., "missing cone" artifacts).[32,33] Beam damage and environmental noise (e.g., airflow, sound, temperature, etc.) also deteriorate image quality and limit the accuracy of HAADF-STEM tomographic reconstruction.[16,29,34] Due to the particle nature of electrons and the collection method, Poisson noise remains the dominant form of noise in STEM imaging.[16,35] To account for these effects within our simulated data, we apply several HAADF-STEM-related noise conditions including Gaussian noise, Poisson noise, and tip-blurring effects to each simulated image similar to the approach implemented by Schwenker et al.[24−26] Parameters such as broadening effects, counts, and additive background noise were adjusted to account for the different levels of noise that may be encountered during image acquisition. Mathematically, each noise-free image, I, generated from the MD simulations is converted into an artificially noisy image, $\tilde{I}$, by corrupting it with articial noise under the noise model

$$\tilde{I} = I + I_{\text{Poisson}} + I_{\text{Gaussian}} + I_{\text{Scan}} \tag{2}$$

Given that Poisson noise is not additive and correlated with voxel intensity, we instead begin by applying a signal-dependent Poisson noise layer on top of each noise-free image using the discrete probability distribution

$$I_{\text{Poisson}} \approx \Pr(N = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3}$$

where N represents the number of photons measured by a given sensor and λ is the expected number of photos per unit time interval. We make the assumption that the number of atoms counted in a given voxel unit ($I_{m,n}$) is similar to photon counting in a classic Poisson process.

STEM images are susceptible to thermal vibrations and electronic noise which can be modeled as a Gaussian process.[36] To account for this, we add a Gaussian noise layer that obeys the distribution

$$I_{\text{Gaussian}} \approx N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(z-\mu)^2/2\sigma^2} \tag{4}$$

where μ is equal to the mean of the image and σ is the standard deviation which represents the broadening (i.e., "spread") of the signal. Similar to the approach by Schwenker et al. to emulate noise and distortion conditions common to the HAADF-STEM imaging mode, we set σ = 0.8.[24−26]

Finally, scan line shifts, I_{Scan}, are random, persistent, time-dependent distortions that occur due to positioning errors of the electron beam that result in shifts in the image perpendicular to the scan lines.[37] We generated this type of noise by introducing approximately a 1 subpixel offset randomly along the x direction and resampling these random shifts via bilinear interpolation

$$I_{\text{Scan}} \approx I_{u_x,u_y} = I_{m,n}(u_x, u_y) \tag{5}$$

where u_x and u_y are the desired shifts across the range $[-1, 1)$.

Denoising Autoencoder (DAE) Architecture. As the name suggests, denoising autoencoders (DAEs) are artificial

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

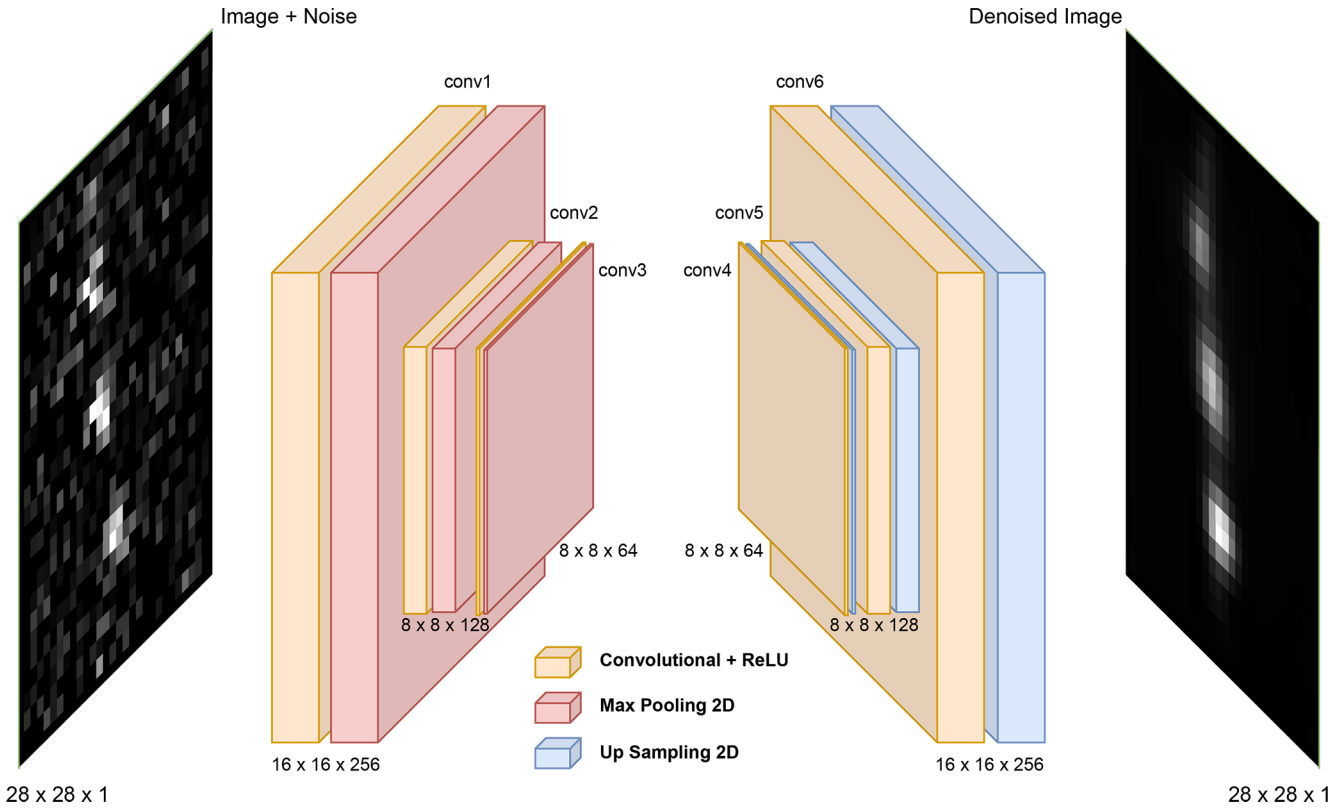


Figure 2. A denoising autoencoder (DAE) comprises an encoder that compresses the noisy image into a low-dimensional latent space embedding and a decoder that decompresses this embedding into a denoised image. The latent space presents an information bottleneck that the trained DAE model uses to reject noise and preserve signal, enabling reconstruction of denoised images. The DAE is trained on noise-free images for which the ground truth is known and which are artificially corrupted by noise under a noise model representative of the intended application domain for the trained DAE. The image illustrates a DAE that performs an encoding of a 28 × 28 pixel grayscale (i.e., single channel) image into a 64-channel 8 × 8 latent space embedding under three convolution plus max pooling layers, followed by decoding under three convolutional plus upsampling layers to generate a denoised 28 × 28 pixel image.[40]

neural networks designed to remove noise from an input signal, frequently images.[38] A typical autoencoder is comprised of two distinct components: an encoder and decoder. The encoder compresses a high-dimensional image into a low-dimensional representation. These representations are called latent representations or encodings which the decoder uses to reconstruct the original input image. During training, the DAE is provided with training images that have been artificially corrupted with noise generated by a model representative of the noise expected to be encountered in the particular application domain. A loss function is applied that minimizes the difference between the reconstructed image and the original noise-free image. Intuitively, the training process teaches the DAE to learn a latent space representation that filters out the noise while preserving the underlying signal within the training data and permits the decoder to reconstruct denoised images.[39] The trained DAE model may then be applied to noisy images outside of the training data for which the ground truth is unknown to predictively reconstruct denoised images. The success and generalizability of the trained model are contingent on the training images and noise model being sufficiently representative of the new images to which it is applied, and it is good practice to perform *post hoc* checks that the model has not introduced artifacts or been applied outside of its domain of applicability.

