

Supporting Information for:

# Uncovering the Conformational Distribution of a Small Protein by Nanoparticle-Aided Cryo-Electron Microscopy Sampling

*Hosung Ki,<sup>†,‡,§</sup> Junbeom Jo,<sup>†,‡,§,@</sup> Youngmin Kim,<sup>†,‡,§</sup> Tae Wu Kim,<sup>†,‡,§,○</sup> Changin Kim,<sup>†,‡,§</sup> Yeeun Kim,<sup>||</sup> Chang Woo Kim,<sup>†</sup> Srinivasan Muniyappan,<sup>†,‡,§</sup> Sang Jin Lee,<sup>†,‡,§</sup> Yonggwan Kim,<sup>†,‡,§</sup> Ho Min Kim,<sup>⊥,#</sup> Yongsoo Yang,<sup>||</sup> Young Min Rhee,<sup>†</sup> and Hyotcherl Ihee<sup>\*,†,‡,§</sup>*

<sup>†</sup>Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

<sup>‡</sup>KI for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

<sup>§</sup>Center for Nanomaterials and Chemical Reactions, Institute for Basic Science (IBS), Daejeon 34141, Republic of Korea

<sup>||</sup>Department of Physics, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

<sup>⊥</sup>Center for Biomolecular & Cellular Structure, Institute for Basic Science (IBS), Daejeon 34126, Republic of Korea

<sup>#</sup>Graduate School of Medical Science & Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

\*Corresponding Author

Hyotcherl Ihee: (E-mail) [hyotcherl.ihee@kaist.ac.kr](mailto:hyotcherl.ihee@kaist.ac.kr) (Tel) +82-42-350-2844

## Supporting Methods

**Supporting Method S1. Determination of the distribution of projection angle,  $\theta_{Au-Au}$ .** In a cryo-EM measurement, the target molecules or particles in a specimen often prefer a particular angular orientation, i.e., adopt a preferred orientation, rather than being evenly distributed over the entire angular orientations<sup>1</sup>. For an accurate analysis of the experimentally obtained distribution of the projected interparticle distance,  $\mathbf{a}_p$ , it is vital to clearly determine the distribution of the angular orientation,  $\theta_{Au-Au}$ , prior to the analysis because such preferred orientation significantly affects the distribution of the projected extra distance,  $\mathbf{c}_p + \mathbf{d}_p$ , and thus the distribution of the actual intersite distance,  $\mathbf{b}$ , reconstructed with the distribution of  $\mathbf{c}_p + \mathbf{d}_p$ .

The distribution of  $\theta_{Au-Au}$  can be determined by directly measuring the orientational distribution with a multitilt analysis, i.e., collecting tilt-series of the projection images at two or more known tilt angles. For a protein labeled with two particles, theoretically,  $\theta_{Au-Au}$  of the sample can be accurately determined by measuring  $\mathbf{a}_p$ s at two or more known tilt angles. Here, we used such a multitilt analysis to determine the distribution of  $\theta_{Au-Au}$ . We prepared the G1C-G45C double mutant of cyt-c labeled with two nanogold particles. The vitrification procedure of folded G1C-G45C mutant for multitilt analysis was the same as described in the Supporting Method S4. Tilt-series images were measured using a Thermo Fisher Glacios (200 kV) transmission electron microscope. Data were collected using a low dose mode (20 e/Å<sup>2</sup>s per image). The samples were measured at three different tilt angles (0°, 30°, -30°) for the multitilt analysis. Each image was acquired at a magnification of 120,000 x under the defocus of up to 1 μm. Exposure time of 1 s was used for acquisition of all images. The total electron dose of the whole tilt series ranged from 40 to 140 e/Å<sup>2</sup>s. We calculated  $\theta_{Au-Au}$  from acquired tilt-series images using the following procedure. First, among the nanogold particles observed in the projection images, we distinguished

the nanoparticle pairs attached to cyt-c by applying the pair selection criteria described in the “Processing of cryo-EM images” section of the Supporting Method (Supporting Method S5). The positions of each pair of nanoparticles were tracked in the tilt-series images. We defined  $(x_{j,n}, y_{j,n})$  as the positions of the  $j$ -th nanoparticles in the  $n$ -th tilt-series projection image measured with a tilt angle  $\theta_n$ . Note that  $j$  ranges from 1 to 2 (nanoparticle pair) and  $n$  ranges from 1 to 2 or 3 (two or three tilt angles) in our case. The first angle (angle index  $n = 1$ ) was chosen as the reference angle. Then, the projected position of the nanoparticle at the tilt angle  $\theta_n$  can be calculated from the 3D nanoparticle position  $(x_{j,1}, y_{j,1}, z_{j,1})$  at the reference angle as follows:

$$\begin{pmatrix} x_{j,n}^{calc} \\ y_{j,n}^{calc} \end{pmatrix} = \text{Proj} \left( R(\theta_n, \mathbf{w}) \begin{pmatrix} x_{j,1} \\ y_{j,1} \\ z_{j,1} \end{pmatrix} \right) + \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \quad (\text{S1})$$

where  $(x_{j,n}^{calc}, y_{j,n}^{calc})$  is the calculated projected  $(x, y)$  positions of the  $j$ -th particle at the  $n$ -th tilt angle, Proj is the projection operator onto the  $xy$  plane,  $R(\theta_n, \mathbf{w})$  is the 3D rotation matrix for the rotation by the tilt angle  $\theta_n$  around the rotation axis  $\mathbf{w}$ , and  $X_n$  and  $Y_n$  are arbitrary translation of the projection image in  $x$  and  $y$  directions which results from a translation of the specimen accompanying the tilting of the specimen. Note that  $\mathbf{w}$  was pre-determined by microscope axis calibration. Since  $x_{j,1}$  and  $y_{j,1}$  are measured values, we only need to determine  $z_{j,1}$  for fully retrieving the 3D positions of the nanoparticles. Therefore, we performed a least-square fitting by minimizing the squared residue  $S$  shown in the following equation:

$$S = \sum_{j=1}^2 \sum_{n=2}^N ((x_{j,n}^{calc} - x_{j,n})^2 + (y_{j,n}^{calc} - y_{j,n})^2) \quad (\text{S2})$$

where  $N$  is the number of tilt angles used to collect the tilt-series projection images ( $N = 2$  or  $3$  in our analysis). Through the least-square fitting, the optimal  $X_n$ ,  $Y_n$ , and  $z_{j,1}$  that minimize  $S$  can be found. It determines the 3D nanoparticle positions of each nanoparticle at the reference angle (i.e.,  $x_{j,1}, y_{j,1}, z_{j,1}$ ), from which the angle  $\theta_{Au-Au}$  between the  $z$  axis and the vector that connects the two

nanogold particles within the pair can be calculated. We determined  $\theta_{Au-Au}$  for 57 pairs of nanogold particles labeled at G1C-G45C mutant of cyt-c in a folded state.

The resulting distribution of  $\theta_{Au-Au}$  is shown in Figure S1. It can be clearly seen that the distribution of  $\theta_{Au-Au}$  is significantly biased toward 90 degrees when compared to an isotropic distribution, indicating that the sample shows a strong preferred orientation. We modeled the preferred orientation using following equation:  $N(\theta_{Au-Au}) = \alpha \cdot \sin^\beta(\theta_{Au-Au})$  where  $N(\theta_{Au-Au})$  is the number of pairs corresponding to  $\theta_{Au-Au}$  and  $\alpha$  and  $\beta$  are constants. We found that the model gives a satisfactory fit to the experimental distribution with the following constants:  $\alpha = 29.4$  and  $\beta = 16.45$  (shown as a red solid line in Figure S1). Accordingly, we used the best-fit model equation to describe the preferred orientation of the sample.

**Supporting Method S2. Preparation of protein sample.** To prepare a synthetic G1C-G45C mutant of equine heart cytochrome c (cyt-c), its corresponding gene was subcloned in the pJRhrs2 vector using EZ-Cloning kit (purchased from Enzynomics™). Initially, the gene was amplified by using PCR. Sequences of primers used for amplification containing KpnI and BamHI restriction enzyme sites are as follows;

5'-GGGGTACCATGTGCGATGTGGAAAAAGGC-3'

5'-CGGGATCCTTATTCGTTGGTCGCTTT -3'

The amplified PCR product treated with BamHI and SalI restriction enzyme was inserted into the pJRhrs2 vector treated with BamHI and SalI restriction enzyme. The G1C-G23C and G1C-E104C mutations were introduced into cyt-c gene using the EZchange® mutagenesis kit (Enzynomics™). The sequence of the oligonucleotides used for the construction of cyt-c mutants

is listed in Supplementary Table 1. The double mutants of cyt-c, namely G1C-G45C, G1C-G23C and G1C-E104C, in the pJRhrs2 vector were over-expressed in *Escherichia coli* BL21(DE3) by IPTG (isopropyl-1-thio- $\beta$ -D-galactopyranoside) induction and purified according to the established procedure<sup>2</sup>.

The cell cultures were grown at 37 °C with ampicillin together with vigorous shaking until a mid-exponential phase (OD 0.6 at 600 nm) was achieved. Then, the expression was induced with 0.2 mM IPTG. After that, cultures were grown for additional 18 hours and were harvested by centrifugation at 6000 g for 10 minutes. The resulting pellet was sonicated in lysis buffer (50 mM of NaPO<sub>4</sub> at pH 7.4 with 0.1 M NaCl) containing protease inhibitors and PMSF (phenylmethylsulfonyl fluoride). The lysate was purified by ammonium sulfate precipitation (55% saturation; 351 g/L). Then, nitrogen gas was purged over the solution for 1 hour followed by 4 hours of dialysis against 10 mM sodium phosphate buffer at pH 7.5 with 1 mM DTT. The protein was applied to a CM Sepharose fast flow ion exchange column (GE Healthcare) equilibrated with 10 mM sodium phosphate at pH 7.5. During the elution procedure, the concentration gradient of NaCl in a range of 40–250 mM was generated in the same buffer. After the purification, the eluted fractions with  $Abs_{410\text{ nm}} / Abs_{280\text{ nm}} > 4$  were pooled. Finally, the selected fraction of cyt-c was diluted, desalted through gel permeation chromatography (GPC) column, and buffer-exchanged to final buffer (20 mM sodium phosphate at pH 6.5 with 150mM NaCl). The purified cyt-c mutants were characterized by UV/VIS absorption, circular dichroism spectroscopy, and mass spectrometry.

**Supporting Method S3. Nanogold labeling of cyt-c.** A 20-fold molar excess of tris(2-carboxyethyl)phosphine hydrochloride (TCEP, purchased from Sigma-Aldrich, USA) was added

to 100  $\mu$ l of 30  $\mu$ M cyt-c mutant solution to reduce disulfide bond. After 15-minute incubation at room temperature, it was mixed with a monomaleimido nanogold (Nanoprobes, USA) solution which was prepared by adding 200  $\mu$ l of distilled water to 6 nmol of nanogold particles lyophilized from 20 mM sodium phosphate at pH 6.5 with 150 mM sodium chloride. The container of the mixed solution was tightly sealed and the gas filled inside the container was removed by using a vacuum pump. Then, the mixture was incubated at room temperature for 5 hours.

To separate the target product from the mixture, size-exclusion chromatography was performed. 100  $\mu$ l of the mixture was loaded onto Enrich<sup>TM</sup> SEC 70 column (Bio-Rad, USA) and eluted through the column with the same buffer. The flow rate of 0.8 ml/min was used, and the eluent was collected for each minute so that the fractional volume of each collected sample was 0.8 ml. The elution profile was monitored with absorbance at four different wavelengths (215 nm, 280 nm, 413 nm and 550 nm). An eluent fraction that corresponds to the first peak of the elution profile was selected for use in the subsequent cryo-EM experiment. A representative elution profile is shown in Figure S2. For all three cyt-c mutants, the same procedures as described above were applied. For the unfolded state sample, 4 M of GdnHCl (Sigma-Aldrich, USA) was adjusted to the selected fraction and used for cryo-EM analysis.

**Supporting Method S4. Acquisition of cryo-EM images.** Cryo-EM samples were prepared using the selected eluent solution. For the folded objects, the original fraction immediately after separation was used without any further dilution or concentration. The solution was deposited on a glow discharged 200 mesh Quantifoil R 2/1 holey-carbon-coated copper grid (Electron Microscopy Science, USA) and blotted with a filter paper inside the vitrification device (FEI Vitrobot<sup>TM</sup> Mark IV). The following blotting conditions were used for folded objects; the loaded

sample volume of 3  $\mu\text{l}$ , the relative humidity of 100% and temperature of 4  $^{\circ}\text{C}$  inside the chamber, the wait time before blotting of 5 s, the blot time of 5 s, the blot force of 7, the wait time of 10 s, and the total number of blotting of 1.

For the unfolded objects, we observed the generation of bubbles when the sample solution containing guanidine hydrochloride (GdnHCl) was loaded on the TEM grid. To remove these bubbles as much as possible, we waited longer than the case without GdnHCl (five seconds for the solution without GdnHCl and ten seconds for the solution with GdnHCl) after loading the sample on the grid and before blotting the grid, and also increased the number of blotting (one for the solution without GdnHCl and three for the solution with GdnHCl). These bubbles are observed only in a small portion of the entire cropped images (37 cropped images out of 162). We confirmed that the bubbles have little effect on the measured projection distances between the particles, from the fact that there was no significant difference between the distribution of the projected distances of 162 particle pairs from all cropped images even containing some bubbles and that of 125 particle pairs only from the cropped images with no bubbles (Figure S7). The blotting conditions for unfolded objects were as follows; the loaded sample volume of 3  $\mu\text{l}$ , the relative humidity of 100% and temperature of 4  $^{\circ}\text{C}$  inside the chamber, the wait time before blotting of 10 s, the blot time of 10 s, the blot force of 12, the wait time of 10 s and the total number of blotting of 3.

Blotted grids were then immediately immersed into liquid ethane and sequentially moved into a grid storage box located in a liquid nitrogen container. The grids were stored inside a liquid nitrogen environment until their usage for electron microscopy. Prepared samples were transferred to cryo-holder (Gatan Multi-Specimen Single Tilt Cryo Transfer Holder, Model 910), and the holder was inserted into a transmission electron microscope (FEI Tecnai G2 Spirit). Data were collected at -174  $^{\circ}\text{C}$  with an acceleration voltage of 120 kV, using a low dose mode ( $< 20 \text{ e}/\text{A}^2\text{s}$ ),

a magnification of 110,000x and the defocus ranges of 0.5–1.5  $\mu\text{m}$ . The exposure time of 1 s was used for the acquisition of all images.

**Supporting Method S5. Processing of cryo-EM images.** To process acquired images automatically, a MATLAB code was written based on the following strategy. First, all particles were selected from background based on the intensity values of the pixels. The pixels located in the particle regions were set as black, and background pixels were set as white. Secondly, only pairs of particles were selected, excluding solitary particles and aggregates composed of three or more particles. In addition to the targeted doubly-labeled cyt-c, the obtained images contained appearances of single particles or aggregated particles because the purification using size exclusion chromatography was not perfect. On average, ~60% of gold particles in a typical image were from isolated pairs of two particles whereas the remaining ~40% were from single particles or aggregated particles (Figure S3). The pairs of particles were selected for each image with the following two criteria: (1) a single particle was excluded if there were no particles other than itself within a radius of 20 nm; (2) a group of three or more particles assembled within a radius of 30 nm was also excluded. All the remaining cases then consisted of two particles, which we regarded as properly labeled cyt-c. The second criterion is a rather severe one and could remove even well-defined pairs of particles, thereby reducing the actually used particles for pairs down to ~30% of all particles observed in the cryo-EM images. Still, we used this criterion to be free from counting potential aggregates at the expense of rejecting good pairs. After filtering out abnormal particle pairs using the criteria, only the groupings composed of two particles remained and these parts of the images were cropped for further automatic process of extracting the projected interparticle distances, which was the last step of the image processing scheme. For each image of a particle

pair, the centroid coordinates for the two particles were determined, and their distance was calculated. The reason for measuring the distance between the centroid of the particles, not the edge-to-edge distance of the particles, is that while the particle edge in an image depends on the edge detection process, the centroid position is rather insensitive to the definition of the process.

**Supporting Method S6. Construction of the projected extra distance distribution.** We performed simulations to estimate the projected extra distance. Briefly, first, the spaces where the nanogold particles can be located were determined considering all spatial constraints imposed by the length of the linker, the size of the particles, and the shape of the protein (Figures S4D and S4F). After that, the positions of the nanogold particles were randomly sampled within the available spaces together with the positions of the labeling sites. The distribution of  $\mathbf{c}_p + \mathbf{d}_p$  was estimated from the sampled positions of the nanogold particles and the labeling sites. There we also considered that the labeling particles have a strong preferred orientation (Supporting Method S1 and Figure S1). The detailed procedure of the simulation is as follows. Four residues (Gly1, Gly23, Gly45 and Glu104) that were selected as the labeling sites were substituted with cysteine residues using PyMOL software and the sphere with a radius of 2 nm (the sum of the linker length, 1.3 nm, and the radius of nanogold particle, 0.7 nm) around the sulfur atoms of cysteine residues were constructed to represent the possible positions of the centroid of the particles. Considering the flexibility of the linker connecting the protein and particle, the position of the centroid of the nanogold particle was limited within 2.0 nm<sup>3</sup> distance from the labeling site. The available positions of the nanogold particles are shown in Figure S4A–C. In Figure S4A–C, cyt-c is represented in black, and the spheres depicting the available positions of the centroid of nanogold particle were colored in red (around Cys1), blue (around Cys23), green (around Cys45), and yellow

(around Cys104). Within the spherical area with 2.0 nm radius, some portion is excluded from the available positions to avoid collision or overlap of the particle with protein (Figure S4D and S4F, marked in gray). First, the portion of the sphere which directly overlaps with the protein was excluded because the particle would not penetrate the protein molecule. In addition, considering the radius of a nanogold particle (0.7 nm), the portion of the sphere within 0.7 nm from the surface of the protein was also excluded because the center of nanogold particles cannot be located in this region. Comparison of each of the remaining, available area for the two labeling particles shows that the overlap between the available area is negligible when the particles are labeled at Cys1 and Cys23 and Cys1 and Cys 45 (Figure S4A and S4B), allowing us to infer that the two particles will only have little interaction with each other. Accordingly, when simulating the projected extra distance for the particles labeled at Cys1 and Cys23 and Cys1 and Cys 45, the position of the particles was determined within the region marked in blue in Figure S4D, considering no interaction between the two particles. By contrast, in the case of labeling at Cys1 and Cys104 sites (Figure S4C), there is a significant overlap between the available area for the two labeling particles due to a short distance (~2.0 nm in the crystal structure) between the two labeling sites. In this case, if we simulate the projected extra distance considering no interaction between the particles (as shown in Figure S4D), the reconstruction of the actual intersite distance yields a significant deviation from the distance theoretically estimated using MD simulations (Figure S4E). The experimental actual intersite distance (blue filled bars, Figure S4E) is reconstructed to be much longer than the distance from the MD simulations (empty bars, Figure S4E), indicating that the projected extra distance is estimated to be shorter than its real value. We hypothesized that this wrong estimation of the projected extra distance is due to an ignorance of the interaction between particles. Here, the distance between two labeling sites is close enough, allowing the particles to

physically touch each other. In reality, the physical collision of two particles is likely to depopulate such cases. To reflect this, we considered the case where each particle avoids positions that potentially can cause collision to the other particle. In other words, each particle was not allowed to be located within a location within 3.4 nm (2.0 nm + particle diameter, 1.4 nm) from the adjacent labeling site (Figure S4F, excluded area considering interparticle interaction is marked in yellow).