We employ a fully convolutional DAE architecture that permits variable input image sizes to allow for potential variability in training and experimental image sizes.[41] Training and validation sets of 12702 and 3176 images (80/20 random split), respectively, with 28 × 28 dimensions at a batch size of 32 were used for training and validation (Figure 2). We guard against overfitting by employing early stopping based on the validation error on a 20% randomly sampled hold-out validation partition. These images were harvested from the 1CPN MD simulations and contain a diversity of conformations of chromatin fibers at a resolution commensurate with that of a typical ChromSTEM imaging experiment. We use a convolution layer of kernel size (3,3) with 256 output filters and stride 1 employing ReLU activation functions and followed by a max pooling layer of pool size (2,2). We follow this with a second ReLU convolutional layer of kernel size (3,3), 128 output filters, and stride 1 followed by a max pooling layer of pool size (2,2), and finally a third ReLU convolutional layer of kernel size (3,3), 64 output filters, and stride 1 followed by a max pooling layer of pool size (2,2). The output of the third convolutional layer produces a low-dimensional latent space embedding of the image that serves as an information bottleneck designed to preserve the image signal and reject noise. The decoder architecture mirrors the encoder structure, employing three convolutional upsampling layers used to rebuild images to their original dimension. Our network employs a fully convolutional architecture that does not use any fully connected layers and enables its deployment on images of arbitrary size. Given that images comprise single

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

channel grayscale pixels with intensities normalized between [0,1], the binary cross-entropy (BCE) loss function is used

$$\mathrm{BCE} = -\frac{1}{\mathrm{output}_{\mathrm{size}}} \sum_{i=1}^{\mathrm{output\ size}} y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \tag{6}$$

where $\hat{y}_i$ is the output prediction and y_i is the corresponding target value. It has been shown that when training autoencoders on image data, minimizing the BCE loss function facilitates gradient steps in data space from low- to high-probability regions under the data-generation distribution.[42]

We constructed and trained our DAE in TensorFlow using Keras.[43] Training took ∼3 min per epoch on an AMD Ryzen 9 3950X 16-core CPU and Nvidia RTX 3090 GPU card. Training was performed using the Adam algorithm with a learning rate of 1×10^{-3}.[44] We guard against overfitting by employing early stopping based on the validation error on a 20% randomly sampled hold-out validation partition. We explored architectures employing 3−6 convolutional layers, first layer filters ranging from 2×2 to 5×5, and latent spaces bottlenecks ranging from $2 \times 2 \times 12$ to $16 \times 16 \times 128$ but found our result to be relatively insensitive to the precise choice of architecture. The source code for our DAE and training/validation data are available at https://github.com/Ferg-Lab/ChromSTEM-Denoising-Autoencoder.

Denoising Performance. Denoising performance was measured using mean-square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM).[45,46] Mean-square error is the total squared error between pixel intensity differences of the original noise-free image, I, and denoised image, $\hat{I}$, defined as

$$\mathrm{MSE} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} [I_{m,n} - \hat{I}_{m,n}]^2}{MN} \tag{7}$$

where M and N are the number of rows and columns in the image and $M = N = 28$ for our training data. The lower the MSE value, the lower the error. Similarly, PSNR measures the quality of reconstruction of lossy compression by measuring the peak error and is calculated as

$$\mathrm{PSNR} = 10 \log_{10}\left(\frac{R^2}{\mathrm{MSE}}\right) \tag{8}$$

where R is the maximum possible pixel value and typically depends on the bit depth of an image (e.g., for 8-bit images $R = 255$).[47] For PSNR, the higher the value, the better the reconstruction.

Whereas MSE and PSNR calculate absolute errors between pixels, the SSIM index considers degradation as the change of perception in structural information by taking into account three key features: luminance, contrast, and structure. An SSIM value can range from −1, indicating images are structurally different, to +1, indicating they are either the same or very similar, and is defined as

$$\mathrm{SSIM}(x, y) = [l(x, y)]^{\alpha} \times [c(x, y)]^{\beta} \times [s(x, y)]^{\gamma} \tag{9}$$

where

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{10}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{11}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{12}$$

The functions $l(x,y)$, $c(x,y)$, and $s(x,y)$ compare luminance, contrast, and structure between two images x and y, where here we set $x = I$ and $y = \hat{I}$ for our ground-truth and denoised images, respectively.[47] The variables μ_x and μ_y are their respective local means over all pixel values and represent the luminance of each images. Contrast is measured by taking the standard deviation σ_x and σ_y of all pixel values, and σ_{xy} is the cross-covariance of the images. The variables α, β, and γ adjust the relative importance of each feature and are typically set to unity. The constants $C_i = (K_iL)^2$ prevent functions from becoming undefined, where L accounts for pixel value range and is set to unity given that our images are normalized in the range of [0,1]. By convention, we adopt $C_3 = C_2/2$ and set $K_1 = 0.01$ and $K_2 = 0.03$.[45]

Denoising performance metrics such as MSE, PSNR, and SSIM are calculated between a ground-truth image (i.e., noise-free image), I, and its denoised counterpart, $\hat{I}$, produced by the DAE from the artificially noisy image $\tilde{I}$. Given that noise-free ChromSTEM images do not exist to serve as a ground-truth comparison, we rely on power spectral density (PSD) plots to compare raw and denoised experimental image sets. PSD represents the total signal power contributed across the frequency domain of a signal. For images, it measures the strength of the features at different resolutions. This allows for comparison of morphological features and noise in the low- and high-wavelength domains, respectively. We compute the PSD by taking the discrete Fourier transform (DFT) of each image which allows for the decomposition of resolutions

$$F(k, l) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{m,n} \exp\left\{-2\pi i \frac{mk}{M}\right\} \exp\left\{-2\pi i \frac{nl}{N}\right\} \tag{13}$$

where $I_{m,n}$ is a representation of the image in the spatial domain corresponding to the grayscale intensity of the pixel at row (m) and column (n) coordinates, $F(k, l)$ is the representation of the image in the Fourier domain corresponding to the Fourier component at discrete row-wise and column-wise "frequencies" k/M and l/N, and $k = 0, ..., (M - 1)$ and $l = 0, ..., (N - 1)$.[48,49] Since we only consider square images for which $M = N$, we simplify this expression to equalize the row and column frequency components by setting $k = l$ so that

$$F(k) = \sum_{m=0}^{M-1} I_{m,n} \exp\left\{-4\pi i \frac{mk}{M}\right\} \tag{14}$$

The PSD follows from the modulus of the DFT as $P(k) = |F(k)|$.

ChromSTEM Sample Preparation, Imaging, and Reconstruction for A549 Cell Nucleus. Adenocarcinoma human lung epithelial cell line A549 (ATCC Manassas, VA) was cultured in Dulbecco's Modified Eagle Medium (Thermo-Fisher Scientific, Waltham, MA, #11965092) and maintained at 5% CO_2 and 37 °C. All culture media were supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA; #16000044) and penicillin−streptomycin (100

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

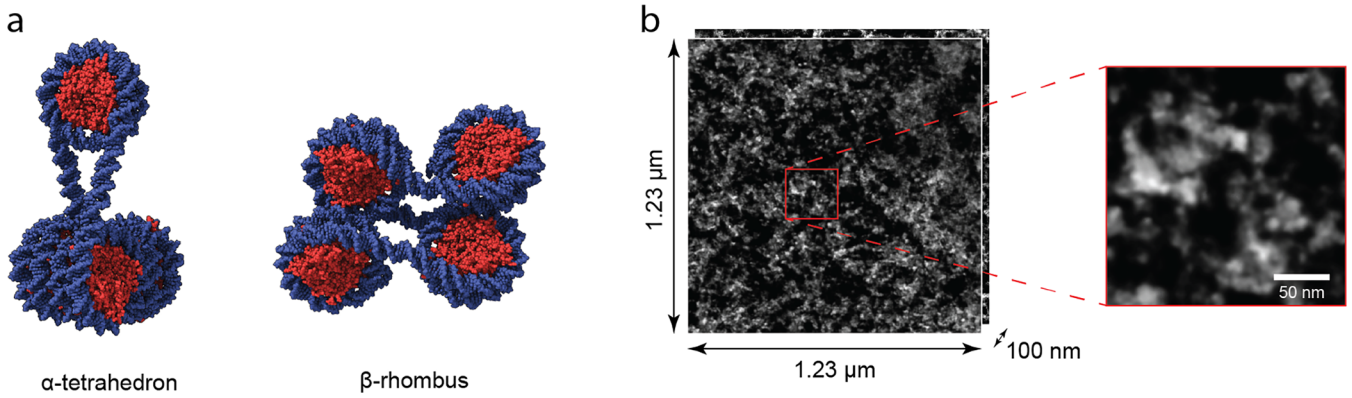


Figure 3. Resolution in dense chromatin regions is obstructed by the intrinsic noise of STEM imaging. (a) The α-tetrahedron and β-rhombus tetranucleosome motifs have been proposed to play a regulatory and epigenetic role in the accessibility of DNA to external cellular machinery. The α-tetrahedron promotes DNA compaction, whereas the β-rhombus results in elongated chomatic structures. Histone proteins are colored in red, and DNA is colored in blue. (b) In this work we employ high-resolution ChromSTEM tomograms comprised of 33 slices at 1.23 μm × 1.23 μm × 100 nm. The structural resolution accessible to experimental ChromSTEM tomograms is limited by the conformational variability of chromatin within chromatin-rich regions, Poisson noise, and the ability of image segmentation approaches to differentiate background and chromatin signal by voxel intensity.