Following the scheme of the distribution of the particles around the labeling site described above, we reconstructed the distribution of the projected extra distance as follows. First, with respect to the direction of the incident electron beam, we randomly rotated the cyt-c by multiplying a random three-dimensional rotation matrix to the atomic coordinates of cyt-c. Next, we extracted 10,201 coordinates of uniformly distributed points within a spherical area around the labeling site for each of the two labeling sites of a mutant. After that, among the 10,201 points in each spherical area, the points within the excluded area (Figure S4D and S4F, marked in gray and yellow) were removed. We randomly choose one of the remaining points for each of the two labeling sites of a mutant. The resulting two points were considered as the position of the centroid of the labeling particles.

Next, we sampled the positions of the centroid of the nanoparticles with an additional probabilistic weight depending on  $\theta_{Au-Au}$ . The probabilistic weight was applied to reflect the fact that the nanogold-labeled cyt-c possesses a preferred orientation (which is described in the Supporting Method S1). We adjusted the orientational probability distribution of  $\theta_{Au-Au}$ ,  $P(\theta_{Au-Au})$ , to follow  $\sin^{16.45}(\theta_{Au-Au})$  by rejecting the positions of the centroids unless  $\sin^{15.45}(\theta_{Au-Au})$  (which is an additional weight to  $\sin(\theta_{Au-Au})$  of the isotropic distribution) is larger than a uniform random number generated within the range between 0 and 1. Lastly, if the positions were accepted, the projected extra distance was calculated by using the coordinates of the two nanogold particle

centroids and the two labeling sites.

Repeating this process for 9,999 times yielded 9,999 values of the projected extra distances for each mutant and these values were used to construct the distribution of the projected extra distance as shown in Figure S4G–I. Negative values of the projected extra distance correspond to the case where the projected interparticle distance is smaller than the projected intersite distance.

**Supporting Method S7. Subtraction of the extra distance from the experimental data.** To obtain the distribution of the projected intersite distance,  $\mathbf{b}_p$ , the projected extra distance,  $\mathbf{c}_p + \mathbf{d}_p$ , should be subtracted from the experimentally obtained projected interparticle distance,  $\mathbf{a}_p$ . If there is no correlation at all between  $\mathbf{b}_p$  and  $\mathbf{c}_p + \mathbf{d}_p$ , for each experimental  $\mathbf{b}_p$ , we can randomly choose and use a value from the distributions of the projected extra distance (Figure S4G–I). If there is some correlation, ignoring it will distort the distribution of  $\mathbf{b}_p$  due to a mispairing of  $\mathbf{a}_p$  and  $\mathbf{c}_p + \mathbf{d}_p$ . Therefore, we first checked if there is any correlation by using simulations. For this, (1) a structure pool consisting of 126 virtual structures for a folded protein (G1C-G23C, G1C-G45C, and G1C-E104C in its folded state) or 162 virtual structures for G1C-G45C in its unfolded state were generated. For each virtual structure, the corresponding  $\mathbf{a}_p$  and  $\mathbf{c}_p + \mathbf{d}_p$  were calculated.  $\mathbf{c}_p + \mathbf{d}_p$  were generated in the same manner that was adopted in generating Figure S4G–I. For the case of the unfolded state, the intersite distance of 7 nm was assumed when generating virtual structures. Representative results are shown in Figures S5A–D. Each dot in the plots contains information about the generated  $\mathbf{a}_p$  (position of the dot in  $x$ -axis) and the corresponding  $\mathbf{c}_p + \mathbf{d}_p$  (position of the dot in  $y$ -axis). (2) Then, the structures in the pool were ranked within the pool by two different criteria,  $\mathbf{a}_p$  and  $\mathbf{c}_p + \mathbf{d}_p$ , so that each structure has two different rankings depending on the criteria. To quantify the correlation, the difference between the two ranks (rank difference) was calculated

for each dot. The rank difference indicates that, for example, a small rank difference indicates there is a strong positive correlation between the two projection distances of a structure. (3) To show the trends in the rank difference more clearly, we fractionated the virtual structures in a pool into ten bins in the descending order of  $\mathbf{a}_p$  (namely, the first bin for 0–10%, the second bin for 10–20%, and so on), and took the maximum rank difference in each bin. This largest maximum rank difference value inversely reflects the degree of correlation. The larger the value, the smaller the correlation. The above three steps were repeated 1,000 times to collect 1,000 such plots of the largest maximum rank difference value as a function of the bin. The plots of the largest maximum rank difference values among 1,000 repetitions for each bin obtained for G1C-G23C, G1C-G45C, G1C-E104C, and the unfolded state of G1C-G45C are shown in Figure S5E–H.

These plots of the largest maximum rank difference values provide a safe guideline for pairing  $\mathbf{c}_p + \mathbf{d}_p$  to  $\mathbf{a}_p$ . In other words, the values of the maximum rank difference for each bin can be used to limit the range of  $\mathbf{c}_p + \mathbf{d}_p$  that can be paired to  $\mathbf{a}_p$ . For example, for G1C-G23C, the first bin with 0–10% of  $\mathbf{a}_p$  has the largest maximum rank difference value of 17, indicating that 17 is the largest maximum rank difference possible for the top 10% ranked  $\mathbf{a}_p$ s. Since there are 135 experimental  $\mathbf{a}_p$ s for G1C-G23C, we can rank these 135 values. Also, we randomly select and rank 135  $\mathbf{c}_p + \mathbf{d}_p$  values from Figure S4G. Then, for example, the first 13 (which corresponds to 10% of 135) ranked experimental  $\mathbf{a}_p$ s can be paired with  $\mathbf{c}_p + \mathbf{d}_p$ s considering that the largest possible rank difference should be at most 17. For example, for the experimental  $\mathbf{a}_p$  at the 5th rank, one from the first rank to the 22nd (which is 5 plus 17) rank values from the 135  $\mathbf{c}_p + \mathbf{d}_p$ s is randomly selected toward being subtracted from the 5th rank experimental  $\mathbf{a}_p$  to yield  $\mathbf{b}_p$ . In the same manner, for treating the 6th rank experimental  $\mathbf{a}_p$ , a value from 23 highest-ranking  $\mathbf{c}_p + \mathbf{d}_p$ s is used. This is repeated until processing the 135th rank experimental  $\mathbf{a}_p$ , which belongs to the 10th bin. For the

10th bin, the largest maximum rank difference is 102, and thus a value from 33 (135 - 102) lowest ranking values from the 135  $\mathbf{c}_p + \mathbf{d}_p$ s is used. If a  $\mathbf{b}_p$  value thus obtained by subtracting  $\mathbf{c}_p + \mathbf{d}_p$  from  $\mathbf{a}_p$  was negative, the whole run was rejected and a new one was attempted. The resulting set of 135  $\mathbf{b}_p$ s was converted into a histogram, representing the probability distribution of  $\mathbf{b}_p$ s over 0 to 10 nm with 0.2 nm bin width. To make the data statistically meaningful, we repeated such successful runs 9,000 times and averaged the resulting histograms of the distribution of the population fraction of  $\mathbf{b}_p$ s to obtain the final  $\mathbf{b}_p$  distributions. We chose this number of cycles (9,000) to reach a smooth result shown in Figures 4E–H. The standard deviation of the population fraction of  $\mathbf{b}_p$  for each bin was also calculated from the 9,000 histograms before averaging and is indicated as an error bar in these figures.

**Supporting Method S8. Reconstruction of the actual intersite distance distribution from the projected intersite distance distribution.** To reconstruct the distribution of the actual intersite distance,  $\mathbf{b}$ , from the distribution of the projected intersite distance,  $\mathbf{b}_p$ , the population fraction values at all bins of the distribution of  $\mathbf{b}$  were treated as the independent parameters. Then, the population fraction value at the  $n$ -th bin of the histogram representing the projected intersite distance distribution,  $P_{\text{proj, calc}}(n)$ , can be calculated using the following equation<sup>4,5</sup>:

$$P_{\text{proj, calc}}(n) = \sum_{i=n+1}^m (P(\theta_{\text{res-res}} = \sin^{-1}(d_n/d_i)) \times P_{\text{recon}}(i) \times 1/\sqrt{d_i^2 - d_n^2}) \quad (\text{S3})$$

where  $P_{\text{recon}}(i)$  is a population at the  $(i - 1)$ -th bin of the histogram representing the reconstructed  $\mathbf{a}$ . This is what we aim to determine. Here,  $d_i$  is the center distance of the  $i$ -th bin of the distribution of  $\mathbf{b}$ ,  $d_n$  is the center distance of the  $n$ -th bin of the distribution of  $\mathbf{b}_p$ ,  $P(\theta_{\text{res-res}})$  is the distribution of the angle between the axis connecting the two labeled sites and the direction of the electron beam,  $P(\theta_{\text{res-res}} = \sin^{-1}(d_n/d_i))$  is the population fraction of  $P(\theta_{\text{res-res}})$  at  $\theta_{\text{res-res}} = \sin^{-1}(d_n/d_i)$ , and  $m -$

1 is the number of bins used for analysis. Please note that the equation  $\theta_{res-res} = \sin^{-1}(d_n/d_i)$  expresses the following relationship,  $\sin(\theta_{res-res}) = d_n/d_i$ . In the case of the analyses on folded objects presented in Figure 4E–G, and the analysis on unfolded objects in Figure 4H, there is no significant population with a distance larger than 8 nm, which is the 40th bin of the histogram. Therefore,  $m - 1 = 50$  is large enough for our analyses on both folded and unfolded objects. For  $P(\theta_{res-res})$ , even though the distribution of another projection angle,  $\theta_{Au-Au}$ , was determined by using cryo-electron multitilt analysis (Figure S1), the distribution of  $\theta_{Au-Au}$  would not be identical to the distribution of  $\theta_{res-res}$  because the linker connecting the protein and labeling particles is flexible. Considering this, we modeled the distribution,  $P(\theta_{res-res})$ , to be proportional to  $\sin^x(\theta_{res-res})$ . In this model,  $x$  is a free parameter which denotes the degree of preferred orientation: a high value of  $x$  indicates that there is a strong preferred orientation of  $\theta_{res-res}$ . With this model, the Equation S3 can be expressed as follows:

$$P_{\text{proj, calc}}(n) = \sum_{i=n+1}^m (P_{\text{recon}}(i) \times (d_n/d_i)^x \times 1/\sqrt{d_i^2 - d_n^2}) \quad (\text{S4})$$

where  $(d_n/d_i)^x$  is equal to  $\sin^x(\theta_{res-res})$ .

The population values at all bins of the distribution of  $\mathbf{b}$ , together with the parameter  $x$  which denotes the degree of preferred orientation in the projection angle,  $\theta_{res-res}$ , were iteratively fit by converting it to its corresponding projected intersite distance distribution,  $P_{\text{proj, calc}}(n)$ , and by comparing it against the experimentally derived distribution of  $\mathbf{b}_p$ . More specifically, the parameters were optimized to minimize the chi-square ( $\chi^2$ ) value which is calculated using the following equation:

$$\chi^2 = \sum_{n=1}^{m-1} ((P_{\text{proj, calc}}(n) - P_{\text{proj, exp}}(n))/\sigma(n))^2 \quad (\text{S5})$$

where  $P_{\text{proj, exp}}(n)$  is a population fraction at the  $n$ -th bin of the histogram representing the distribution of experimental  $\mathbf{b}_p$ s, and  $\sigma(n)$  is the standard deviation of the population fraction at the

$n$ -th bin of the histogram representing the distribution of experimental  $\mathbf{b}_p$ s. This has led us to the final reconstructed actual distance distribution. Considering the size of the nanogold particles, we assumed that  $\mathbf{b}$  should always be longer than the diameter of the nanogold particle ( $1.4 \pm 0.14$  nm)<sup>3</sup>. In other words, we applied a constraint that  $P_{\text{recon}}(i) = 0$  for the indices  $i$  with  $d_i < 1.26$  nm ( $1.4 - 0.14$  nm). The error of the reconstructed distribution of  $\mathbf{b}$  and  $x$  were estimated using the MINUIT software package with the MINOS error estimation method<sup>6</sup>.

We also tried to reconstruct the distribution of  $\mathbf{b}$  under a different assumption that the distribution of  $\theta_{\text{res-res}}$ ,  $P(\theta_{\text{res-res}})$ , is identical to that of another projection angle,  $\theta_{\text{Au-Au}}$ . The resulting distribution of  $\mathbf{b}$ , together with its corresponding projected intersite distance distribution, is shown in Figure S6. Here, the procedure of the reconstruction is the same as described above except that the parameter  $x$  is fixed at 16.45 throughout the iterative fitting. A comparison of  $\chi^2$  values for the fixed  $x$  (shown in Figures S6M–S6P) and free  $x$  (shown in Figures S6E–S6H) shows that  $\chi^2$  is much higher with  $x$  fixed at 16.45 than with free  $x$ . The significantly lower  $\chi^2$  with free  $x$  indicates that the pair of labeling nanogold particles and the pair of labeled residues in the protein have different degrees of preferred orientation with respect to the direction of the incident electron beam. The reason for the higher  $\chi^2$  values for the fixed  $x$  is that the calculated projected intersite distance distribution with the high  $x$  value ( $x = 16.45$ ) cannot describe the populations of the experimentally derived projected intersite distance distribution at short  $\mathbf{b}_p$  values (at  $\mathbf{b}_p < 1.0$  nm, see Figures S6M–S6P) unless the populations in the distribution of  $\mathbf{b}$  are allowed even for the short  $\mathbf{b}$  values,  $\mathbf{b} < 1.26$  nm, which are physically unrealistic. In contrast, when  $x$  is small, the experimental distribution of  $\mathbf{b}_p$  at short  $\mathbf{b}_p$  values can be well described without a population at short  $\mathbf{b}$  values (See Figures S6E–S6H). The optimized  $x$  values are close to 1 in all four cases shown in Figures S6A–S6D, implying that the protein has an orientation close to random unlike the labeled

nanoparticle pairs that show a strong preferred orientation.

**Supporting Method S9. Calculation of the intersite distance distribution from the MD simulations.** We adopted MD simulations to sample the structural ensembles of both the folded and the unfolded states of horse heart cyt-c. The simulations were started from the X-ray crystal structure<sup>7</sup> by employing the Gromos96-43a1 force field<sup>8</sup>. The simulations were performed with the GROMACS 4.5.5 program package<sup>9</sup>. For the folded state, the protein was placed in a cubic box with 6 nm side lengths and solvated by TIP3P water<sup>10</sup>. The net charge in the box was neutralized by replacing some water molecules with sodium cations. A cutoff distance of 1.6 nm was used for treating both the electrostatic and dispersion interactions, while the particle mesh Ewald (PME) method<sup>11,12</sup> was utilized toward treating long-range electrostatics. Before the production run, the potential energy of the system was minimized by using the steepest descent algorithm, and the system was then equilibrated for 20 ps at 293 K with the velocity-rescale thermostat<sup>13</sup>. This was followed by the production run over a 1 ns duration. The structures were collected at every 1 ps. The intersite distance distributions obtained from the MD simulations for the folded state are plotted with empty bars in Figure 5A–C.

Sampling the vast conformational space with explicit solvent for the unfolded state, especially at room temperature, is computationally too burdensome. To circumvent this issue, we employed a high-temperature condition<sup>14,15</sup> of 800 K in combination with the GB/SA implicit solvent model<sup>16</sup>. While this temperature was somewhat higher than the one used in other unfolding simulations<sup>14</sup>, we did not find any instabilities during simulations. In this case, the stochastic Langevin dynamics was employed with the solvent collision rate of  $0.5 \text{ ps}^{-1}$ . This unfolding simulation was also started from the crystal structure, with an initial equilibration for 1 ns. The

production run was then performed for 500 ns, and the structures were sampled at every 20 ps. The intersite distance distribution obtained from the simulation for the unfolded state is plotted with empty bars in Figure 5D.

We are aware of the possibility of beam-induced motion, which can be corrected by CMOS-based direct cameras and software such as Motioncor. According to relevant studies<sup>17,18</sup>, individual particles can move 0.1–1 nm in the ice during cryo-EM image acquisitions. We did not take such beam-induced movement into account, but the good agreement between the MD simulation and the experimental distribution indicates that the beam-induced motion did not significantly affect our measurements.

**Supporting Method S10. Estimation of the intersite distance distribution by using random coil polymer models.** For comparison with the distance distribution of the unfolded state, we also estimated the distributions by assuming that the unfolded protein can be regarded as a random coil polymer. Specifically, we considered two models: (i) a random walk model with a completely flexible chain but with an excluded volume for avoiding intra-chain steric clashes<sup>19</sup>, and (ii) a worm-like chain model where the chain is not completely flexible but has a degree of stiffness due to the residue-residue interaction<sup>20</sup>.

The random walk model estimates the distribution of the polymer end-to-end distance  $r$  as<sup>19</sup>:

$$p(r) \propto 4\pi r^2 \times \exp[-((r - b)/\sigma)^2] \quad (\text{S6})$$

with  $b$  and  $\sigma$  defined as:

$$\begin{aligned} b &= 5/3^{3/5} \times (V/(3\pi l))^{1/5} \times L^{3/5} \\ \sigma &= \sqrt{20Ll/27} \end{aligned} \quad (\text{S7})$$

Here,  $L = nl$  represents the length of the polymer with  $n = 44$  denoting the number of chain

segments (i.e., the number of residues between the 1st and the 45th residues of cyt-c) and  $l$  serving as the length of one chain segment (which is  $3.8 \text{ \AA}$  for a protein residue). The excluded volume  $V$  can be used as a free parameter for fitting, and we assumed that this volume can be approximated as a sphere whose radius  $r_e$  does not deviate significantly from the half of the monomer size,  $l/2 = 1.9 \text{ \AA}$ . With this assumption, we inspected whether there are  $r_e$  values that could well explain the distributions obtained from the MD simulations and the NACS experiment. We found that the distribution from the first random walk model (Figure 5D, black dashed line) agreed well with the distribution from MD simulations when  $r_e = 1.3 \text{ \AA}$ , which is close to  $l/2$ . In contrast, to explain the distance distribution obtained from the NACS experiment well,  $r_e$  needed to be over  $10.0 \text{ \AA}$ . Such a large value is unreasonable considering the size of an amino acid residue.

For the worm-like chain model, we used the following end-to-end distance distribution function<sup>20</sup>:

$$p(r) \propto 4\pi r^2 \times (1 - r^2/L^2)^{-9/2} \times \exp[-3/4 \times (3L/2l_p) \times 1/(1 - r^2/L^2)] \quad (\text{S8})$$

Here,  $l_p$  is the persistence length of the polymer, which was used as an adjustable parameter. A distribution that agrees well with the experimental NACS result was obtained with  $l_p = 22.8 \text{ \AA}$ , which corresponds to an extension over about six residues. Another value of  $l_p$  was also explored to see if the distribution can match the MD simulation result. A good match was found but  $l_p$  was close to  $l$ , the length of one amino acid residue. Not surprisingly, this case corresponds to the situation with no residue-residue interactions, and it basically becomes equivalent to the random walk model.

## Supporting References

- (1) Noble, A. J.; Dandey, V. P.; Wei, H.; Braschi, J.; Chase, J.; Acharya, P.; Tan, Y. Z.; Zhang, Z. N.; Kim, L. Y.; Scapin, G.; Rapp, M.; Eng, E. T.; Rice, W. J.; Cheng, A. C.; Negro, C. J.; Shapiro, L.; Kwong, P. D.; Jeruzalmi, D.; des Georges, A.; Potter, C. S.; Carragher, B. Routine Single Particle CryoEM Sample and Grid Characterization by Tomography. *Elife* **2018**, *7*, e34257.
- (2) Rumbley, J. N.; Hoang, L.; Englander, S. W. Recombinant Equine Cytochrome c in Escherichia coli: High-level Expression, Characterization, and Folding and Assembly Mutants. *Biochemistry* **2002**, *41*, 13894-13901.
- (3) Hainfeld, J. F.; Furuya, F. R. A 1.4-nm Gold Cluster Covalently Attached to Antibodies Improves Immunolabeling. *J. Histochem. Cytochem.* **1992**, *40*, 177-184.
- (4) Busson, M. P.; Rolly, B.; Stout, B.; Bonod, N.; Larquet, E.; Polman, A.; Bidault, S. Optical and Topological Characterization of Gold Nanoparticle Dimers Linked by a Single DNA Double Strand. *Nano Lett.* **2011**, *11*, 5060-5065.
- (5) Lermusiaux, L.; Sereda, A.; Portier, B.; Larquet, E.; Bidault, S. Reversible Switching of the Interparticle Distance in DNA-Templated Gold Nanoparticle Dimers. *ACS Nano* **2012**, *6*, 10992-10998.
- (6) James, F.; Roos, M. Minuit - a System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.* **1975**, *10*, 343-367.
- (7) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. High-Resolution Three-Dimensional Structure of Horse Heart Cytochrome c. *J. Mol. Biol.* **1990**, *214*, 585-595.
- (8) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of Aliphatic CH<sub>n</sub> United Atoms of GROMOS96 Force Field. *J. Comput. Chem.* **1998**, *19*, 535-547.
- (9) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C.

GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701-1718.

(10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(11) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(12) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

(13) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*.

(14) Toofanny, R. D.; Daggett, V. Understanding Protein Unfolding from Molecular Simulations. *Wires Comput. Mol. Sci.* **2012**, *2*, 405-423.

(15) Wang, T.; Wade, R. C. On the Use of Elevated Temperature in Simulations to Study Protein Unfolding Mechanisms. *J. Chem. Theory Comput.* **2007**, *3*, 1476-1483.

(16) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A* **1997**, *101*, 3005-3014.

(17) Li, X. M.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunschweig, M. B.; Gubbens, S.; Agard, D. A.; Cheng, Y. F. Electron Counting and Beam-Induced Motion Correction Enable Near-Atomic-Resolution Single-Particle cryo-EM. *Nat. Methods* **2013**, *10*, 584-590.

(18) Scheres, S. H. W. Beam-Induced Motion Correction for Sub-Megadalton Cryo-EM Particles. *Elife* **2014**, *3*, e03665.

(19) Edwards, S. F. The Statistical Mechanics of Polymers with Excluded Volume. *Proc. Phys. Soc., London* **1965**, *85*, 613-624.

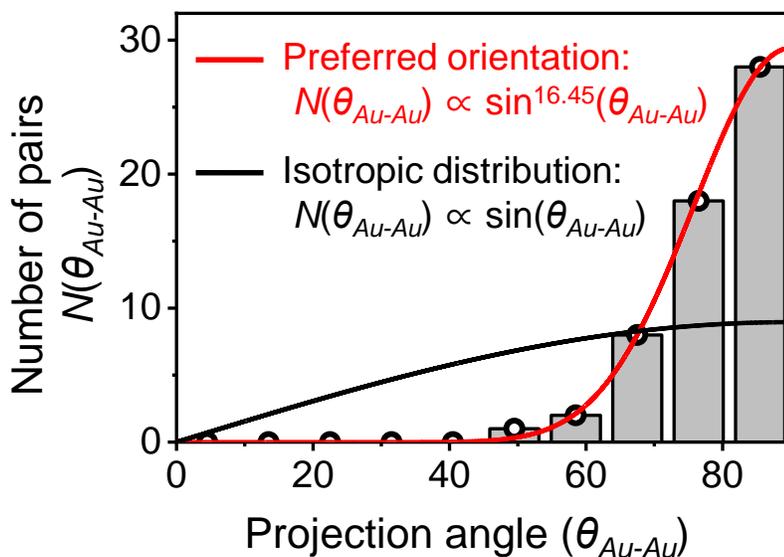
(20) Thirumalai, D.; Ha, B.-Y., Statistical Mechanics of Semiflexible Chains: A Mean-Field Variational Approach. In *Theoretical and Mathematical Models in Polymer Research*; Grosberg, A., Ed; Academic Press: San Diego, 1998; pp 1-35.

**Table S1. The sequence of the oligonucleotides used for construction of cyt-c mutants.**

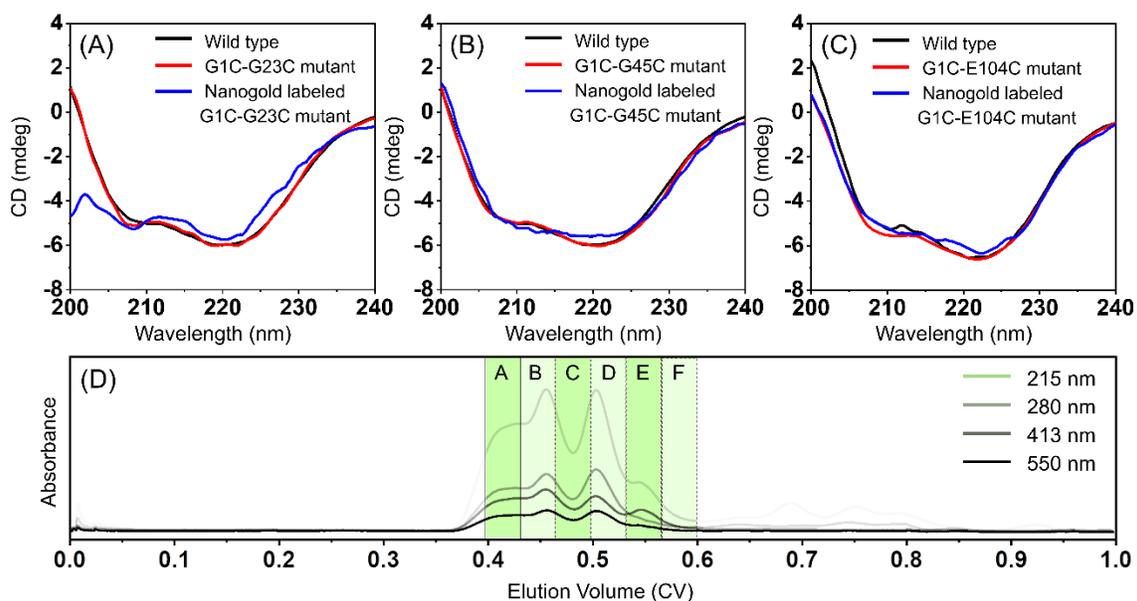
Mutant	Oligonucleotide Sequence (5'-3')*
G23C	Sense, CACTGTCGAAAAAT <u>GTGG</u> TAAGC Anti-sense, GCTTACC <u>CAT</u> TTTTTCGACAGTG
E104C	Sense, CACTAACT <u>GTTA</u> AGGCCTGG Anti-sense, CCAGGCCTTAA <u>CAG</u> TTAGTAG

\*Mutation sites are underlined.

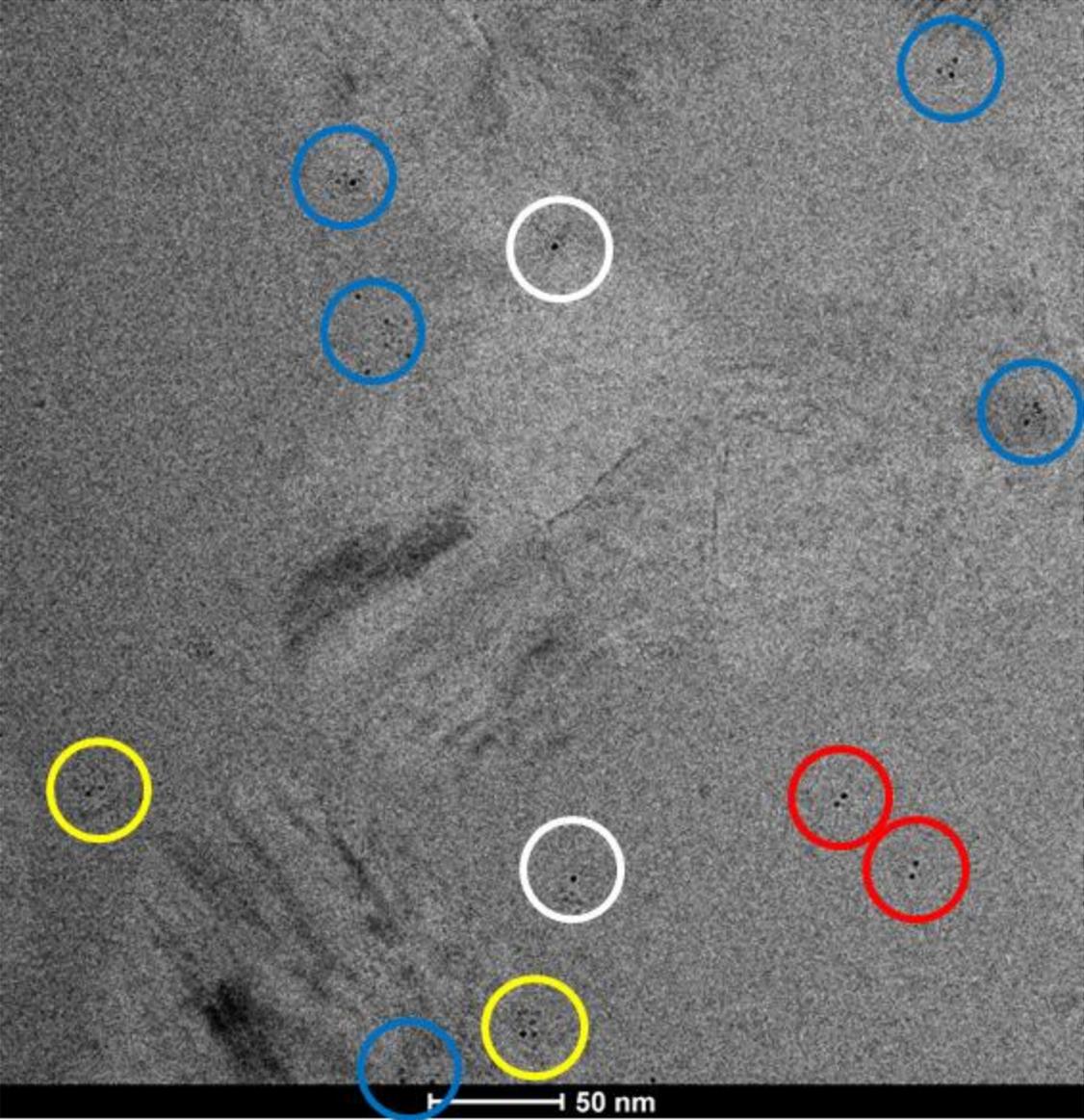
## Supporting Figures



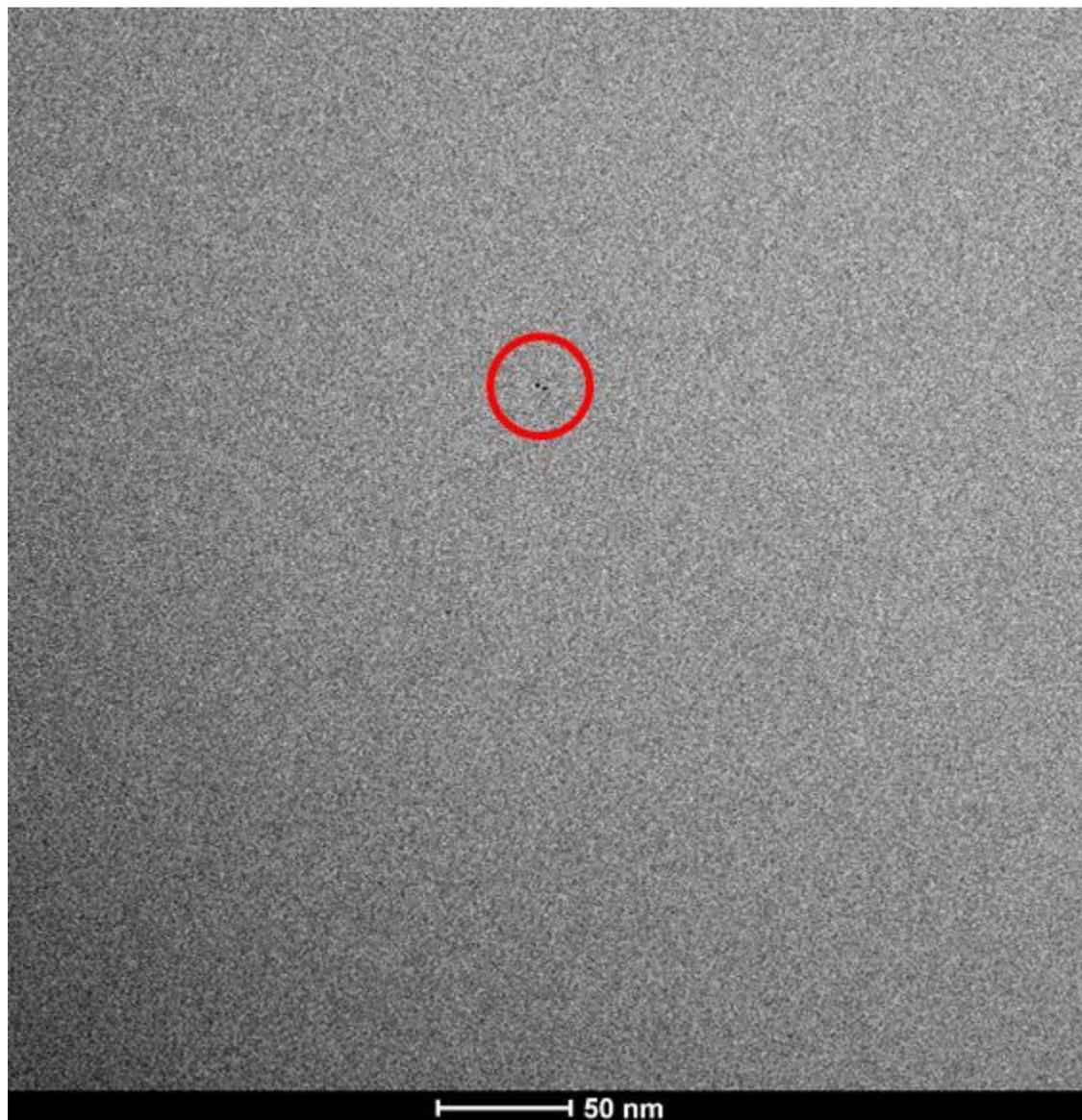
**Figure S1.** The distribution of  $\theta_{Au-Au}$ , the angle between the axis connecting two particles and the direction of the electron beam. We used cryo-electron multitilt analysis to determine  $\theta_{Au-Au}$  for 57 pairs of nanogold particles labeled at G1C-G45C mutant of cyt-c in a folded state. The resulting distribution is shown as a histogram with gray filled bars and empty circles. It is clear that the distribution of  $\theta_{Au-Au}$  is significantly biased toward 90 degrees when compared to an isotropic distribution (black solid line), indicating that the sample shows a strong preferred orientation. We modeled the preferred orientation using following equation:  $N(\theta_{Au-Au}) = \alpha \cdot \sin^\beta(\theta_{Au-Au})$  where  $N(\theta_{Au-Au})$  is the number of pairs corresponding to  $\theta_{Au-Au}$  and  $\alpha$  and  $\beta$  are constants. The best-fit model to the experimental distribution with  $\alpha = 29.4$  and  $\beta = 16.45$  is shown as a red solid curve.



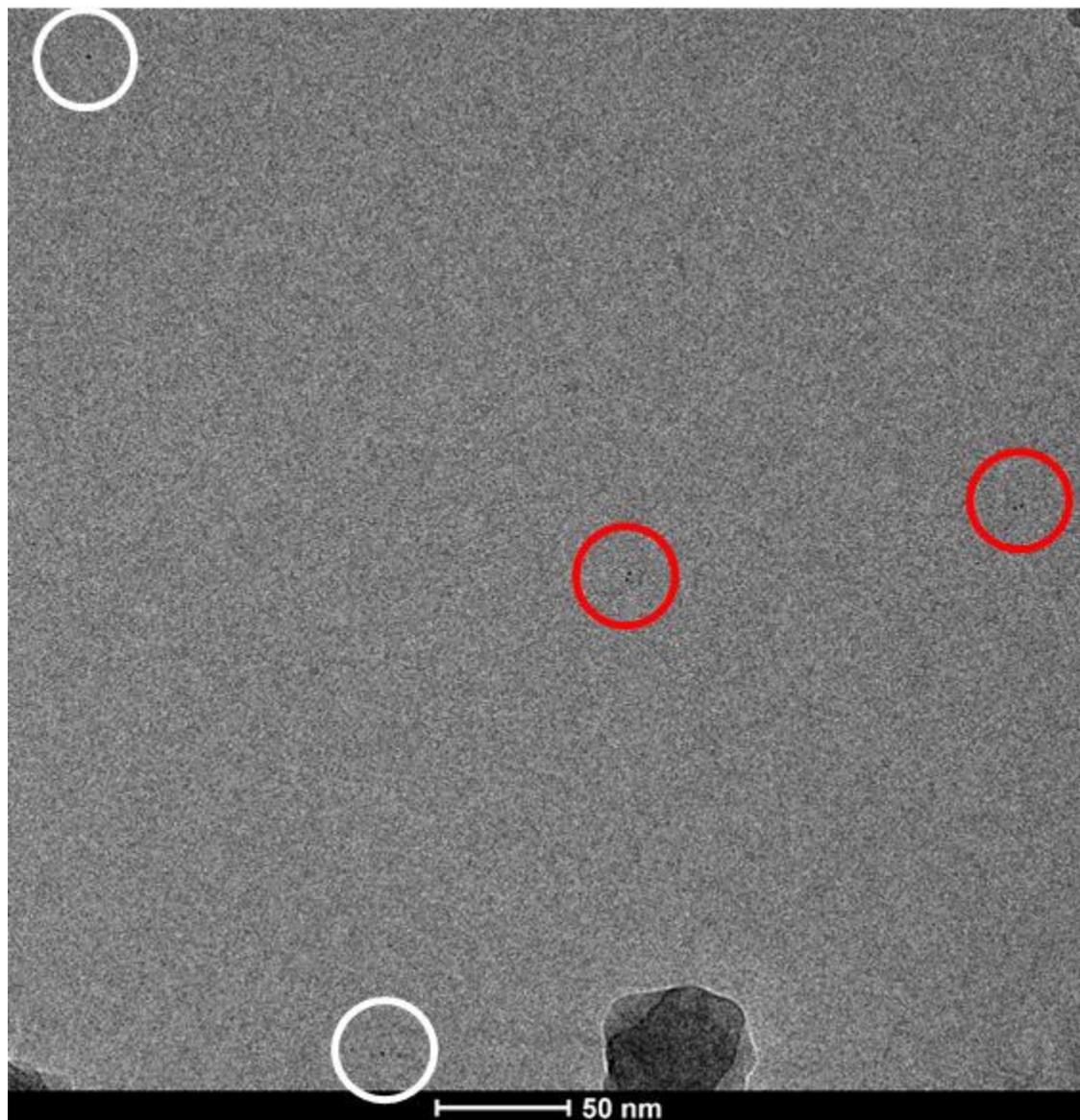
**Figure S2.** Results of CD measurement and size-exclusion chromatography. (A–C) CD spectra of wild type cyt-c, three mutants, and three mutants labeled with nanogold particles. CD spectra of nanogold-labeled mutants were measured from a mixture of nanogold particles and cyt-c before the size-exclusion chromatography. (A) CD spectra from wild type (black), G1C-G23C mutant (red), and G1C-G23C mutant labeled with nanogold (blue). (B) CD spectra from wild type (black), G1C-G45C mutant (red), and G1C-G45C mutant labeled with nanogold (blue). (C) CD spectra from wild type (black), G1C-E104C mutant (red), and G1C-E104C mutant labeled with nanogold (blue). (D) Representative elution profile of size-exclusion chromatography measured from the nanogold-labeled G1C-G45C mutant of cyt-c. An eluent fraction of A was taken as the target product to measure cryo-EM images.