μg/mL; Thermo Fisher Scientific, Waltham, MA; #15140122). The cell line was tested for mycoplasma contamination with Hoechst 33342. Cells were seeded on 35 mm glass-bottom Petri dishes (MatTek Corp.) until approximately 40−50% confluent and were given at least 24 h to adhere to the dish before fixation.

For ChromSTEM sample preparation, the previously published protocol was adapted.[13] A549 cells cultured on the glass-bottom dishes were thoroughly rinsed three times in Hank's balanced salt solution without calcium and magnesium (EMS). A fixation solution (2.5% EM grade glutaraldehyde, 2% paraformaldehyde, 2 mM $CaCl_2$ in 0.1 M sodium cacodylate buffer, pH = 7.4) was prepared. Cells were then fixed at room temperature for 5 min and then replaced with fresh fixative and fixed on ice for 1 h. All the succeeding steps, unless mentioned otherwise, were performed on ice. After fixation, the cells were then washed with 0.1 M sodium cacodylate buffer five times on the ice. The samples were incubated in a blocking buffer (10 mM glycine, 10 mM potassium cyanide in 0.1 M sodium cacodylate buffer, pH = 7.4) for 15 min. Next, the samples were stained with 10 μM DRAQ5 (Thermo Fisher) and 0.1% saponin solution in 0.1 M sodium cacodylate buffer, pH = 7.4 for 10 min. The cells were washed with a blocking buffer twice, and then incubated in the blocking buffer on ice before photobleaching. The blocking buffer was replaced with 2.5 mM of 3−5′-diaminobenzidine (DAB) solution (Sigma-Aldrich) in 0.1 M sodium cacodylate buffer, pH = 7.4, during photobleaching which was performed on a cold stage developed in-house from a wet chamber and equipped with humidity and temperature control.

A continuous epi-fluorescence illumination (150 W xenon lamp) with a Cy5 red tilter with a 100× objective was used to bleach a spot—a random field of view with several cells—on the dish for 7 min on the cold stage. After photobleaching, the cells were washed five times with 0.1 M sodium cacodylate buffer. Reduced osmium solution (EMS) containing 2% osmium tetroxide, 1.5% potassium ferrocyanide, and 2 mM $CaCl_2$ in 0.15 M sodium cacodylate buffer, pH = 7.4, was then used to stain the cells for 30 min on ice. The cells were then washed five times with double-distilled water on ice. Next, serial ethanol dehydration (30%, 50%, 70%, 85%, 95%, 100% twice) was performed on ice, and the last 100% ethanol wash was performed at room temperature. Durcupan resin (EMS) was used for infiltration and embedding. Resin mixture 1 was prepared by mixing (i) 10 mL of Durcupan ACM single-component A, M, epoxy resin, (ii) 10 mL Durcupan ACM single component B, hardener 964, and (iii) 0.15 mL of Durcupan ACM single component D. A 1:1 infiltration mixture containing equal proportions of 100% ethanol and Durcupan resin mixture 1 was used to infiltrate cells for 30 min at room temperature. Next, a 2:1 infiltration mixture containing 5 mL of 100% ethanol and 10 mL of Durcupan resin mixture 1 was used to infiltrate the cells for 2 h at room temperature. Durcupan resin mixture 1 was used to infiltrate the cells at room temperature for 1 h. Resin mixture 2 was prepared by adding 0.2 mL of Durcupan ACM, single component C, accelerator 960 to mixture 1 (10 mL of component A, 10 mL of component B, and 0.15 mL of component D). Durcupan resin mixture 2 was used to infiltrate the cells at 50 °C in a drying oven for 1 h.

The cells were embedded flat with fresh Durcupan resin mixture 2 in BEEM capsules and cured at 60 °C in a drying oven for 48 h. An ultramicrotome (UC7, Leica) was used to prepare 100 nm thick sections that were deposited onto a copper slot grid with carbon/Formvar film. Then, 10 nm colloidal gold fiducial markers were deposited on both sides of the sample. A 200 kV cFEG STEM (HD2300, HITACHI) with HAADF mode was used to collect all images. While keeping the field of view constant, the sample was tilted from −60 to 60° with 2° increments on two roughly perpendicular axes, with a pixel dwell time of ∼5 μs during image acquisition. Each tilt series was aligned with fiducial markers in IMOD and reconstructed using Tomopy with a penalized maximum likelihood for 40 iterations independently.[50,51] The final tomogram was a 3D image size of 1230 × 1230 × 100 nm with a nominal voxel size of 2.9 nm.

RESULTS AND DISCUSSION

Tetranucleosomes are widely considered the building block of the chromatin fiber and have been crystallized and observed in cryo-EM images of longer chromatin fibers.[9] Recent studies have suggested the existence of two tetranucleosome motifs

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

that regulate gene expression—the α-tetrahedron and β-rhombus (Figure 3a).[10,11] Experiments and modeling studies have indicated that these two energetically stable conformations may induce local chromatin compaction (α-tetrahedron) or the formation of elongated aggregates (β-rhombus) and are therefore proposed to play important regulatory and epigenetic roles in the accessibility of DNA to external machinery such as transcription factors.[10,11,52,53] While ChromSTEM has been able to resolve variably packed nucleosomes and linker DNA segments at ∼2 nm spatial resolution, the variation of size, density, and shape of chromatin rich regions can obstruct finer-scale resolution of the structural arrangement of nucleosomes (Figure 3b). The structural resolution is also degraded by Poisson (i.e., shot) noise associated with electron counting statistics and the relatively poorer performance of segmentation (i.e., differentiation of background and chromatin signal by voxel intensity) within chromatin-rich regions relative to regions where nucleosomes are well-separated and have uniform intensity.[16] We develop a machine-learning-assisted computational denoising platform by training a denoising autoencoder (DAE) over coarse-grained molecular dynamics simulations and apply the DAE to *in situ* high-resolution HAADF ChromSTEM microscopy images of chromatin within mammalian cell lines to resolve tetranucleosome motifs.

Testing on Synthetic Data. To validate our trained DAE, we first tested its performance against standard denoising techniques in an application to synthetic ChromSTEM images to which artificial noise was added and the ground truth (i.e., noise-free) images were exactly known. We collected 3000 test images harvested from 1CPN MD simulations of chromatin fibers varying from 150 to 200 nucleosome repeat lengths (NRLs) and comprised of 4−16 nucleosomes and converted these into noise-free images I and noisy images $\tilde{I}$ using eqs 1 and 2. Importantly, the test set data were never exposed to the DAE at any point during their training. We report in Table 1 the denoising performance of our DAE compared to the popular nonlocal means (NLM) and block-matching and 3D filtering (BM3D) techniques.[54,55] Performance is assessed using the mean square error (MSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) metrics that are commonly used to benchmark denoising methods.[46] Better performance is associated with a reduction in cumulative squared error between the compressed and the original image (lower MSE), an increase in the ratio between the maximum possible power of an image and the power of corrupting noise (higher PSNR), and preservation of structural information between the reference and denoised image (higher SSIM). We present in Figure 4 an illustrative example of the application of each of the three denoising approaches to a representative snapshot taken from the 3000 test images.

Table 1. Mean and Standard Deviation for 3000 Synthetic ChromSTEM Test Images Calculated to Compare the Denoising Performance of Our DAE against Nonlocal Means (NLM) and Block-Matching and 3D Filtering (BM3D)[a]

denoiser	MSE	SSIM	PSNR (dB)
NLM	0.011 ± 0.003	0.15 ± 0.04	20 ± 1
BM3D	0.007 ± 0.004	0.55 ± 0.17	22 ± 2
DAE	**0.003 ± 0.001**	**0.83 ± 0.04**	**26 ± 2**

[a]Snapshots were harvested from 1CPN MD simulations of chromatin fibers varying from 150 to 200 nucleosome repeat lengths (NRLs) and comprised of 4−16 nucleosomes and converted into noise-free images I and noisy images $\tilde{I}$ using eqs 1 and 2. Denoising performance is compared using the mean square error (MSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) metrics. The DAE outperforms nonlocal means and BM3D along all three performance metrics (low MSE, high PSNR, high SSIM).

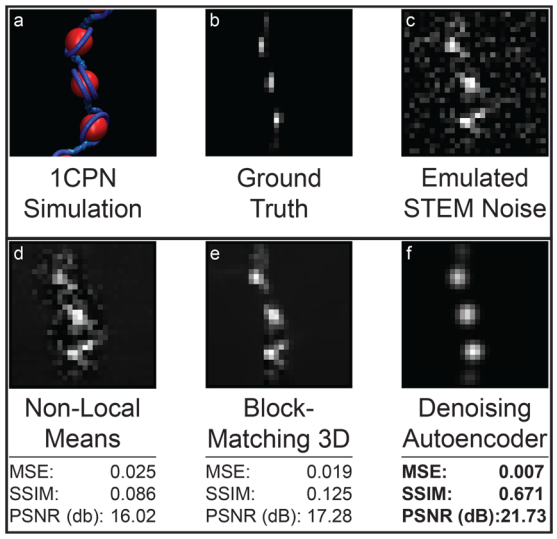


Figure 4. Illustrative example of DAE denoising performance to one selected synthetic ChromSTEM test image harvested from the 1CPN MD simulations. (a) The selected snapshot was harvested from 1CPN MD simulations of chromatin fibers varying from 150 to 200 nucleosome repeat lengths (NRLs) comprised of 4−16 nucleosomes. (b) The noise-free synthetic ChromSTEM image I was constructed from the MD snapshot using eq 1. This constitutes the ground truth image against which we evaluate denoising performance. (c) The noisy image $\tilde{I}$ was generated by adding artificial noise representative of that found in angle annular dark field (HAADF) STEM experiments to the noise-free image using eq 2. The denoised image $\hat{I}$ produced from the noisy test image by (d) nonlocal means (NLM), (e) block-matching and 3D filtering (BM3D), and (f) the DAE. The DAE outperforms NLM and BM3D along all three performance metrics (low MSE, high PSNR, high SSIM) for this particular image and over all 3000 test images (cf. Table 1).

Our DAE performed the best in all three denoising performance metrics (MSE = 0.003, SSIM = 0.83, PSNR = 26 dB), followed by BM3D (MSE = 0.007, SSIM = 0.55, PSNR = 22 dB) and nonlocal means (MSE = 0.011, SSIM = 0.15, PSNR = 20 dB). This represents a 57% improvement in MSE relative to BM3D and 72% improvement over nonlocal means (Table 1). From the example in Figure 4, we can see that our denoising autoencoder is not only able to remove the applied Gaussian and Poisson noise but also has the ability to account for distortions which are typical to STEM experiments by virtue of the fact that it was trained on 1CPN molecular dynamics training data that preserve the physically representative structure of the chromatin strand. Given that denoising autoencoders are inherently lossy compression methods, some fuzzy imaging or loss of information is expected during the encoding process which can lead to broader output signals. The primary goal of our DAE method is to achieve a balance between noise reduction and preservation of structural features in the ChromSTEM images. While it might be possible to

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

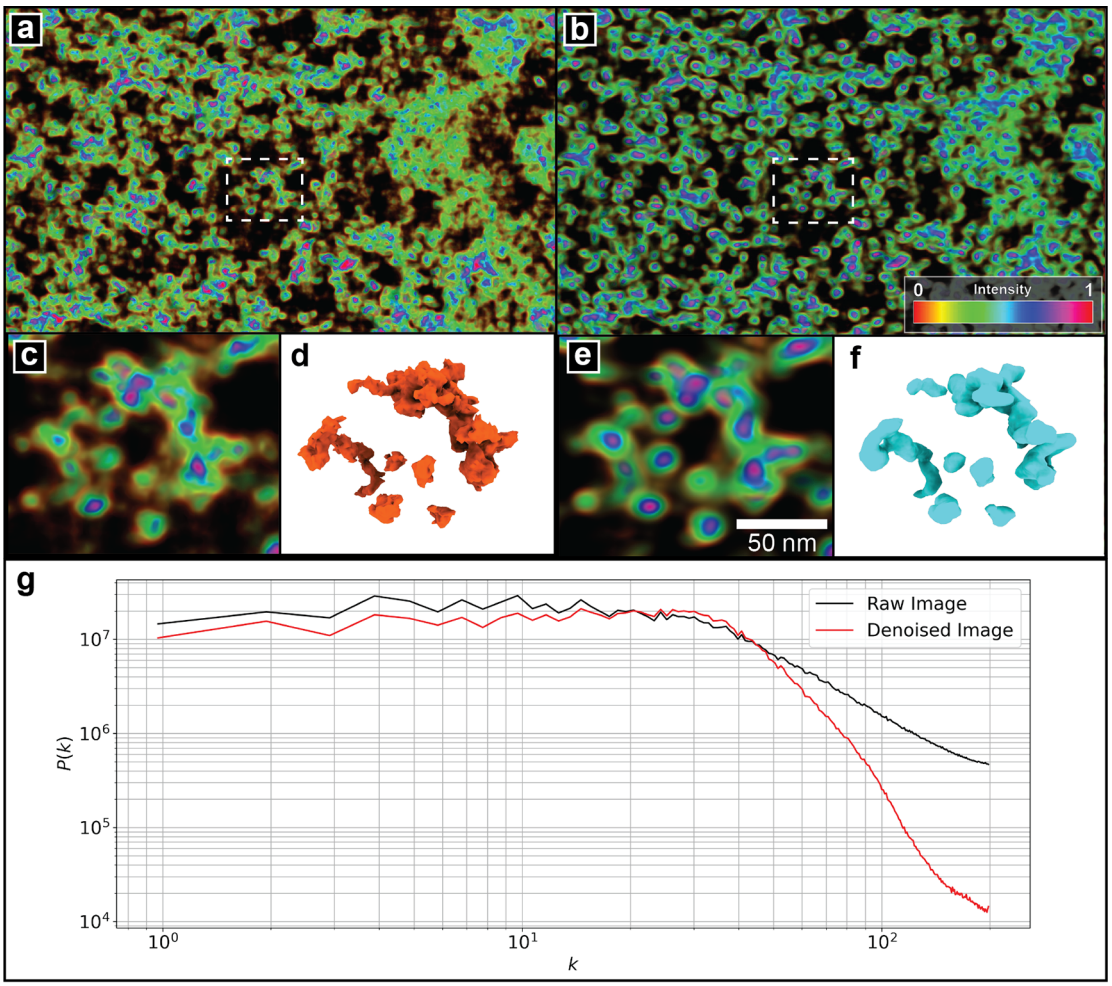


Figure 5. Application of the DAE to denoise the experimental tomogram of an imaged A549 cell. The (a) original experimental image and (b) image generated after passage through the trained DAE. To improve visual clarity and better highlight features of the images, the pixel intensities are normalized to a [0,1] scale and colored by a pseudocolor gradient indicated by the colorbar as opposed to a single grayscale channel. The denoised image achieves improved resolution of nucleosome-level features within chromatin-rich regions of the experimental image. A subsection comparison between the original (c) and denoised experiment (e) shows the reduction of noise and results in a smoother 3D reconstruction of the chromatin fiber from the denoised image (f) compared to the original (d). (g) A comparison of the power spectral density (PSD), $P(k)$, between the raw and denoised images shows the denoised image to preserve the large-scale, low-frequency energy density at small wavenumbers k corresponding to the morphological structure of the chromatin fiber and attenuate the small-scale, high-frequency components at high k that can be primarily attributed to noise.

reduce these broader signals further, doing so could compromise the performance of the DAE or lead to overfitting.