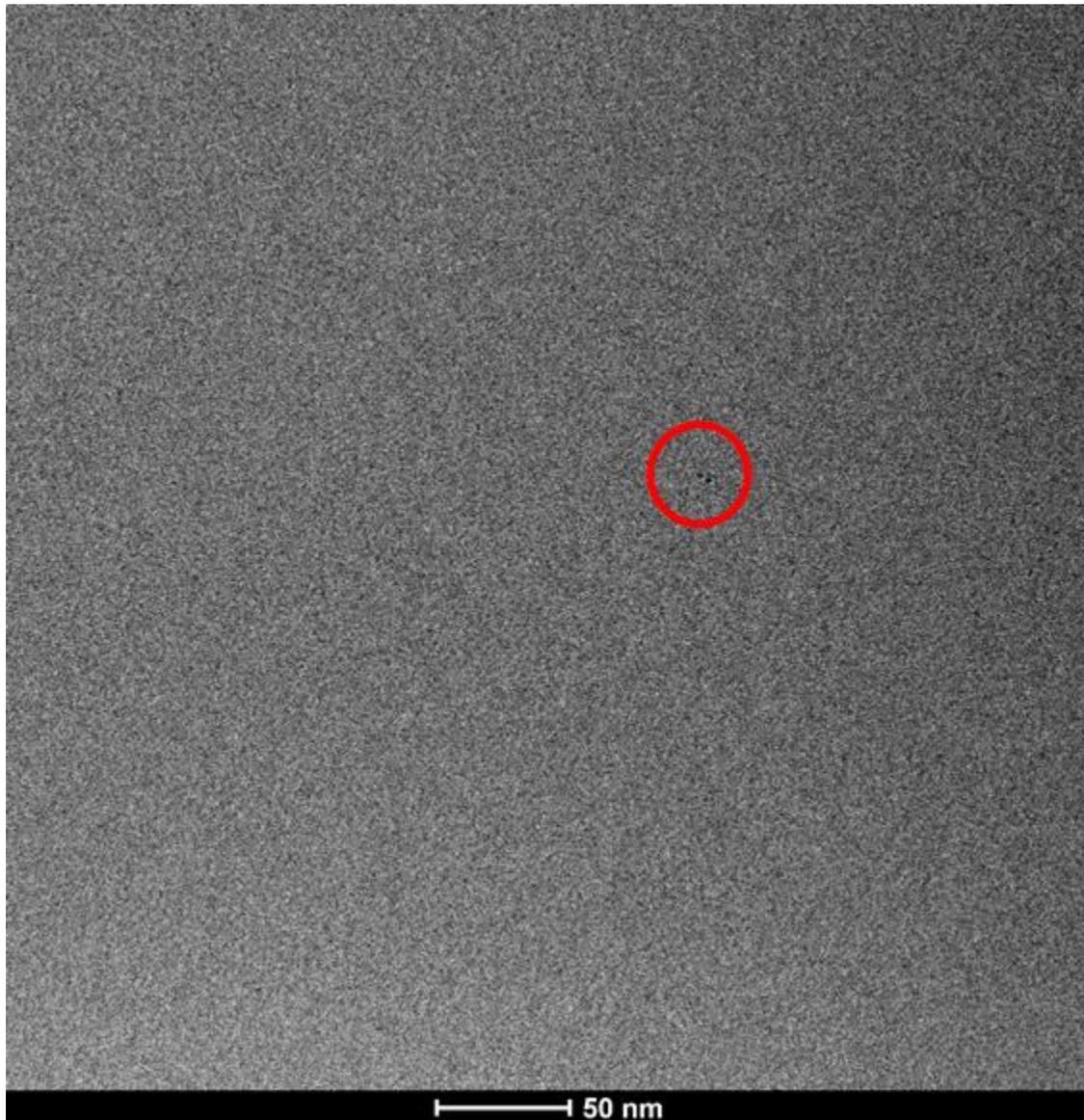
(Continued from the previous page, cryo-EM images for nanogold labeled G1C-G45C cyt-c)



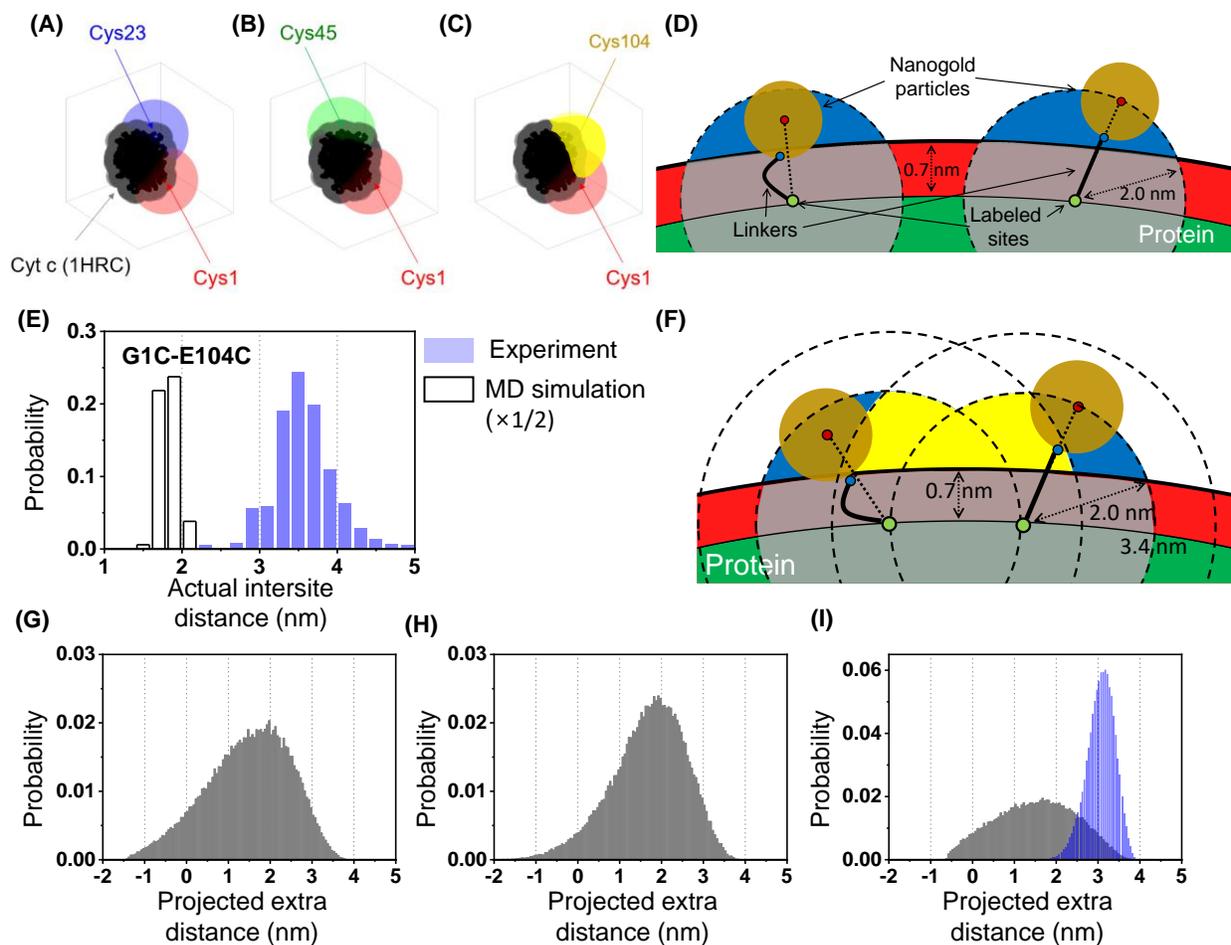
(Continued from the previous page, cryo-EM images for nanogold labeled G1C-G45C cyt-c)



(Continued from the previous page, cryo-EM images for nanogold labeled G1C-G45C cyt-c)

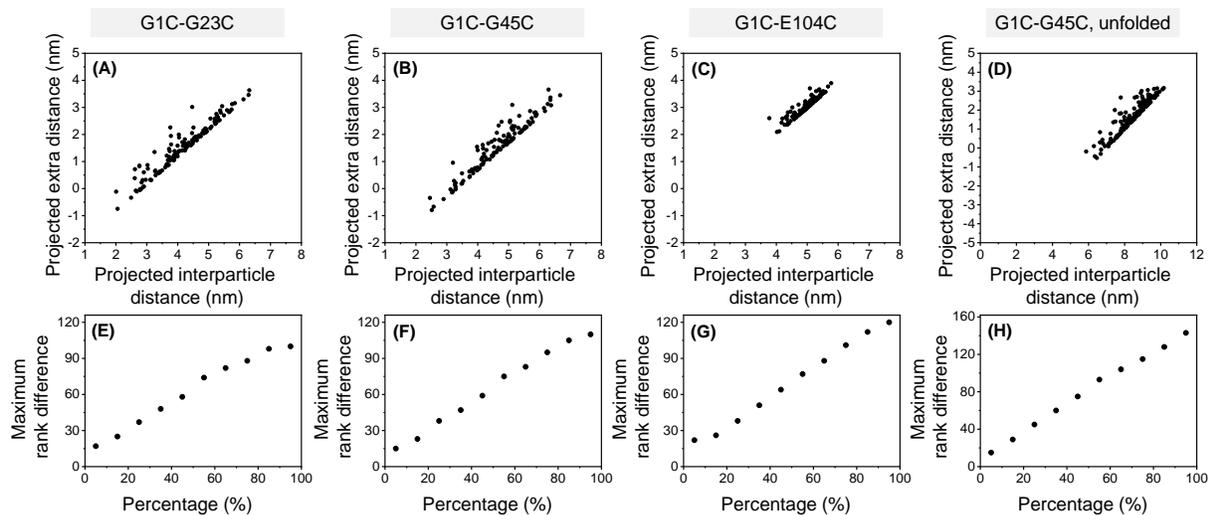


**Figure S3.** Representative raw cryo-EM images for nanogold labeled G1C-G45C cyt-c. The well-isolated pairs of nanoparticles which eventually passed the criteria for selecting good pairs are indicated with red circles. The pairs of nanoparticles which failed to pass the criteria are indicated with yellow circles. The solitary particles and aggregates are indicated with white and blue circles, respectively.



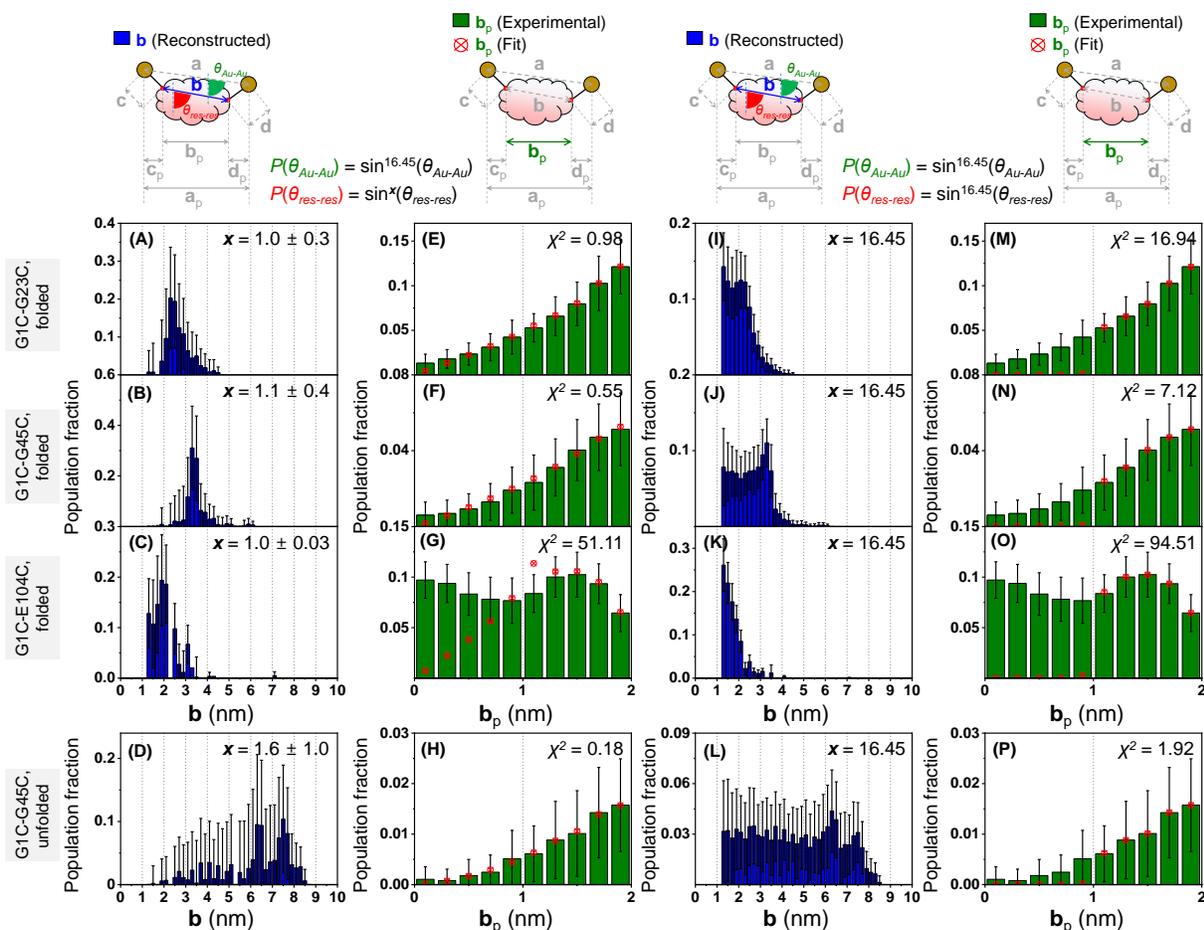
**Figure S4.** Construction of the projected extra distance distribution. (A–C) The position of the labeled residues and possible positions of labeling particles. The red, blue, green, and yellow spheres depict the possible position of the centroid of the labeling nanogold particle around the labeling sites, Cys1, Cys23, Cys45, and Cys104, respectively. The labeled protein, cyt-c, are colored with black. Due to a collision with the protein, the centroid of a nanogold particle is inaccessible to the gray-colored region near the surface of the protein. (D) Pictorial representation of available positions of particles around the labeling sites in the case when there is enough distance between two labeling sites. A cross-sectional view around the labeling sites is shown. The protein is colored green, and the positions where the centroid of the nanogold particle can and cannot be located are marked in blue and gray, respectively. The linker on the left is shown as a

black curve to indicate that a linker can be flexible, and the linker on the right is shown as a straight line to show the case of the maximum linker distance. (E) In the case shown in panel C (particle labeling at Cys1 and Cys104), the experimental actual intersite distance (blue filled bars) is reconstructed to be much longer than the distance from the MD simulations (empty bars) if we estimate the projected extra distances by ignoring the interaction between the two particles (i.e., reconstructed using the projected extra distance distributions shown in panel I as gray bars). (F) Pictorial representation of the distribution of potential locations of particles around the labeling sites in the case when the distance between two labeling sites is close enough, allowing the particles to physically touch each other. The physical collision of two particles is likely to depopulate such cases. To reflect this, it was assumed that each particle avoids positions that potentially can cause collision to the other particle. In other words, each particle was not allowed to be located within a location within 3.4 nm (2.0 nm + particle diameter, 1.4 nm) from the adjacent labeling site. The color scheme is identical to that of panel D except for the additional region, where the particles can cause collision to the other particle, colored in yellow. A cross-sectional view around the labeling sites is shown. (G–I) Distributions of the projected extra distances calculated for (G) G1C-G23C, (H) G1C-G45C, and (I) G1C-E104C mutants of cyt-c. The bin size of 0.05 nm was used for the plot. The distributions were estimated under the assumption of the preferred orientation ( $\sin^{16.45}(\theta_{Au-Au})$ , see Figure S1.) of particles relative to the electron beam propagation. In panel I, the two distributions estimated by ignoring the interaction between the two particles (gray bars) and avoiding collisions (blue bars) are shown. The distribution of the projected intersite distance shown in Figure 4G, and the distributions of the actual intersite distance shown in Figures 4K, 5C, S6C, S6G are calculated and reconstructed by using the distribution shown in blue bars.



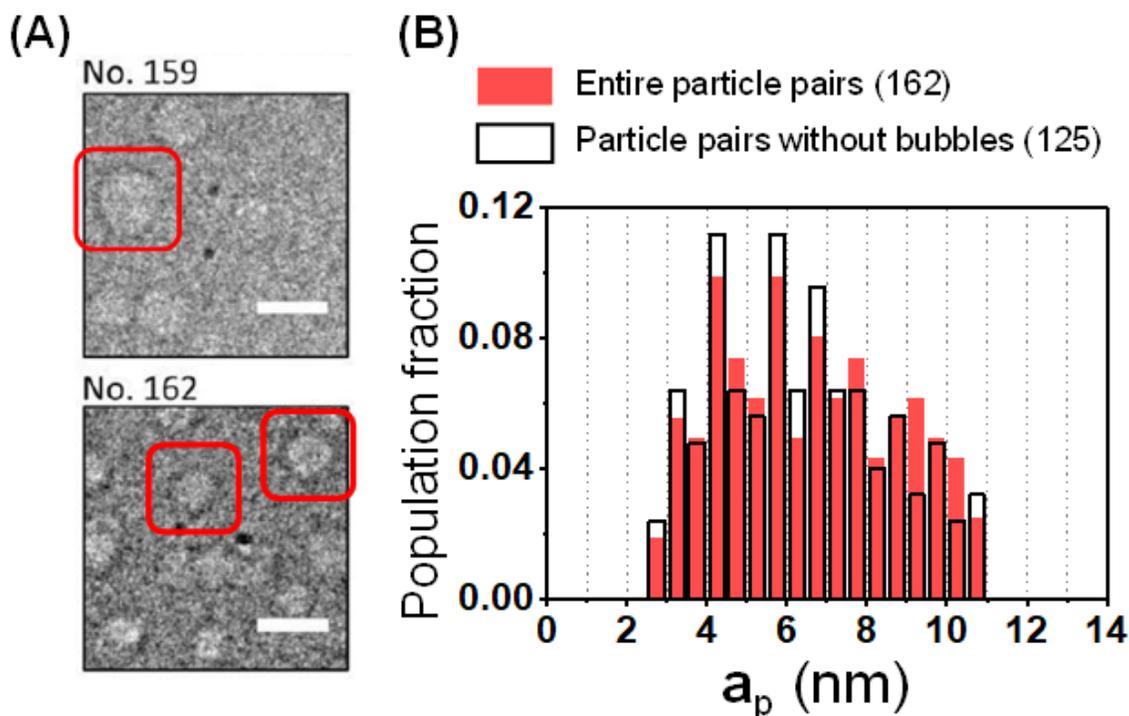
**Figure S5.** Estimation of the range of the projected extra distance values and their correlation with the projected interparticle distance values. (A–D) Representative plots of the projected interparticle distances and the projected extra distances measured from randomly generated structures of nanogold labeled G1C-G23C, G1C-G45C, G1C-E104C, and unfolded G1C-G45C mutant of cyt-c, respectively. The position of a dot in  $x$  and  $y$  axes represents the projected interparticle distance and the projected extra distance, respectively, measured from one of randomly generated structures. The projected distances measured from 126 structures for each of the folded protein (A–C) and from 162 structures for the unfolded protein (D) are shown as the representatives of the distances measured for 10,000 structures generated for each protein. (E–H) The value of the largest maximum rank difference as a function of the ranked projected interparticle distance represented in terms of percentage. To generate these plots, we generated 1,000 structure pools, each consisting of 126 structures for a folded protein or 162 structures for the unfolded protein. For each pool, the structures in the pool were ranked within the pool by two different criteria, the projected interparticle distance and the projected extra distance, so that each

structure has two different rankings depending on the criteria. The difference between the two rankings (rank differences) was calculated for each structure to figure out the correlation between the projected interparticle distances and the projected extra distances. For example, a small rank difference indicates that there is a strong positive correlation between the two projection distances of a structure. To show the trends in the rank difference more clearly, we fractionated the structures in a pool into ten subgroups in the projected interparticle distance order and took the maximum rank difference in each subgroup. The maximum rank differences were compared among 1,000 structure pools, and their maxima are shown in this plot.



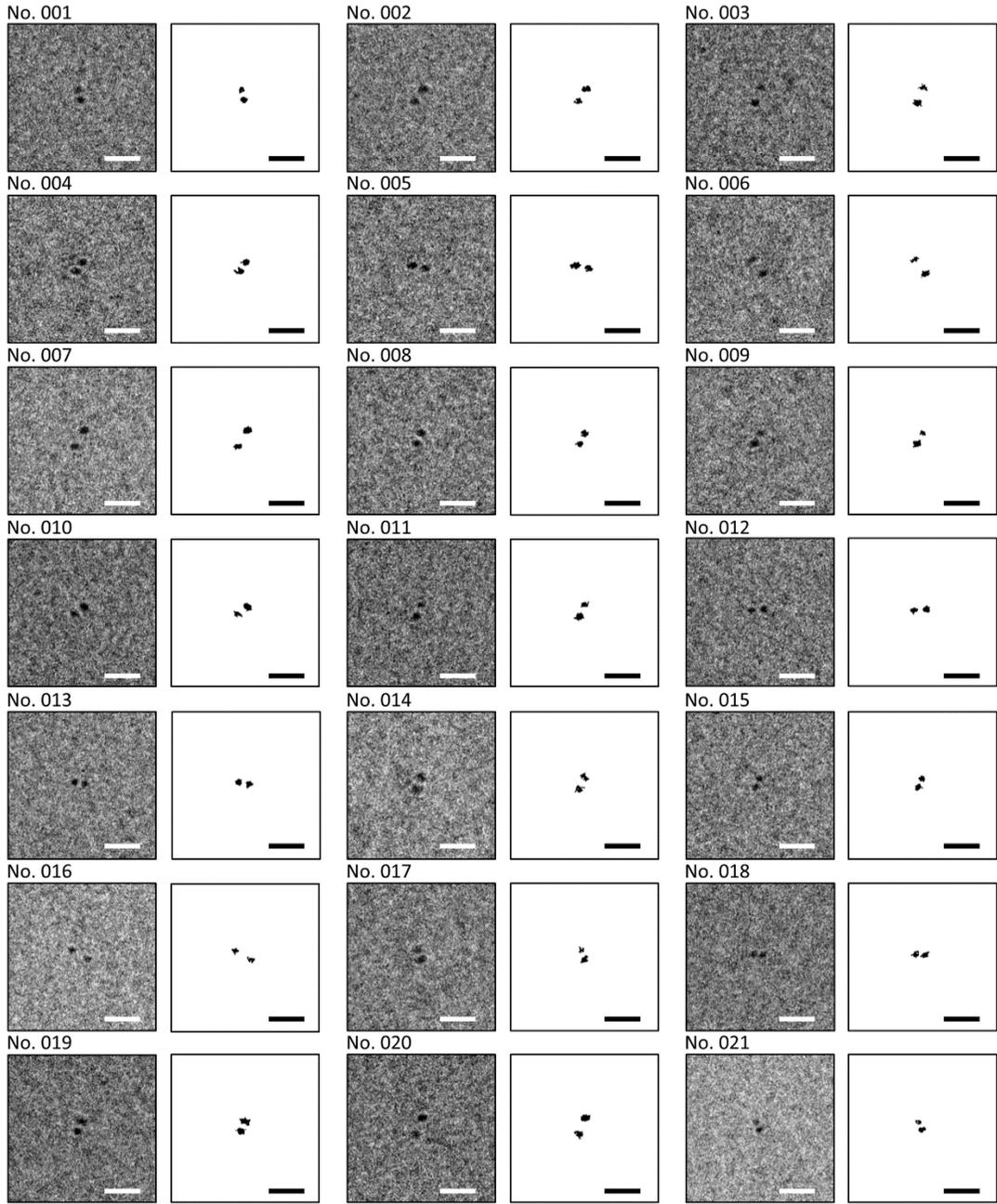
**Figure S6.** Comparison of the distribution of the actual intersite distance,  $\mathbf{b}$ , reconstructed under two different assumptions of the distribution of projection angle,  $P(\theta_{res-res})$ , and the corresponding best-fit distribution of the projected intersite distance,  $\mathbf{b}_p$ , to the experimentally derived distribution of  $\mathbf{b}_p$ . (A–H) Reconstructed distribution of  $\mathbf{b}$  (A–D) and best-fit distributions of  $\mathbf{b}_p$  (E–H, red circles with crosses) under the assumption that the pair of labeling nanogold particles and the pair of labeled residues in the protein have different degrees of preferred orientation with respect to the direction of the incident electron beam. In other words, we assumed that the distribution of  $\theta_{res-res}$ ,  $P(\theta_{res-res})$  is different from the distribution of another projection angle,  $\theta_{Au-Au}$ .  $P(\theta_{res-res})$  is modeled

to be proportional to  $\sin^x(\theta_{res-res})$ . In this model,  $x$  denotes the degree of preferred orientation: a high value of  $x$  indicates that there is a strong preferred orientation of  $\theta_{res-res}$ . The best-fit values of  $x$  to the experimental data are shown along with their errors in the panels A–D. The experimental distribution of  $\mathbf{b}_p$  (E–H, green bars) is shown in the panels E–H together with the best-fit distribution of  $\mathbf{b}_p$  for a comparison. The figures in the panels E–H are identical to the Figures 4E–4H in the main text except that the range of the  $x$ -axis,  $\mathbf{b}_p$ , is narrower in this figure to show more emphasis on the distributions of  $\mathbf{b}_p$  at low  $\mathbf{b}_p$  values. (I–P) Reconstructed distribution of  $\mathbf{b}$  (I–L) and corresponding best-fit distribution of  $\mathbf{b}_p$  (M–P, red circles with crosses) to the experimental distribution of  $\mathbf{b}_p$  under the assumption that the pair of labeling nanogold particles and the pair of labeled residues in the protein have the same degree of preferred orientation. In this case,  $P(\theta_{res-res})$  is proportional to  $\sin^{16.45}(\theta_{res-res})$ . A comparison of the  $\chi^2$  for the panels E–H and I–L shows that  $\chi^2$  is much higher with  $x$  fixed at 16.45 than with free  $x$ . The significantly lower  $\chi^2$  with free  $x$  indicates that the pair of labeling nanogold particles and the pair of labeled residues in the protein have different degrees of preferred orientation with respect to the direction of the incident electron beam. The optimized  $x$  values are close to 1 in all four cases shown in panels A–D, which further implies that the protein has an orientation close to random unlike the labeled nanoparticle pairs that show a strongly preferred orientation.

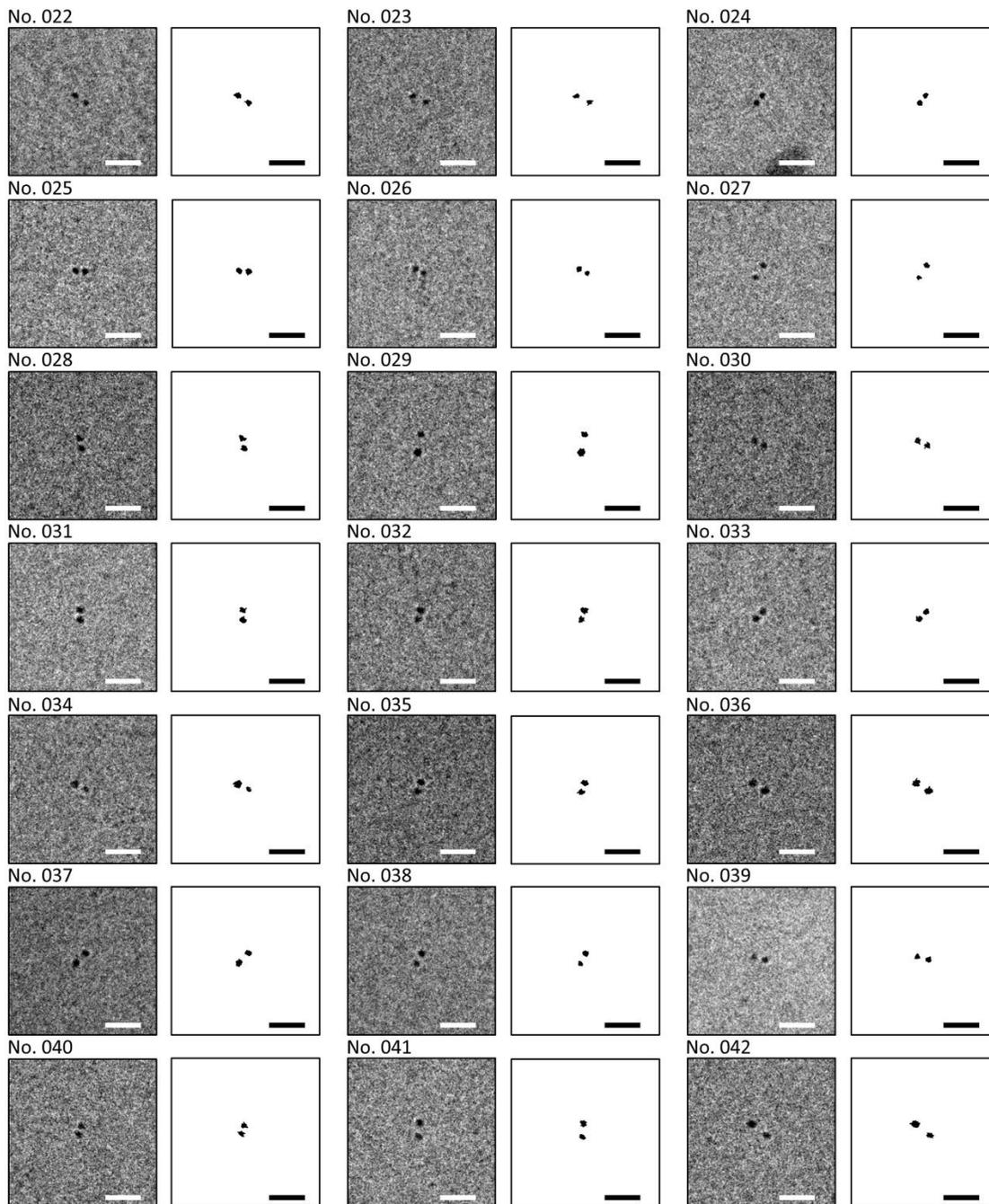


**Figure S7.** Effect of bubbles in the sample solution containing guanidine hydrochloride (GdnHCl).

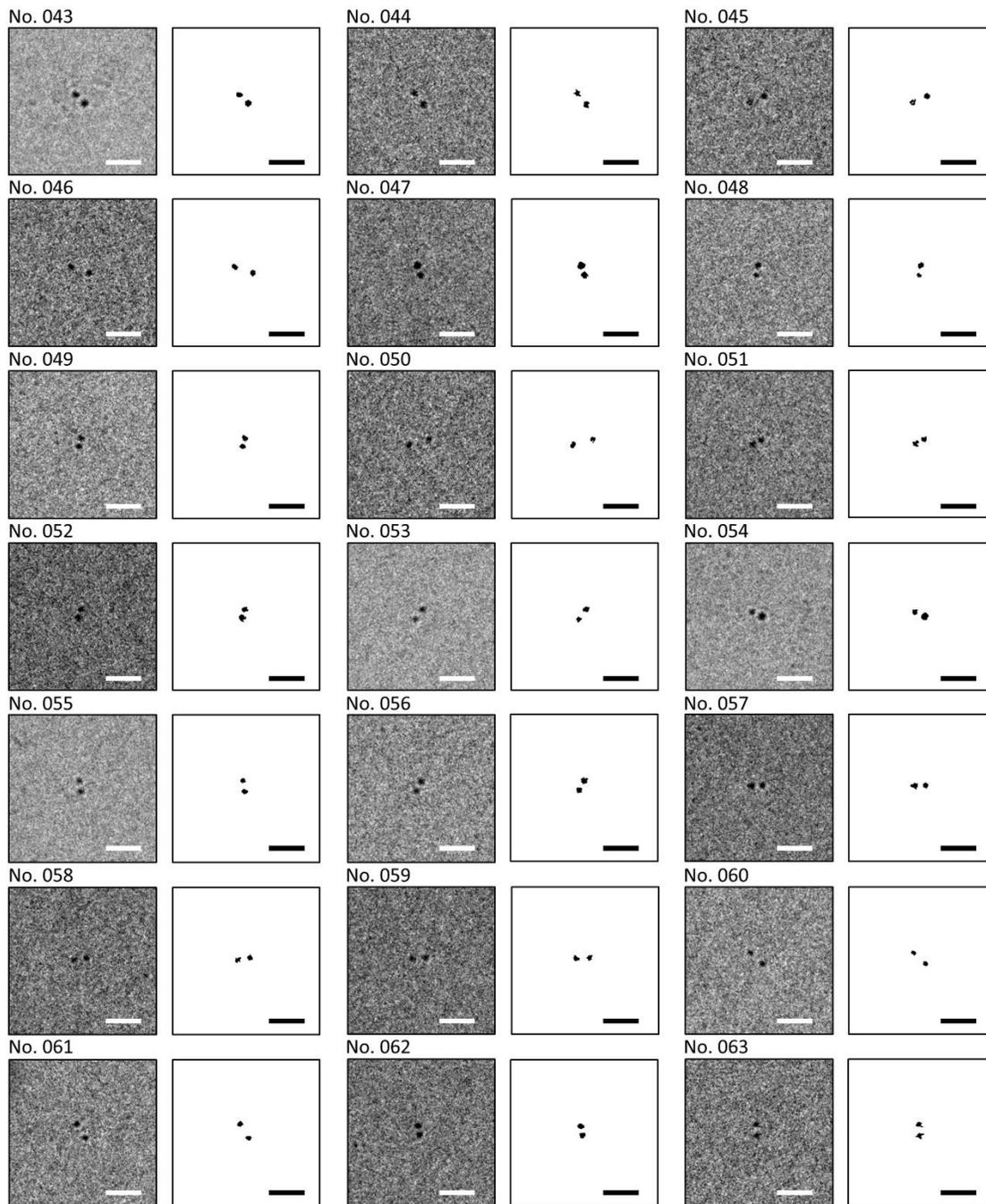
(A) Two representative cropped cryo-EM images showing bubbles in the cryo-EM samples. We observed the generation of bubbles when the sample solution containing GdnHCl was loaded on the TEM grid. These bubbles are observed only in a small portion of the entire cropped images (37 images out of 162, Nos. 076–078, 091–092, 131–162 in Figure S11). (B) The distribution of the projection distances of the 125 particle pairs (empty bars) from the cropped images with no bubbles is compared with those of the 162 particle pairs (red filled bars) from all cropped images shown in Figure S11. There is no significant difference, indicating that these bubbles have little effect on the distribution of the distances between the particles.



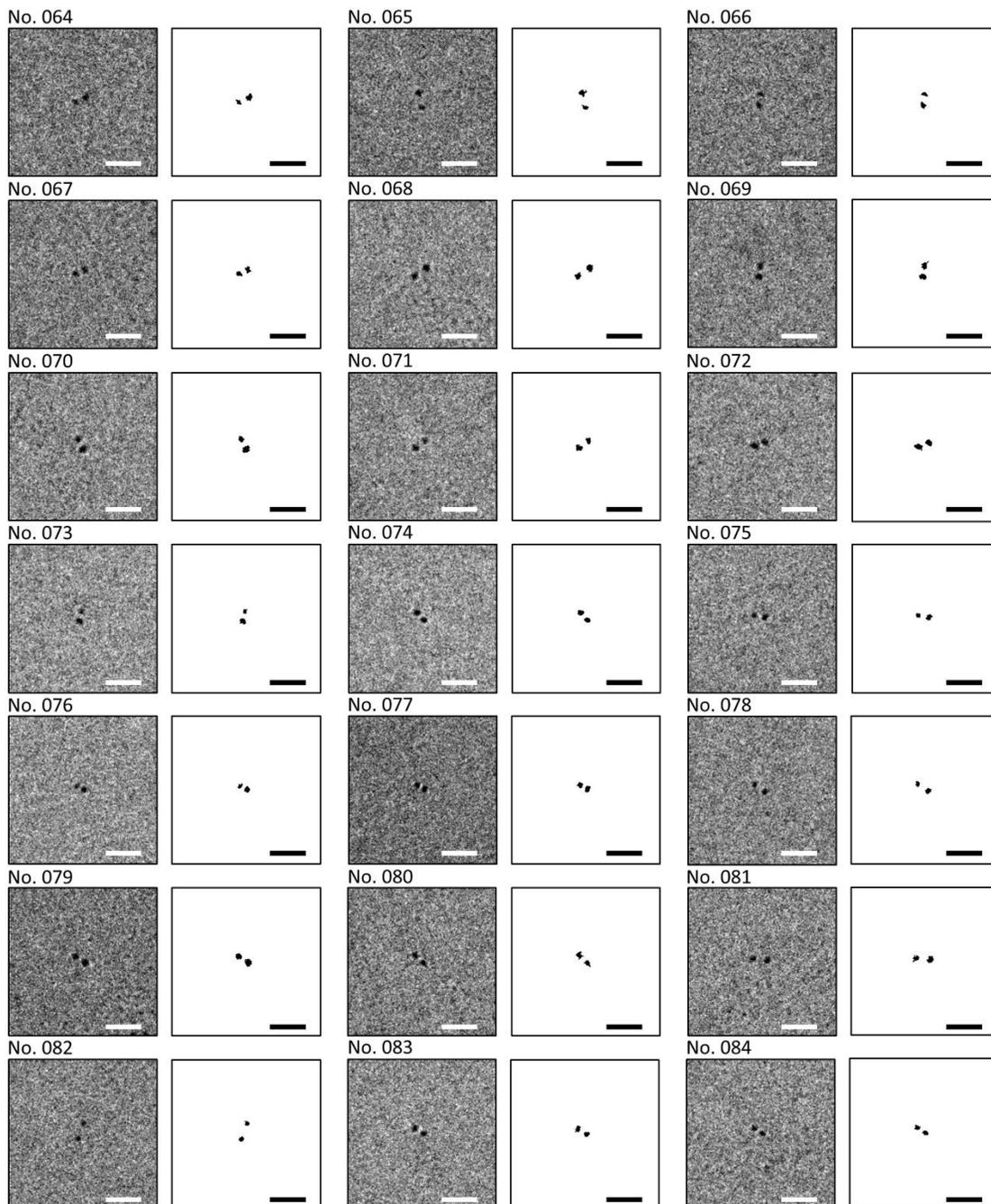
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G23C cyt-c)