We do observe that although our test does expose the DAE to novel synthetic ChromSTEM images it has not before encountered, they are generated using the same model as the training data. Conversely, the nonlocal means and BM3D approaches are standard algorithms that are not trained over images from a particular domain and are more general-purpose denoising tools. As expected, the DAE appears to have learned to distinguish the physical arrangement of nucleosomes along the chromatin fiber within the physics-based simulation training data from the applied noise model and can use these learned patterns to effectively denoise new synthetic ChromSTEM images that it has not previously encountered. A possible cost of this learning is, of course, that the DAE will likely not serve as a good general-purpose, application-agnostic denoising algorithm in the same manner as nonlocal means and BM3D.

Application to Experimental Data. After validating that our DAE was capable of removing noise while preserving local structural features from our synthetic data set, we move to apply it to experimental ChromSTEM images of chromatin. Figure 5 shows the difference between a raw and denoised experimental tomogram of an imaged human pulmonary adenocarcinoma epithelial cell (A549 cell). A pseudocolor gradient as opposed to a single grayscale channel is employed to display pixel intensity for better visibility and to more clearly highlight the features within the image. Visual inspection of the denoised experiment confirms the ability of our DAE to remove noise and its ability to better resolve nucleosomes within chromatin-dense regions. Closer inspection of a randomly selected region of the denoised image (Figure 5b,e,f) clearly reveals the existence of clusters of a few nucleosomes that previous studies have suggested may play a role in the formation of topologically associated domains (TADs) in chromatin biology and which are much less clearly resolved in the original image (Figure 5a,c,d).[10] We also

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

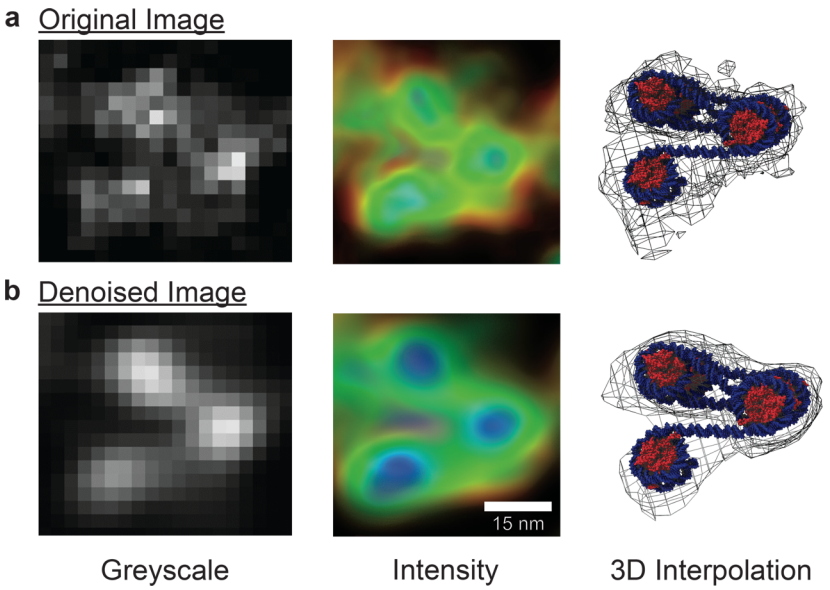


Figure 6. Denoised ChromSTEM images reveal tetranucleosome motifs within a dense chromatin cluster. Analysis of nucleosome clusters extracted from chromatin-rich regions of the (a) raw experimental tomogram and (b) after passing through the DAE. The denoised image clearly shows the presence of α-tetrahedron motifs that are difficult to discern in the raw image. Using Chimera, we construct a prototypical tetranucleosome motif (PDB: 1KX5) within the extracted volume of our denoised tomogram and find an optimal fit with an average high correlation score of 0.87.[56] The construction of the 3D interpolation from the 2D imaging slices is computationally expensive but can, in principle, be extended to large sections of chromatin using high-performance computing resources.

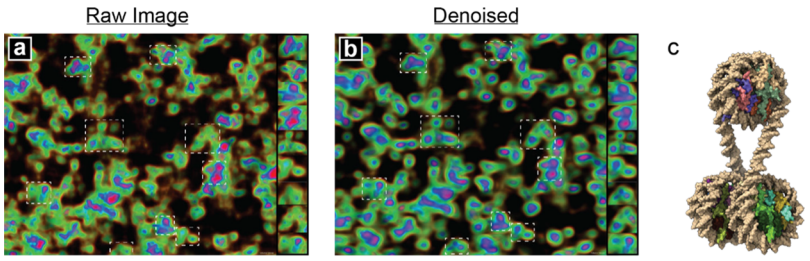


Figure 7. Denoised ChromSTEM images reveal tetranucleosomes motifs within dense chromatin clusters. Analysis of nucleosome clusters extracted from chromatin-rich regions within a 200 × 200 nm^2 section of the (a) raw experimental tomogram and (b) after passing through the DAE. The denoised image clearly shows the presence of (c) α-tetrahedron motifs that are difficult to discern in the raw image. We find no evidence for β-rhombus motifs or for the 30 nm fiber.

compare the power spectral densities (PSDs) of the raw and denoised image stacks (Figure 5g). We see good agreement of the PSD at lower wavenumbers, which correspond to the large-scale (i.e., low-frequency) structural and morphological features of the image. At higher wavenumbers, the PSD of the denoised image exhibits a linear decrease relative to the raw image, which can be interpreted as the attenuation of small-scale (i.e., high-frequency) noise in the experimental image. Taken together, these results indicate that the important structural signal within the experimental ChromSTEM image is preserved by our denoising approach and produces superior resolution of nucleosome-level features within the chromatin-rich regions of the image.

To determine whether these small nucleosomal clusters are comprised of either of the two recently identified folding motifs (α-tetrahedron or β-rhoumbus), we visually inspect a number of nucleosome clusters extracted from chromatin-rich regions within a 50 × 50 nm section of the experimental tomogram (Figure 6). It is challenging to discern from inspection of the raw image, but after passage through the DAE it is visually apparent that these chromatin-dense regions are primarily composed of tetranucleosome motifs (Figure 7). To quantify our assertion, we construct a density map from our denoised STEM image stack and fit a prototypical α-tetrahedral tetranucleosome folding motif reconstructed from a single atomic nucleosome structure (PDB: 1KX5).[23] To find an optimal fit, the cross-correlation coefficient (CCC) score was used to maximize the fit of a simulated map from the atomic structure and our volume map using the density mapping algorithm from the Chimera software.[56] We find an improved optimal fit with an average high correlation score of 0.87 versus a correlation score of 0.85 for the original tomogram (Figure 6). Though comparatively small, incremental quantitative improvements can provide insightful details about the chromatin structure. Detecting and quantifying tetranucleosome motifs in raw and denoised images remains an important task and a significant challenge in the field, and expert experimentalists are crucial for interpreting results due to their deep understanding of the biological context and ability to assess image quality and identify relevant features.[57,58] Our denoising method improved the detection of tetranucleosome motifs primarily based on visual cues, resulting in a more accurate representation of chromatin structure in denoised chromSTEM images (Figure 7).

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

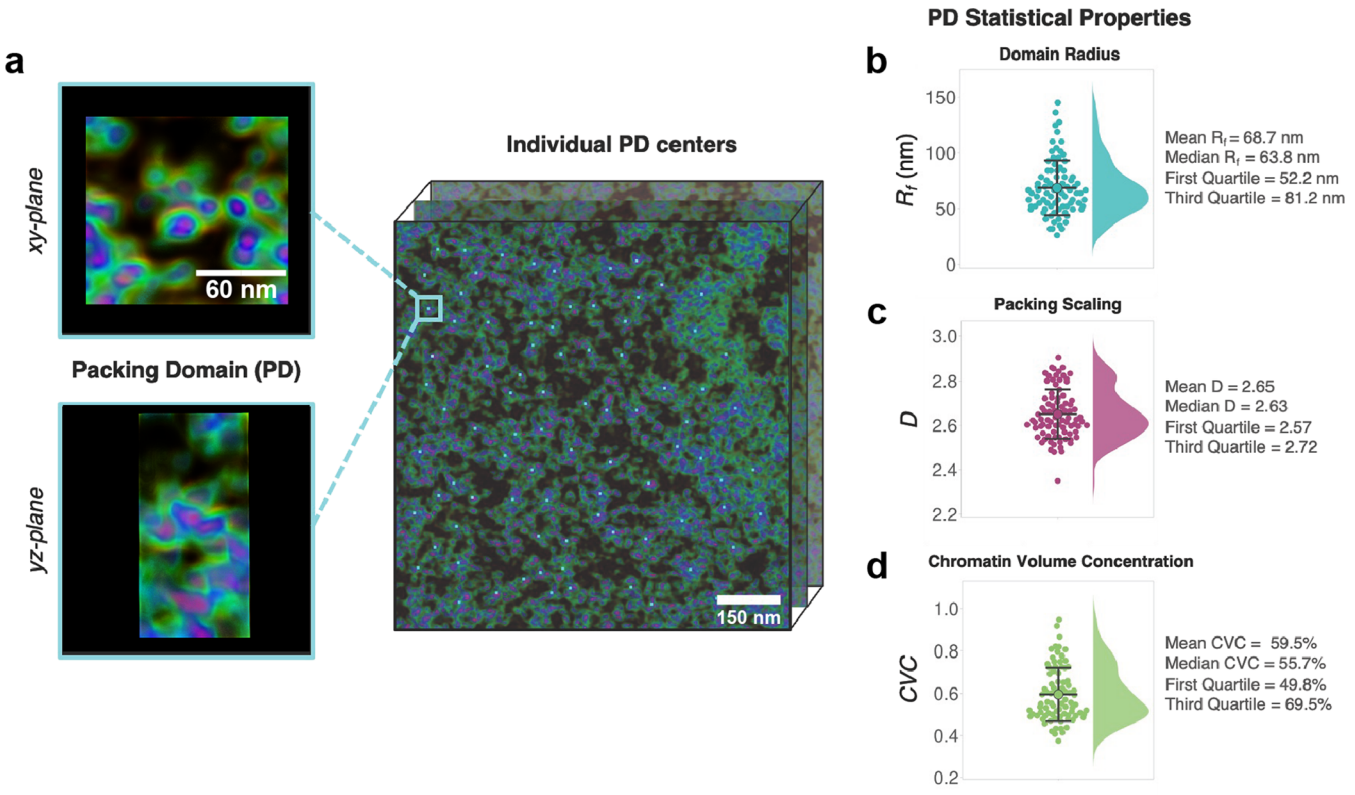


Figure 8. Structural analysis of chromatin-rich packing domains from the DAE-denoised A549 3D ChromSTEM tomogram. (a) A 3D conformation of a packing domain identified from the denoised ChromSTEM tomogram (Figure 5b). Statistical distribution of (b) domain size R_f, (c) packing scaling exponent D, and (d) cluster volume concentration CVC, over the 85 chromatin-rich packing domains identified from the denoised ChromSTEM tomogram. Denoising enables identification of ∼12% more domains and domains more closely associated in space relative to analysis of the raw 3D ChromSTEM tomograms.

These tetranucleosome motifs are known to promote DNA compaction and lead to chromatin condensation, and the preponderance of these structural elements observed within chromatin-dense regions is consistent with prior experiment and simulation.[10,11] In contrast, we do not observe any zigzag β-rhombus motifs or find any evidence for the formation of the postulated 30 nm fiber.[59] These results support a model in which the *in situ* structural organization of chromatin within chromatin-dense regions in the cell is not a 30 nm fiber, but rather largely composed of smaller tetranucleosome motifs.

Identifying Packing Domains and Their Statistical Properties From Denoised ChromSTEM Stack. The denoised images produced by the DAE enable more robust resolution of chromatin-rich packing domains and improved estimation of statistical distribution of their structural properties such as size, packing scaling exponents, and chromatin volume concentration. We first describe these analyses in the context of the raw ChromSTEM images and then demonstrate how our statistical resolution improves within the denoised images.

Considering first the raw 3D ChromSTEM tomogram presented in Figure 5a, we extracted 76 chromatin-rich packing domains and then subjected them to structural analysis to determine the distribution of domain sizes R_f. To do so, we adopted two complementary definitions of domain size. First, we identified the centroid of each domain by creating a local chromatin intensity map by applying Gaussian filtering and local contrast enhancement to the grayscale ChromSTEM z-stacks. We appeal to the fact that ChromSTEM intensity is approximately linearly proportional to mass to fit a scaling law between mass M and distance r from the centroid of each domain.[15] Following classical power-law polymer scaling relations, mass and distance are expected to be related as $M(r) \propto r^D$, where M is defined as the integrated mass (i.e., intensity) lying at a particular radial distance r from the domain centroid and D is the packing scaling exponent for the polymer that is anticipated to be approximately constant over a particular range of length scales.[60] We computed best-fit values of the packing scaling exponent D by fitting power laws over the range of $[0,r]$ at increasing r and defined the domain size $R_f^{(1)}$ as the distance r at which we observe more than 5% deviation from the best-fit power law. This demarcates the length scale at which a single power-law relationship no longer holds and constitutes our first definition of R_f (Figure S1a,b). Second, we calculated the radial density profile of chromatin as a function of distance r from the centroid of the domain. This profile is expected to monotonically decrease until the distance r reaches the boundary of the domain and then increase again as it begins to encroach upon a neighboring domain (Figure S1c). The minimum in the radial density profile defines our second definition of domain size $R_f^{(2)}$. Finally, we defined the domain size $R_f = \min(R_f^{(1)}, R_f^{(2)})$. We observe that the two complementary definitions of domain size over which we take the minimum are necessary to properly account for the environment in which the domains may be found: in chromatin-poor environments where the domains are isolated, we expect domain-size to be dictated by the mass distribution of the single domain under consideration and $R_f^{(1)} < R_f^{(2)}$; in

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

chromatin-rich environments, we anticipate $R_f^{(2)} < R_f^{(1)}$ and domain size should be more appropriately defined as an multibody property that defines the boundary between domains.

Having defined R_f and D for each domain, we compute the chromatin volume concentration, CVC, which correlates with the binding efficiency of transcriptional reactants and is defined as the fraction of volume occupied by chromatin.[13,15] The CVC was calculated as the total number of nonzero voxels over the total number of voxels per domain.[15] The distributions of these three quantities for the 76 chromatin-rich domains extracted from the raw A549 3D ChromSTEM tomograms are presented in Figure S2, for which we report means and standard deviations of $R_f = 71 \pm 26$ nm, $D = 2.46 \pm 0.18$, and CVC = 42 ± 14%. We previously demonstrated that chromatin forms spatially well-defined higher-order domain structures with radii ranging between an interquartile range of 60−90 nm in A549 cells and observe that our present measure of mean domain size lies squarely within this range.[15]

A concern of applying this structural analysis to the raw ChromSTEM tomograms is the introduction of errors into both the definition of the domains and their structural properties due to the noise inherent in the experimental images. Accordingly, we repeated this analysis for the DAE denoised 3D ChromSTEM tomogram presented in Figure 5b. In doing so, our procedure identified 85 chromatin-rich packing domains, 9 more than were identified in the raw images. An analysis reveals that application of the domain identification procedure to the denoised image enables identification of more domains and better resolves domains more closely packed in space (Figure S3). The improvement in signal-to-noise ratio in the denoised tomogram appears to assist in the identification of domain centers that cannot be resolved in the raw tomogram and which are confirmed by manual visual analysis. To assess the possibility of introducing artifacts through the DAE denoising, we present in Figure 8 the statistical analysis of R_f, D, and CVC over the 85 denoised ChromSTEM domains. The mean reported values of $R_f = 69 \pm 24$ nm, $D = 2.65 \pm 0.11$, and CVC = 60 ± 13% are all in good agreement with the analysis of both the raw ChromSTEM images and our prior analyses[15] but are now based on better statistics enabled by the identification of ∼12% more domains in the denoised images.