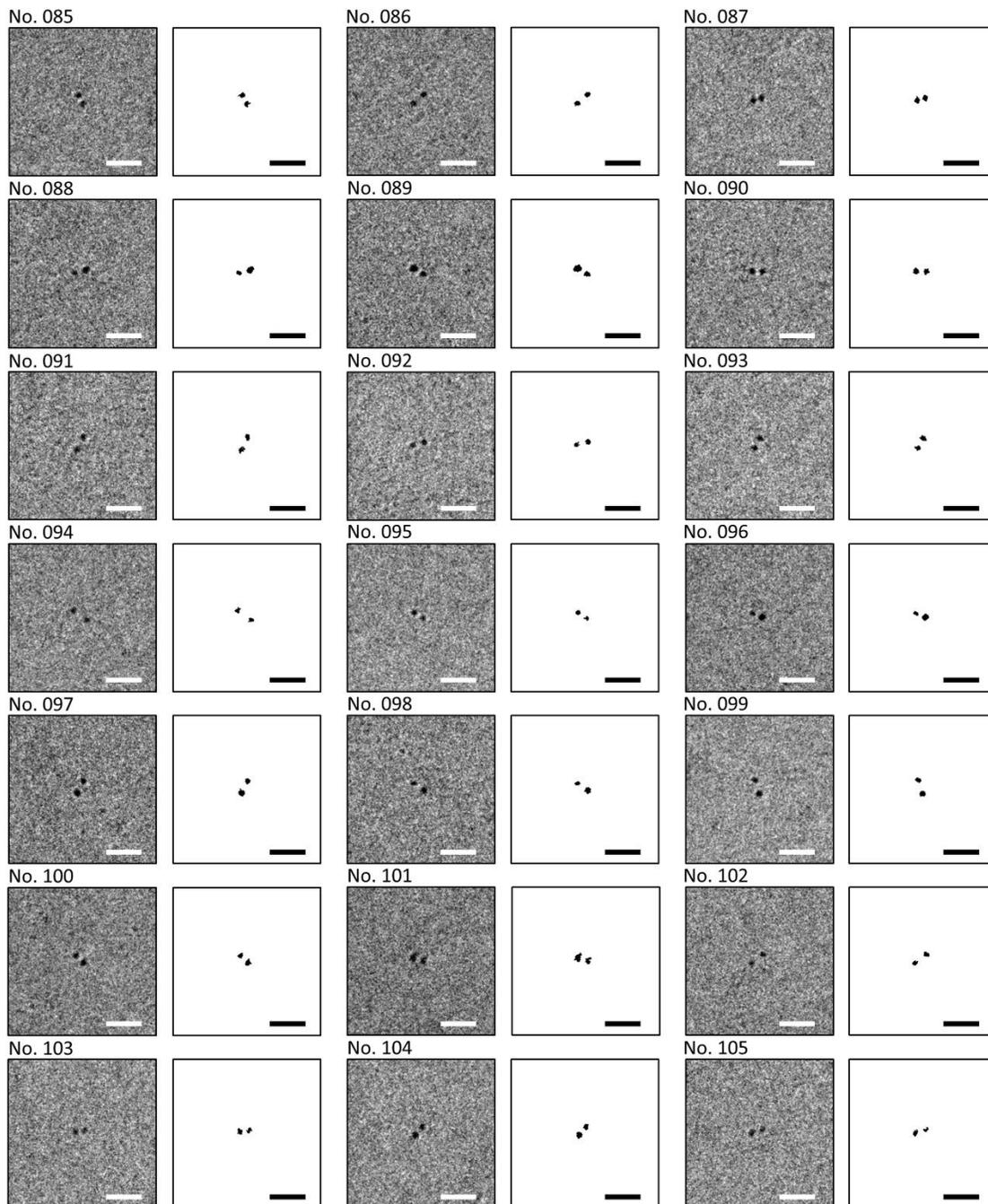
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G23C cyt-c)



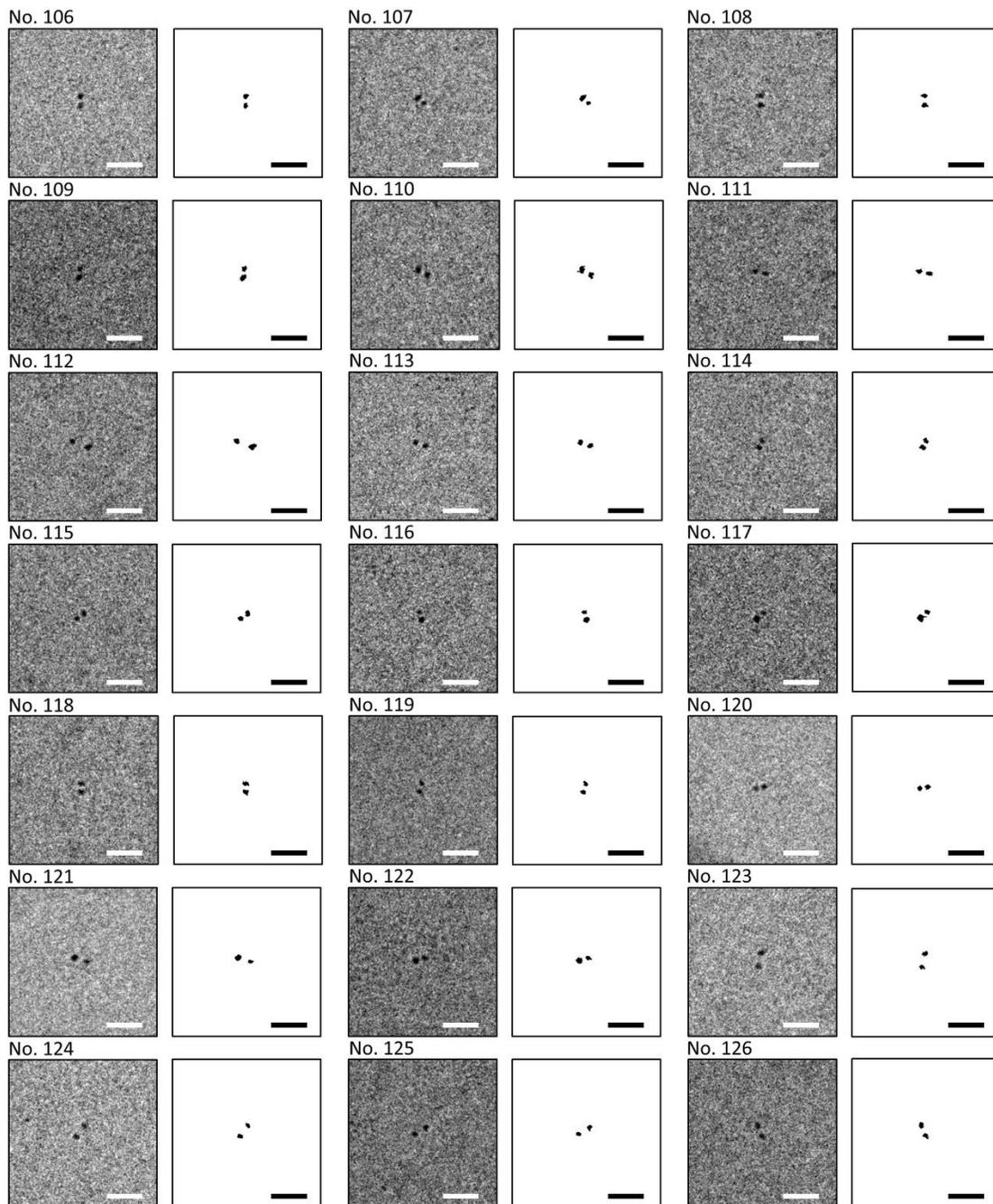
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G23C cyt-c)



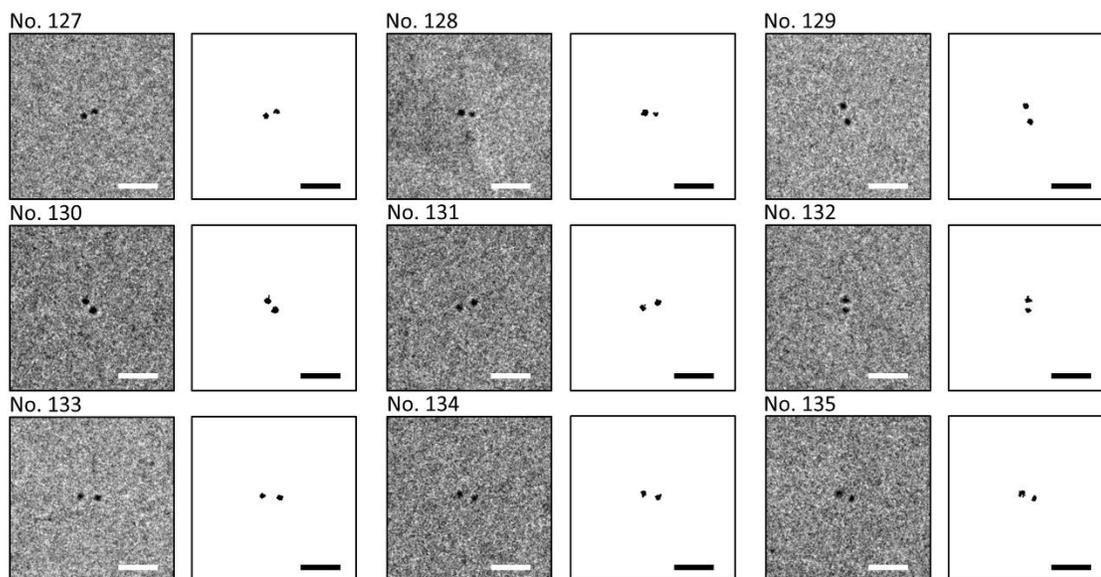
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G23C cyt-c)



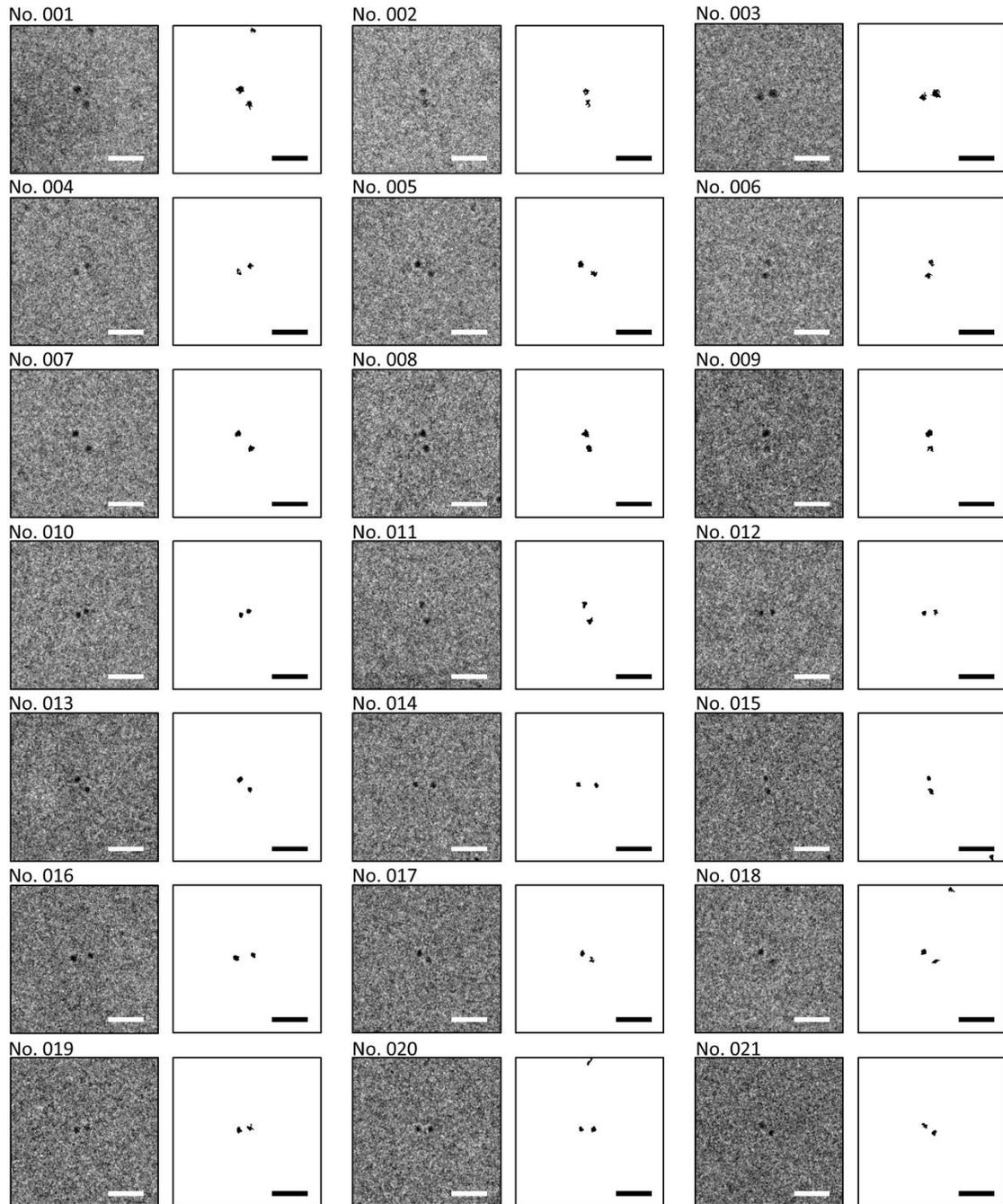
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G23C cyt-c)



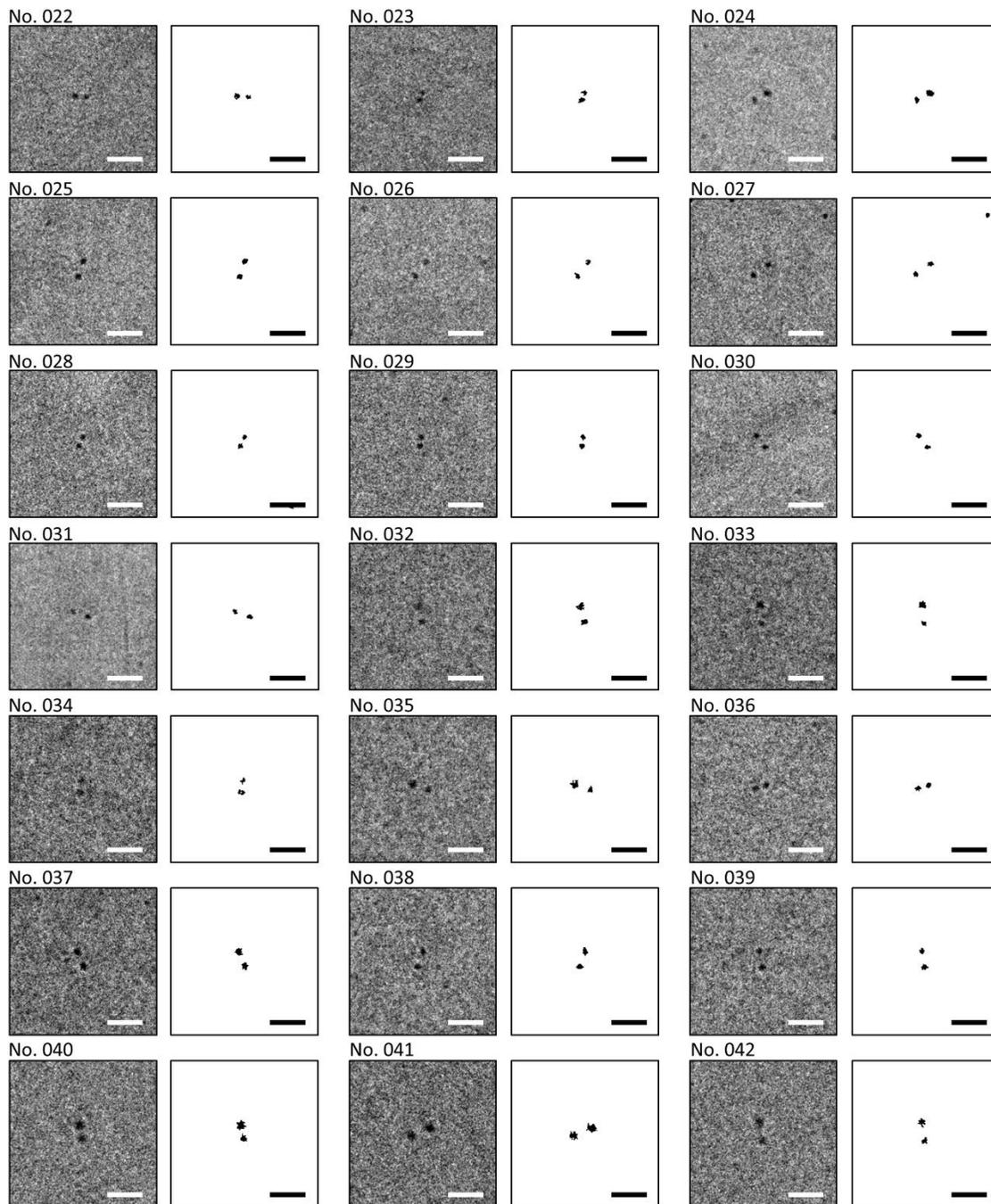
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G23C cyt-c)



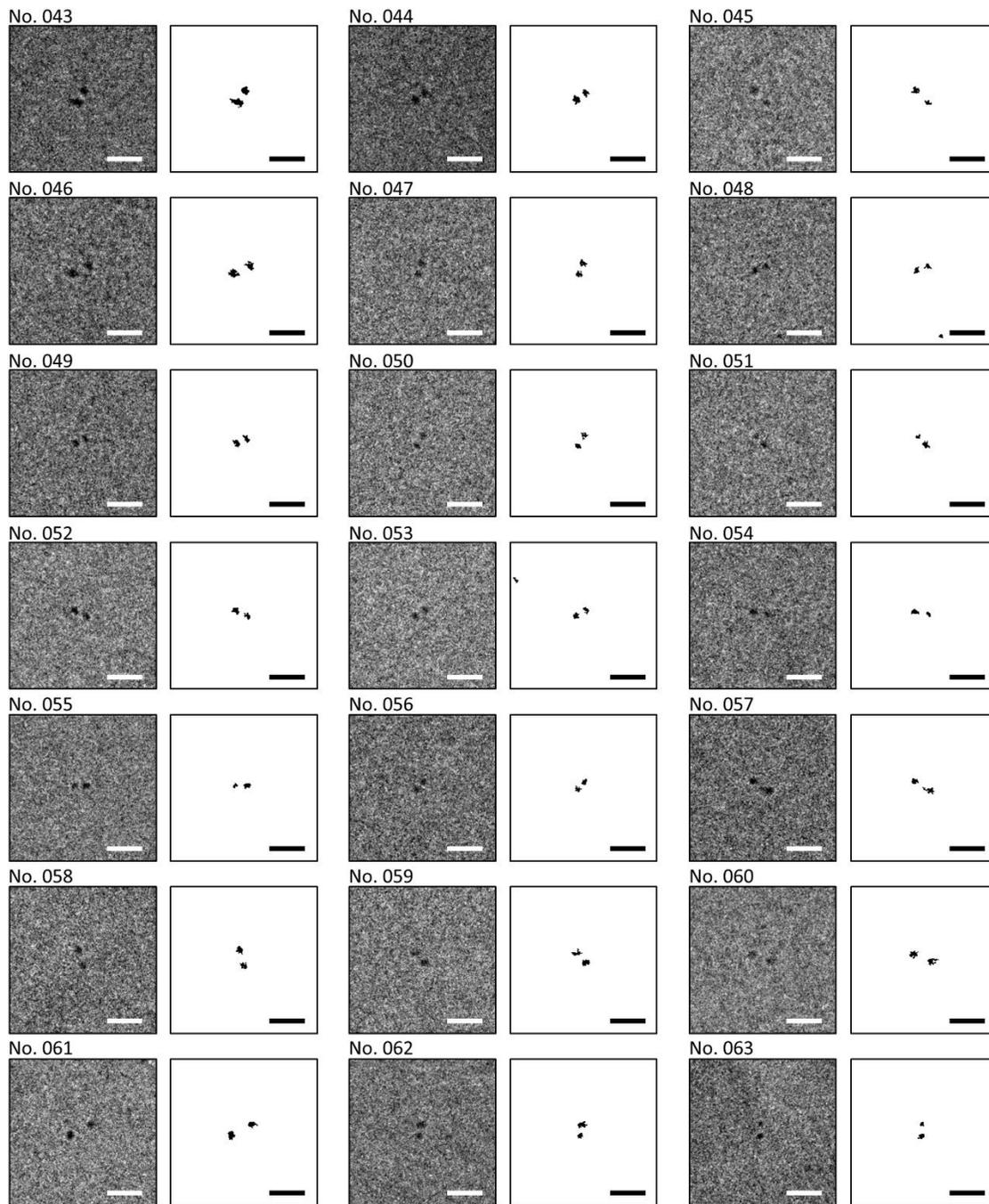
**Figure S8.** Cryo-EM images obtained from nanogold labeled G1C-G23C cyt-c. All the cropped images and corresponding background-removed images are presented. Scale bars are 10 nm.