CONCLUSIONS

By leveraging molecular dynamics and machine-learning approaches, we constructed and trained a denoising autoencoder (DAE) capable of removing noise commonly found in scanning transmission electron microscopy tomography with ChromEM staining (ChromSTEM) imaging. The model is trained over physics-based coarse-grained molecular dynamics simulations using the 1CPN model and learns to distinguish the signal from ground truth chromatin structures from artificial noise mimicking the noise profile inherent to experimental STEM imaging. In tests on synthetic ChromSTEM images generated by molecular simulations for which the ground truth is exactly known, the training outperforms standard denoising approaches, offering a 57% improvement in the mean squared error relative to block-matching and 3D filtering and a 72% improvement over nonlocal means. In applications to *in situ* experimental ChromSTEM images of chromatin within human pulmonary adenocarcinoma epithelial cells (A549 cells), we demonstrate that the DAE eliminates high-frequency noise while preserving the large-scale signal characterizing the chromatin organizational structure. The denoised images enable identification of tetranucleosome motifs at a resolution inaccessible within the raw images and expose the α-tetrahedron as the predominant organizational subunit within chromatin-dense regions in the cell and which have been suggested to play a role in chromatin compaction and regulation of gene expression. Notably, we find no evidence for the presence of β-rhombus tetranucleosome motifs or for the 30 nm fiber. The denoised images also permit the identification of ∼12% more chromatin-rich packing domains that are obscured by noise within the raw images, enabling improved statistical resolution of the distribution of domain sizes, packing scaling exponents, and chromatin volume concentrations without apparently introducing statistical artifacts. The domain size distributions are consistent with, but have higher statistical resolution and smaller uncertainties than, our prior analyses.[15]

The nucleosome motifs exposed by this approach enable a new understanding and insight into the small-scale structural organization of chromatin within the cell and how these structures can influence DNA accessibility and gene regulation. The present work focused primarily on the analysis of tetranucleosome motifs, but in future work we hope to expand our focus to smaller di- and trinucleotide motifs. We anticipate that the approaches reported in this study may be applied to ChromSTEM imaging to advance our understanding of how stress and epigenetic factors affect chromatin conformation and gene regulation and may also be applied to other imaging techniques such as cryogenic electron microscopy (cryo-EM). Our study also exemplifies a generic paradigm wherein experimental imaging and theoretical modeling may be bridged via machine-learning approaches to enable high-resolution exploration of structural organization within biological systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.3c00178.

Mass scaling and radial chromatin density of chromatin domains, morphology of A549 cells characterized in a noisy tomograph, and denoised tomographs revealing additional domains (PDF)

AUTHOR INFORMATION

Corresponding Authors

Vadim Backman − *Department of Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States; Department of Applied Physics, Northwestern University, Evanston, Illinois 60208, United States*; Email: v-backman@northwestern.edu

Juan J. de Pablo − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States*; orcid.org/0000-0002-3526-516X; Email: depablo@uchicago.edu

Andrew L. Ferguson − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States*; orcid.org/0000-0002-8829-9726; Email: andrewferguson@uchicago.edu

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212

Authors

Walter Alvarado − *Biophysical Sciences, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-9027-9951

Vasundhara Agrawal − *Department of Biomedical Engineering, Northwestern University, Evanston, Illinois 60208, United States*

Wing Shun Li − *Department of Applied Physics, Northwestern University, Evanston, Illinois 60208, United States*

Vinayak P. Dravid − *Department of Materials Sciences and Engineering, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-6007-3063

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.3c00178

Notes

The authors declare the following competing financial interest(s): A.L.F. is a co-founder and consultant of Evozyne, Inc. and a co-author of US Patent Applications 16/887,710 and 17/642,582, US Provisional Patent Applications 62/853,919, 62/900,420, 63/314,898, and 63/479,378 and International Patent Applications PCT/US2020/035206 and PCT/US2020/050466.

■ ACKNOWLEDGMENTS

We thank Dr. Yue Li, Eric Roth, and Dr. Reiner Bleher at the Biological-Cryogenic Electron Microscopy (BioCryo) facility at Northwestern University for their assistance in Chrom-STEM staining, sectioning, and imaging. We also thank Dr. Tobin Sosnick, Dr. Rebecca Willett, Aria Coraor, Soren Kyhl, Eric Schultz, Yiheng Wu, Mike Jones, Fabian Bylehn, and Kirill Shmilovich for their helpful discussions. This study was primarily supported by NSF Grant EFRI CEE 1830969. The authors also gratefully acknowledge support from NSF grant EFMA-1830961, NIH grants U54CA268084, R01CA228272, and R01CA225002, and philanthropic support from Rob and Kristin Goldman. This work was completed in part with resources provided by the University of Chicago Research Computing Center. We gratefully acknowledge computing time on the University of Chicago high-performance GPU-based cyberinfrastructure supported by the National Science Foundation under Grant No. DMR-1828629. This work made use of the BioCryo facility of Northwestern University's NUANCE Center, which has received support from the SHyNE Resource (NSF ECCS-2025633), the IIN, and Northwestern's MRSEC program (NSF DMR-1720139).

■ REFERENCES

(1) Van Holde, K. E. *Chromatin*; Springer Science & Business Media: 2012.

(2) McGinty, R. K.; Tan, S. Nucleosome Structure and Function. *Chem. Rev.* **2015**, *115*, 2255−2273.

(3) Warnecke, T.; Becker, E. A.; Facciotti, M. T.; Nislow, C.; Lehner, B. Conserved Substitution Patterns around Nucleosome Footprints in Eukaryotes and Archaea Derive from Frequent Nucleosome Repositioning through Evolution. *PLOS Computational Biology* **2013**, *9*, e1003373.

(4) Razin, S. V.; Gavrilov, A. A. Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. *epigenetics* **2014**, *9*, 653−657.

(5) Finch, J. T.; Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 1897−1901.

(6) McGhee, J. D.; Rau, D. C.; Charney, E.; Felsenfeld, G. Orientation of the nucleosome within the higher order structure of chromatin. *Cell* **1980**, *22*, 87−96.

(7) Tremethick, D. J. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **2007**, *128*, 651−654.

(8) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **2005**, *436*, 138−141.

(9) Song, F.; Chen, P.; Sun, D.; Wang, M.; Dong, L.; Liang, D.; Xu, R.-M.; Zhu, P.; Li, G. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **2014**, *344*, 376−380.

(10) Ohno, M.; Ando, T.; Priest, D. G.; Kumar, V.; Yoshida, Y.; Taniguchi, Y. Sub-nucleosomal genome structure reveals distinct nucleosome folding motifs. *Cell* **2019**, *176*, 520−534.

(11) Alvarado, W.; Moller, J.; Ferguson, A. L.; de Pablo, J. J. Tetranucleosome Interactions Drive Chromatin Folding. *ACS Central Science* **2021**, *7*, 1019−1027.

(12) Annunziato, A. DNA packaging: nucleosomes and chromatin. *Nature Education* **2008**, *1*, 26.

(13) Ou, H. D.; Phan, S.; Deerinck, T. J.; Thor, A.; Ellisman, M. H.; O'shea, C. C. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **2017**, *357*, eaag0025.

(14) Li, Y.; Eshein, A.; Virk, R. K.; Eid, A.; Wu, W.; Frederick, J.; VanDerway, D.; Gladstein, S.; Huang, K.; Shim, A. R.; Anthony, N. M.; Bauer, G. M.; Zhou, X.; Agrawal, V.; Pujadas, E. M.; Jain, S.; Esteve, G.; Chandler, J. E.; Nguyen, T.-Q.; Bleher, R.; de Pablo, J. J.; Szleifer, I.; Dravid, V. P.; Almassalha, L. M.; Backman, V. Nanoscale chromatin imaging and analysis platform bridges 4D chromatin organization with molecular function. *Science Advances* **2021**, *7*, eabe4310.

(15) Li, Y.; Agrawal, V.; Virk, R. K.; Roth, E.; Li, W. S.; Eshein, A.; Frederick, J.; Huang, K.; Almassalha, L.; Bleher, R.; Carignano, M. A.; Szleifer, I.; Dravid, V. P.; Backman, V. Analysis of three-dimensional chromatin packing domains by chromatin scanning transmission electron microscopy (ChromSTEM). *Sci. Rep.* **2022**, *12*, 12198.

(16) Seki, T.; Ikuhara, Y.; Shibata, N. Theoretical framework of statistical noise in scanning transmission electron microscopy. *Ultramicroscopy* **2018**, *193*, 118−125.

(17) Ziatdinov, M.; Maksov, A.; Kalinin, S. V. Learning surface molecular structures via machine vision. *npj Computational Materials* **2017**, *3*, 1−9.

(18) Lequieu, J.; Cordoba, A.; Moller, J.; de Pablo, J. J. 1CPN: A coarse-grained multi-scale model of chromatin. *J. Chem. Phys.* **2019**, *150*, 215102.

(19) Moller, J.; Lequieu, J.; de Pablo, J. J. The free energy landscape of internucleosome interactions and its relation to chromatin fiber structure. *ACS Central Science* **2019**, *5*, 341−348.

(20) Eltsov, M.; MacLellan, K. M.; Maeshima, K.; Frangakis, A. S.; Dubochet, J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19732−19737.

(21) Maeshima, K.; Hihara, S.; Eltsov, M. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr. Opin. Cell Biol.* **2010**, *22*, 291−297.

(22) Huynh, V. A.; Robinson, P. J.; Rhodes, D. A method for the in vitro reconstitution of a defined 30 nm chromatin fibre containing stoichiometric amounts of the linker histone. *Journal of molecular biology* **2005**, *345*, 957−968.

(23) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. *J. Mol. Biol.* **2002**, *319*, 1097−1113.

(24) Schwenker, E. *Image Matching for Computer Vision in Atomic-Resolution Electron Microscopy*. 2020; https://github.com/MaterialEyes/atomagined.
(25) Ophus, C. A fast image simulation algorithm for scanning transmission electron microscopy. *Advanced Structural and Chemical Imaging* **2017**, *3*, 1−11.
(26) Pryor, A.; Ophus, C.; Miao, J. A streaming multi-GPU implementation of image simulation algorithms for scanning transmission electron microscopy. *Advanced Structural and Chemical Imaging* **2017**, *3*, 1−14.
(27) Sohlberg, K.; Pennycook, T. J.; Zhou, W.; Pennycook, S. J. Insights into the physical chemistry of materials from advances in HAADF-STEM. *Phys. Chem. Chem. Phys.* **2015**, *17*, 3982−4006.
(28) Kübel, C.; Voigt, A.; Schoenmakers, R.; Otten, M.; Su, D.; Lee, T.-C.; Carlsson, A.; Bradley, J. Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications. *Microscopy and Microanalysis* **2005**, *11*, 378−400.
(29) Binev, P.; Blanco-Silva, F.; Blom, D.; Dahmen, W.; Lamby, P.; Sharpley, R.; Vogt, T. In *Modeling Nanoscale Imaging in Electron Microscopy*; Vogt, T., Dahmen, W., Binev, P., Eds.; Springer US: 2012; pp 127−145.
(30) Buban, J. P.; Ramasse, Q.; Gipson, B.; Browning, N. D.; Stahlberg, H. High-resolution low-dose scanning transmission electron microscopy. *Journal of Electron Microscopy* **2010**, *59*, 103−112.
(31) Ercius, P.; Alaidi, O.; Rames, M. J.; Ren, G. Electron Tomography: A Three-Dimensional Analytic Tool for Hard and Soft Materials Research. *Adv. Mater.* **2015**, *27*, 5638−5663.
(32) Midgley, P.; Weyland, M. 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography. *Ultramicroscopy* **2003**, *96*, 413−431.
(33) Dahmen, T.; Kohr, H.; Lupini, A. R.; Baudoin, J.-P.; Kübel, C.; Trampert, P.; Slusallek, P.; de Jonge, N. Combined Tilt- and Focal-Series Tomography for HAADF-STEM. *Microscopy Today* **2016**, *24*, 26−31.
(34) Luther, P. K. In *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*; Frank, J., Ed.; Springer New York: 2006; pp 17−48.
(35) Mevenkamp, N.; Binev, P.; Dahmen, W.; Voyles, P. M.; Yankovich, A. B.; Berkels, B. Poisson noise removal from high-resolution STEM images based on periodic block matching. *Advanced Structural and Chemical Imaging* **2015**, *1*, 1−19.
(36) Jondral, F. K. White gaussian noise−models for engineers. *Frequenz* **2018**, *72*, 293−299.
(37) Maraghechi, S.; Hoefnagels, J. P.; Peerlings, R. H.; Geers, M. G. Correction of scan line shift artifacts in scanning electron microscopy: An extended digital image correlation framework. *Ultramicroscopy* **2018**, *187*, 144−163.
(38) Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research* **2010**, *11*, 3371−3408.
(39) Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th international conference on Machine learning*, 2008; pp 1096−1103.
(40) JGraph, Diagrams.net. 2021; https://github.com/jgraph/drawio.
(41) Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
(42) Creswell, A.; Arulkumaran, K.; Bharath, A. A. *On denoising autoencoders trained to minimise binary cross-entropy*. arXiv preprint arXiv:1708.08487, 2017.
(43) Chollet, F. *Keras*. https://keras.io, 2015.
(44) Kingma, D. P.; Ba, J. *Adam: A Method for Stochastic Optimization*. 2014; https://arxiv.org/abs/1412.6980.
(45) Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **2004**, *13*, 600−612.
(46) Fan, L.; Zhang, F.; Fan, H.; Zhang, C. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art* **2019**, *2*, 1−12.
(47) Horé, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. *20th International Conference on Pattern Recognition*, 2010; pp 2366−2369.
(48) Solomon, C.; Breckon, T. *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*; Wiley: 2011.
(49) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357−362.
(50) Kremer, J. R.; Mastronarde, D. N.; McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **1996**, *116*, 71−76.
(51) Gürsoy, D.; De Carlo, F.; Xiao, X.; Jacobsen, C. TomoPy: a framework for the analysis of synchrotron tomographic data. *Journal of Synchrotron Radiation* **2014**, *21*, 1188−1193.
(52) Ekundayo, B.; Richmond, T. J.; Schalch, T. Capturing Structural Heterogeneity in Chromatin Fibers. *J. Mol. Biol.* **2017**, *429*, 3031−3042.
(53) Takizawa, Y.; Ho, C.-H.; Tachiwana, H.; Matsunami, H.; Kobayashi, W.; Suzuki, M.; Arimura, Y.; Hori, T.; Fukagawa, T.; Ohi, M. D.; Wolf, M.; Kurumizaka, H. Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. *Structure* **2020**, *28*, 44−53.e4.
(54) Buades, A.; Coll, B.; Morel, J.-M. Non-local means denoising. *Image Processing On Line* **2011**, *1*, 208−212.
(55) Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising with block-matching and 3D filtering. *Image processing: algorithms and systems, neural networks, and machine learning*; SPIE: 2006; pp 354−365
(56) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera−a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(57) Cardona, A.; Tomancak, P. Current challenges in open-source bioimage informatics. *Nat. Methods* **2012**, *9*, 661−665.
(58) Meijering, E.; Carpenter, A. E.; Peng, H.; Hamprecht, F. A.; Olivo-Marin, J.-C. Imagining the future of bioimage analysis. *Nature biotechnology* **2016**, *34*, 1250−1255.
(59) Grigoryev, S. A.; Woodcock, C. L. Chromatin organization - The 30nm fiber. *Exp. Cell Res.* **2012**, *318*, 1448−1455.
(60) Pethrick, R. *Polymer physics*; Rubinstein, M., Colby, R. H., Eds.; Oxford University Press: 2003; p 440.

https://doi.org/10.1021/acscentsci.3c00178
ACS Cent. Sci. 2023, 9, 1200−1212