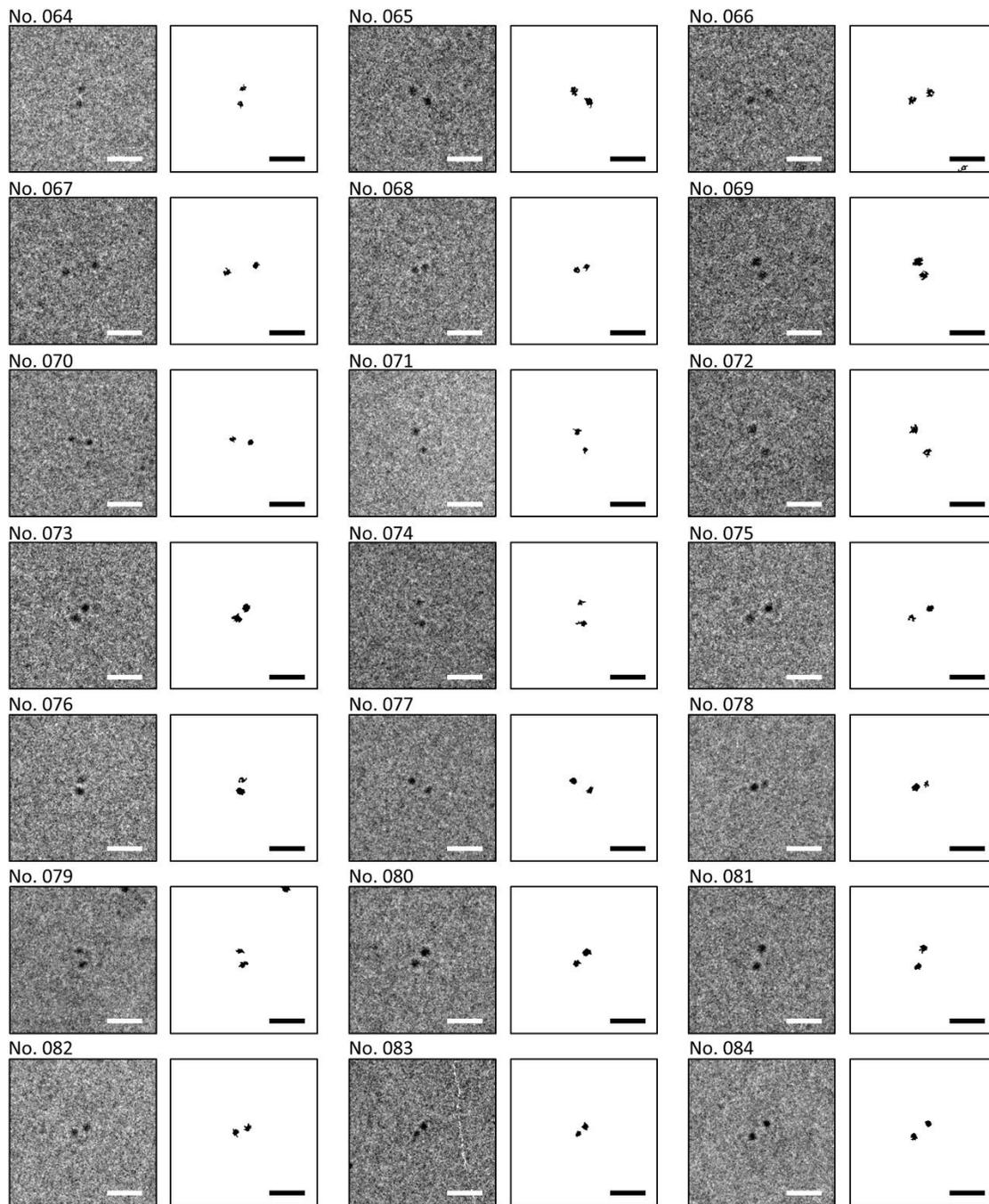
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c)



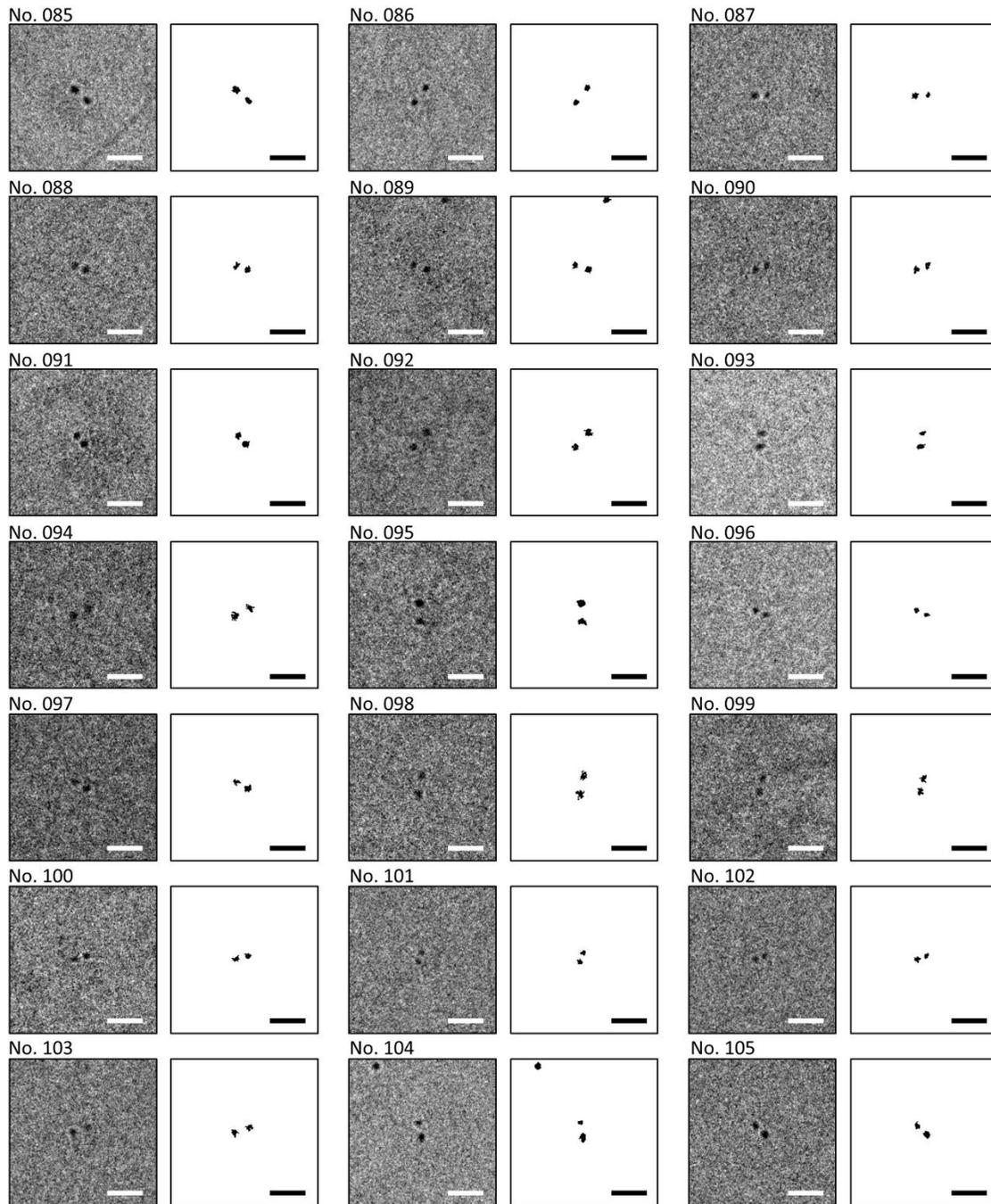
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c)



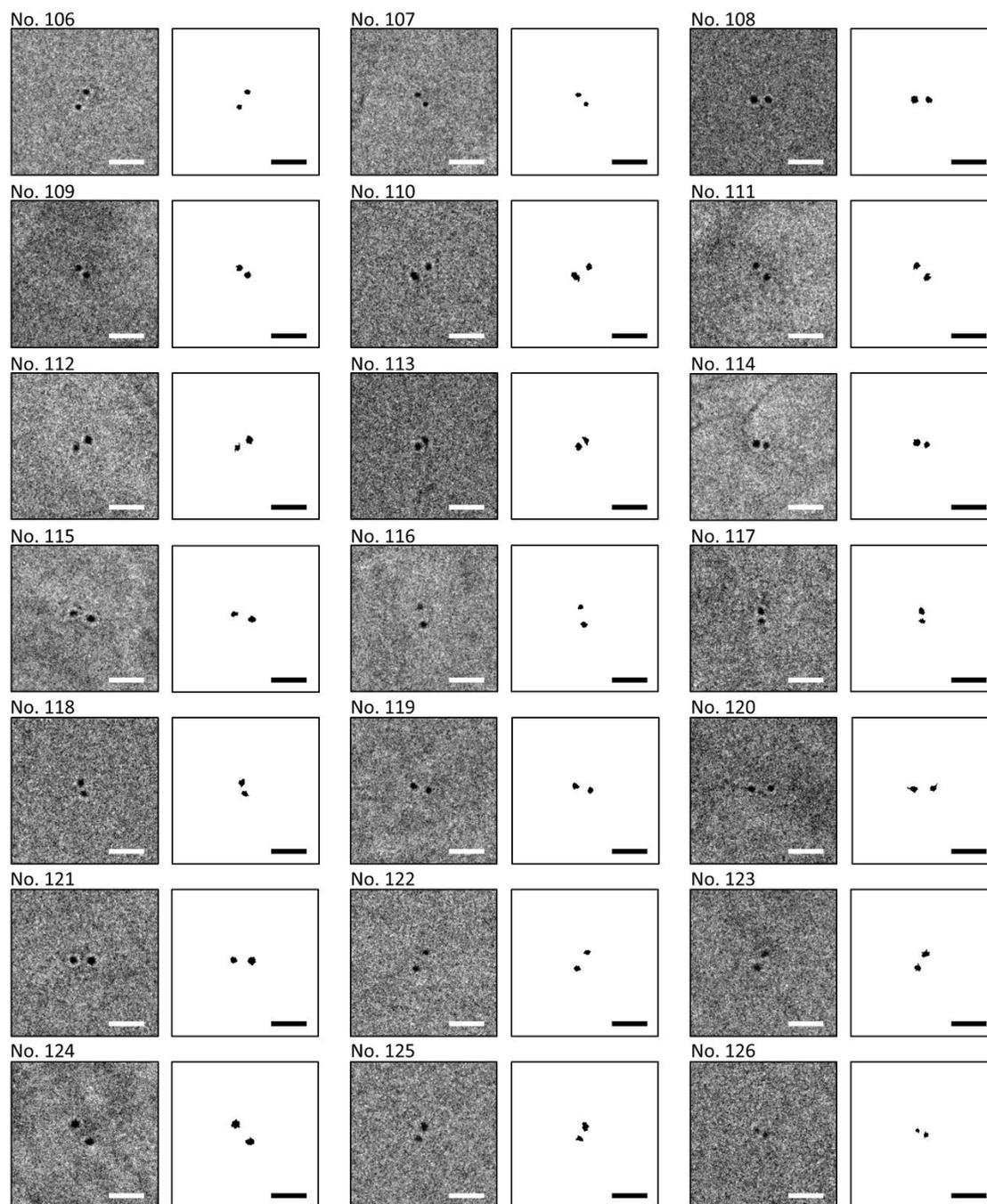
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c)



(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c)

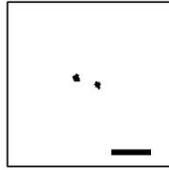
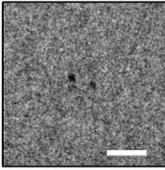


(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c)

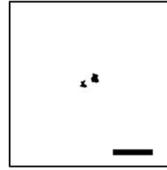
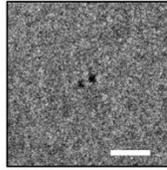


**Figure S9.** Cryo-EM images obtained from nanogold labeled G1C-G45C cyt-c. All the cropped images and corresponding background-removed images are presented. Scale bars are 10 nm.

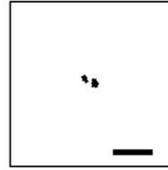
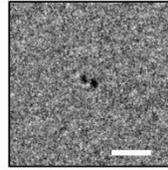
No. 001



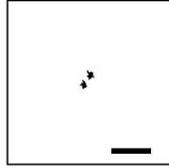
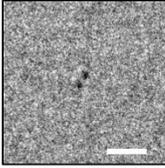
No. 002



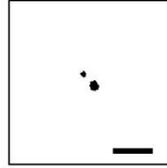
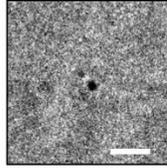
No. 003



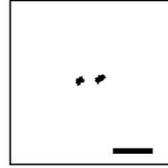
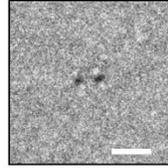
No. 004



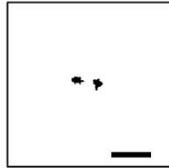
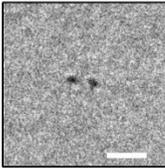
No. 005



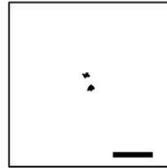
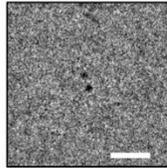
No. 006



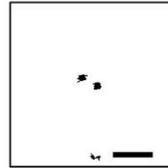
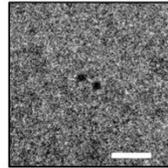
No. 007



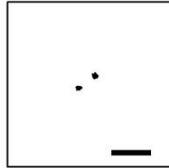
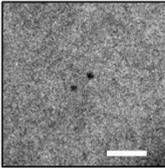
No. 008



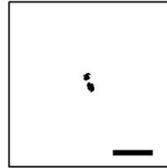
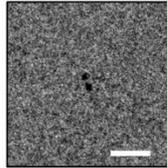
No. 009



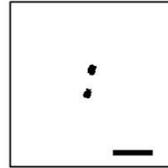
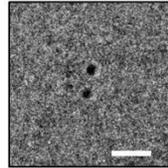
No. 010



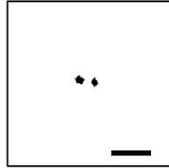
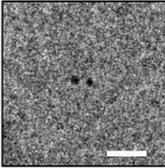
No. 011



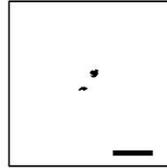
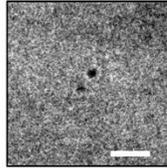
No. 012



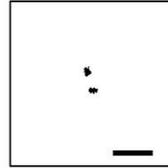
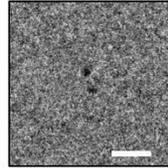
No. 013



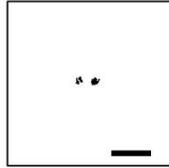
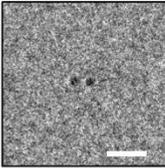
No. 014



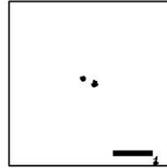
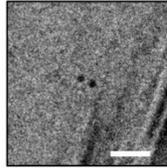
No. 015



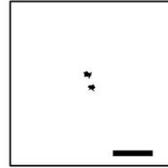
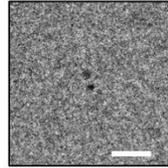
No. 016



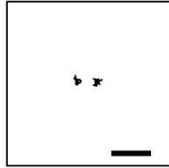
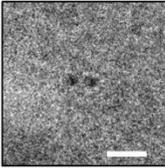
No. 017



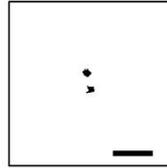
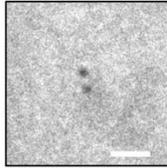
No. 018



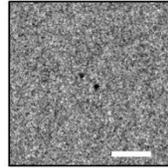
No. 019



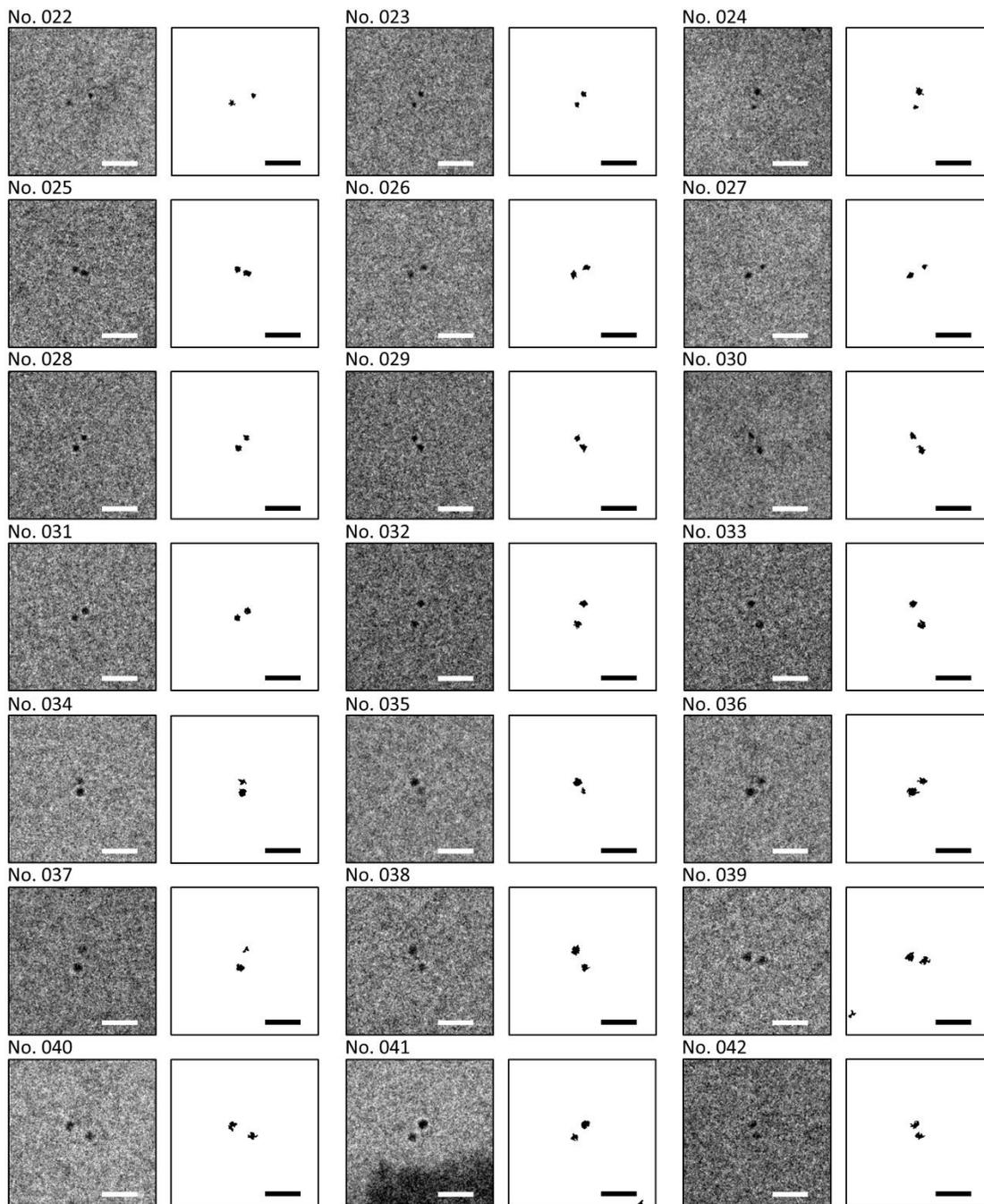
No. 020



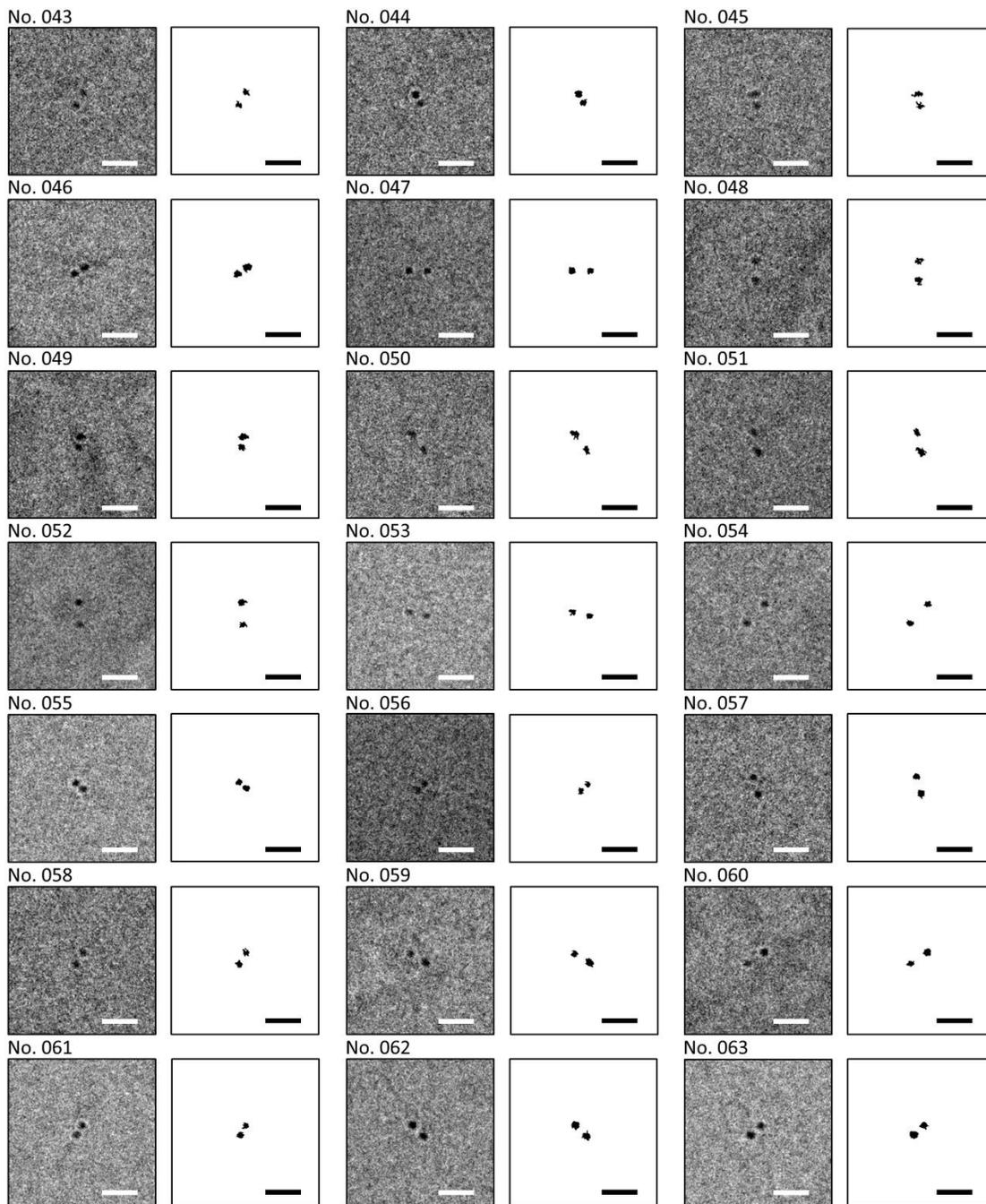
No. 021



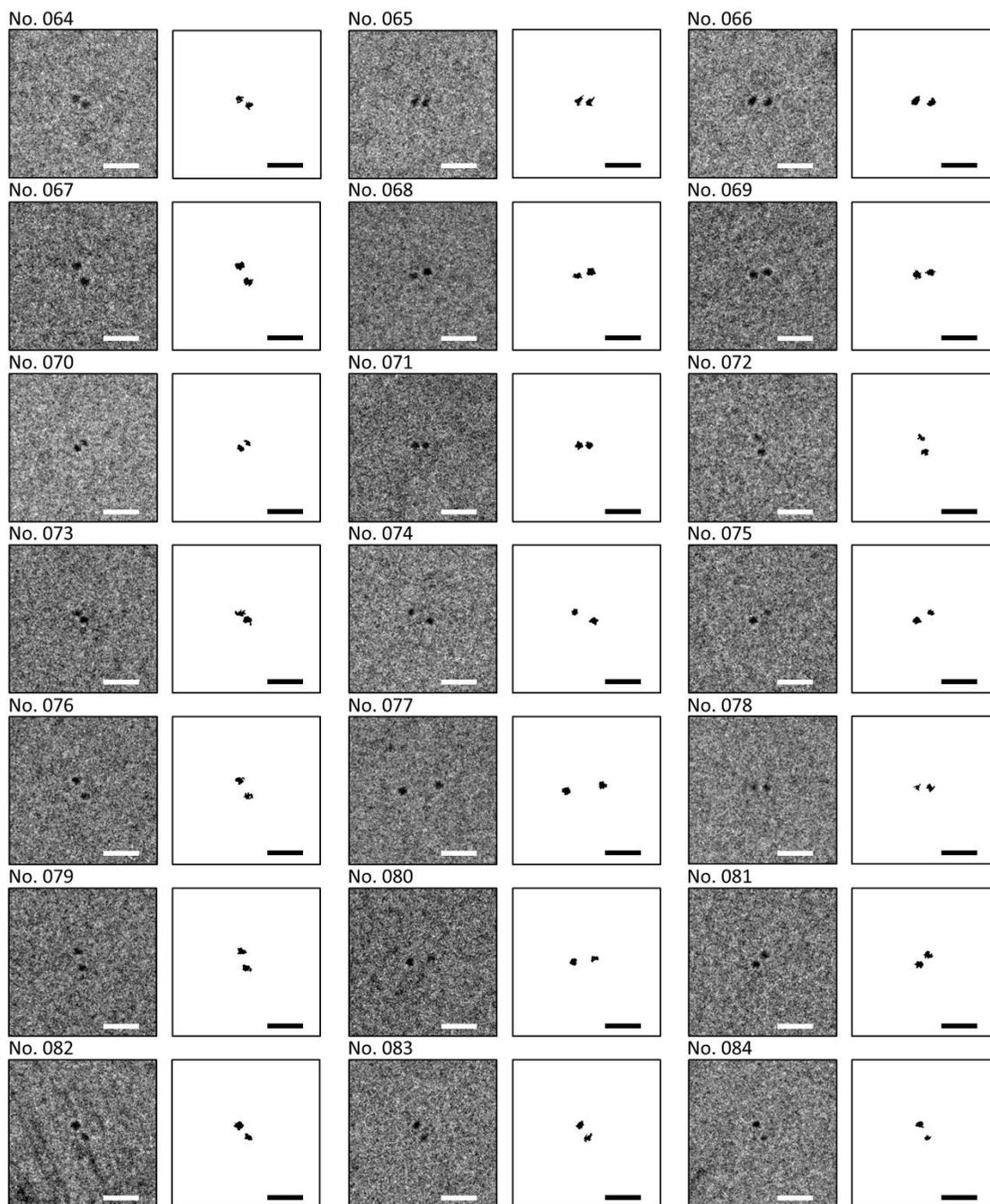
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



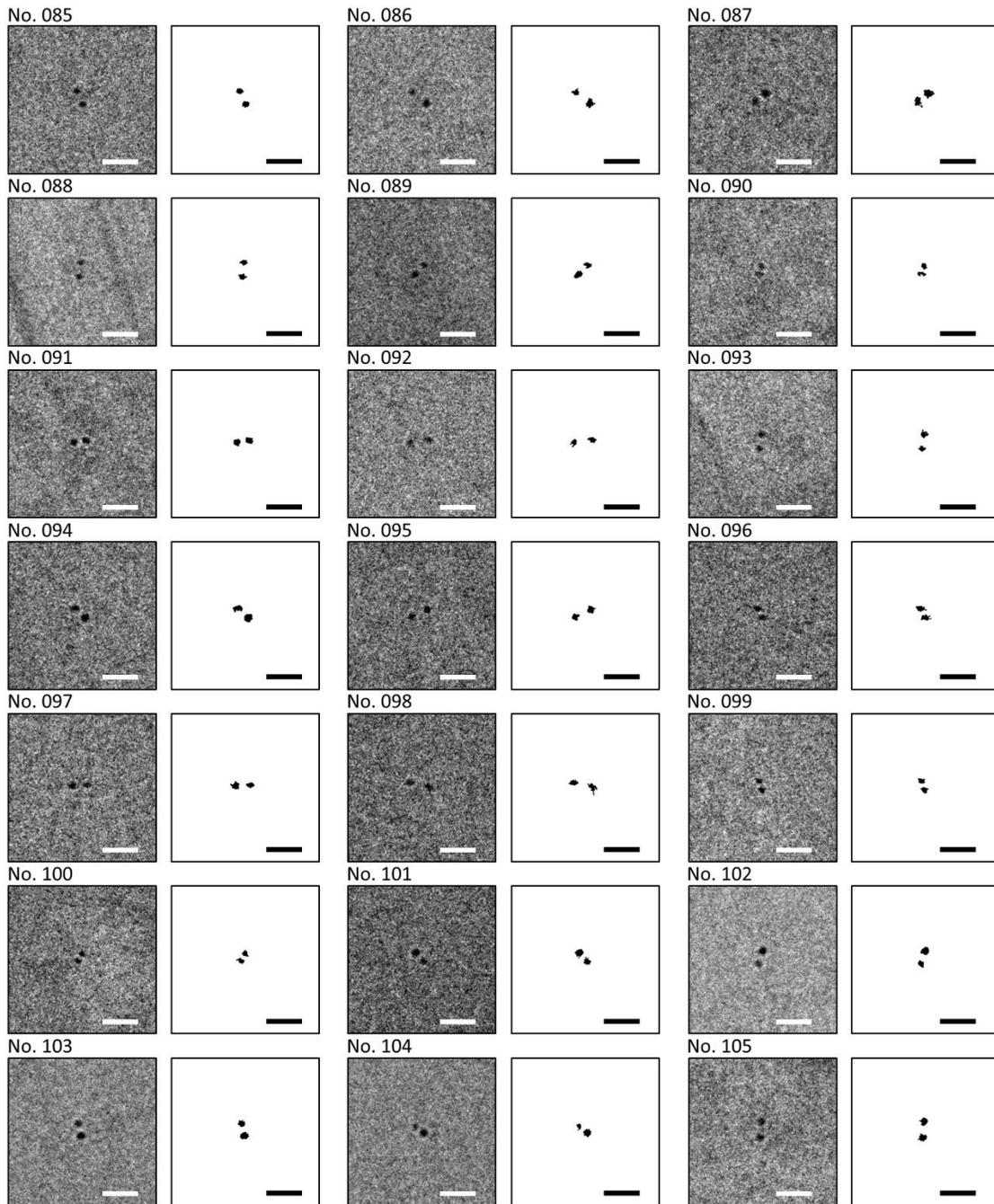
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



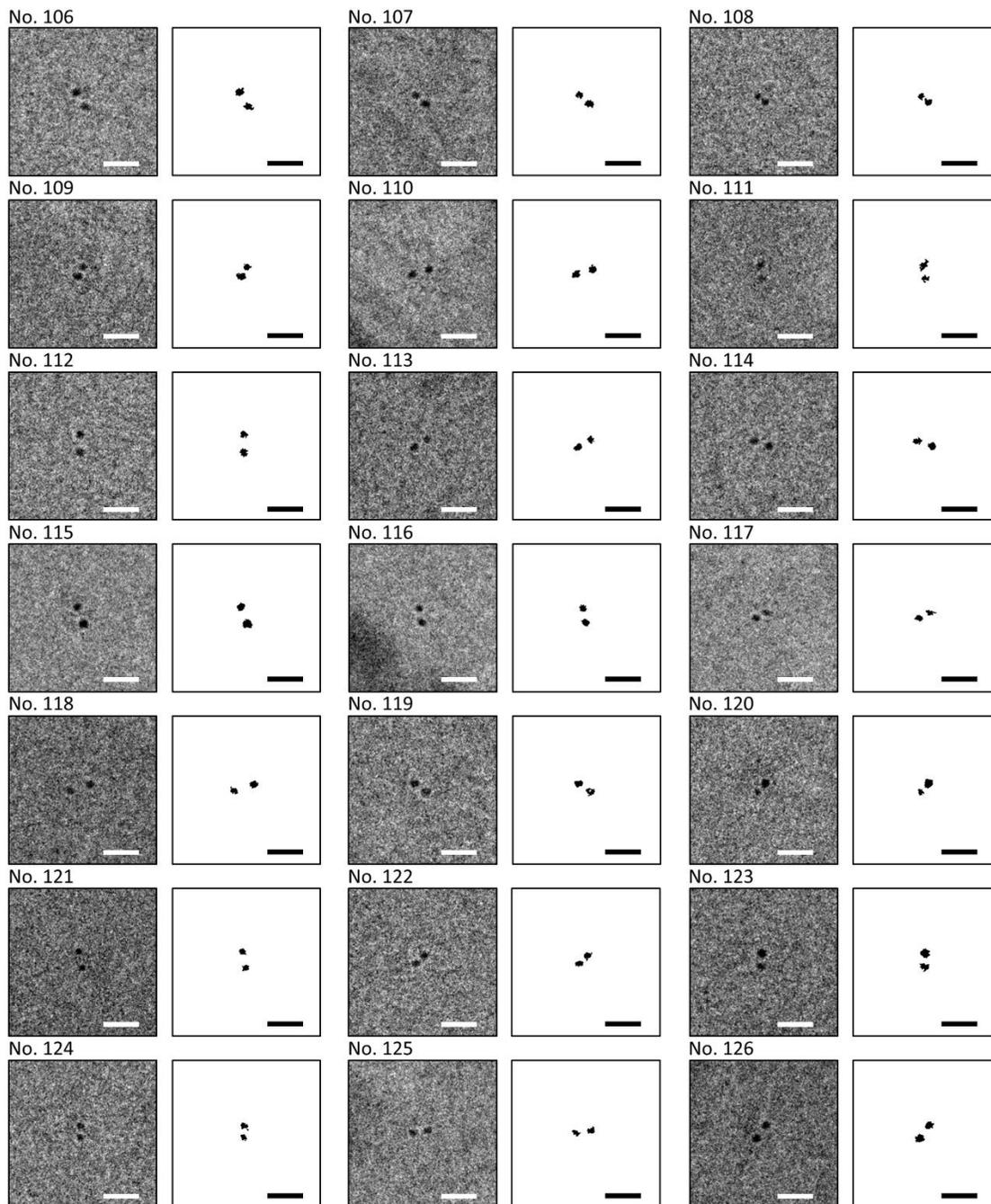
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



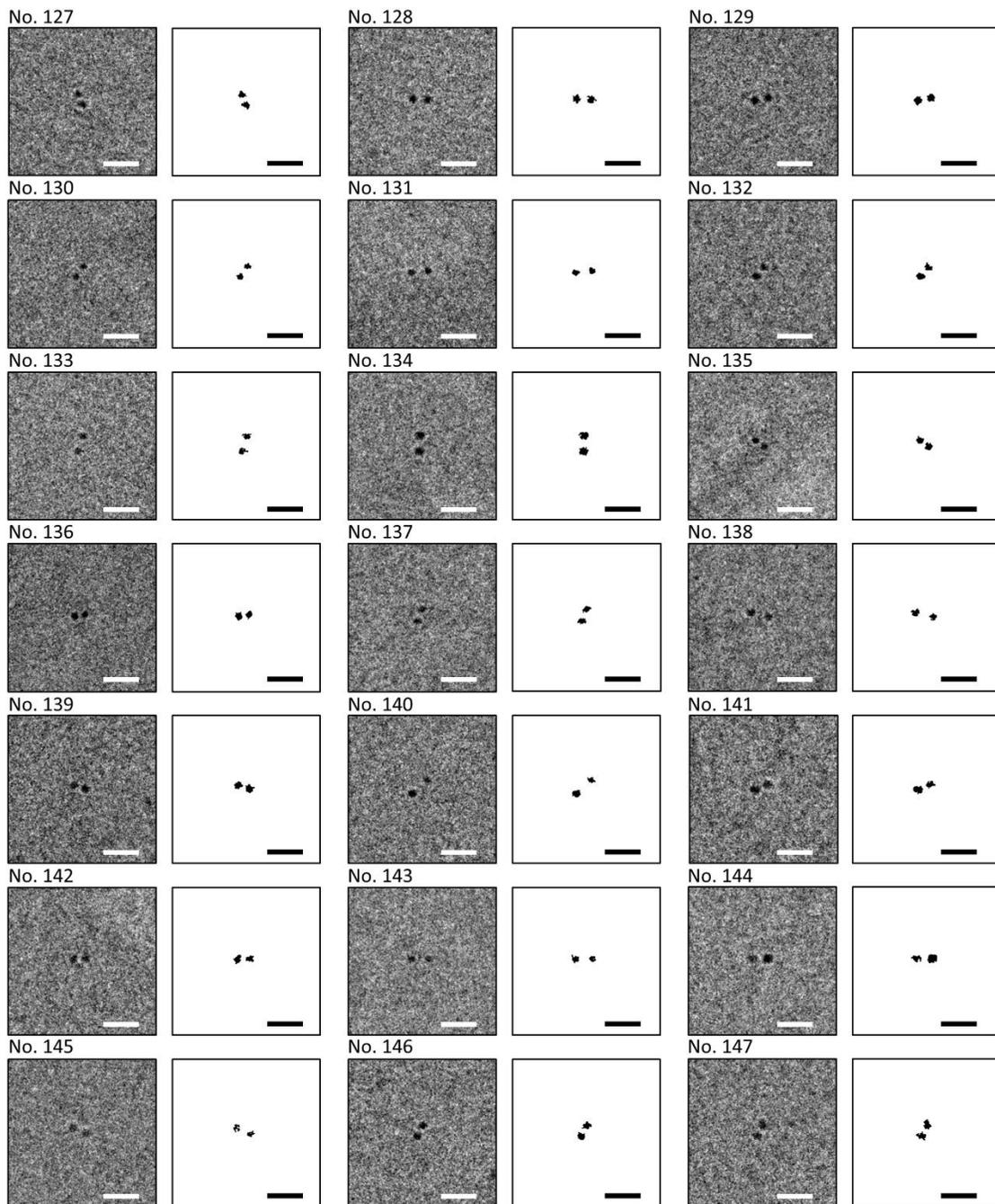
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



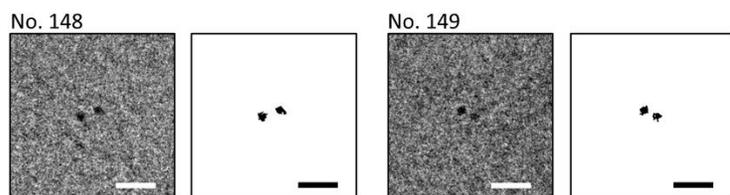
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



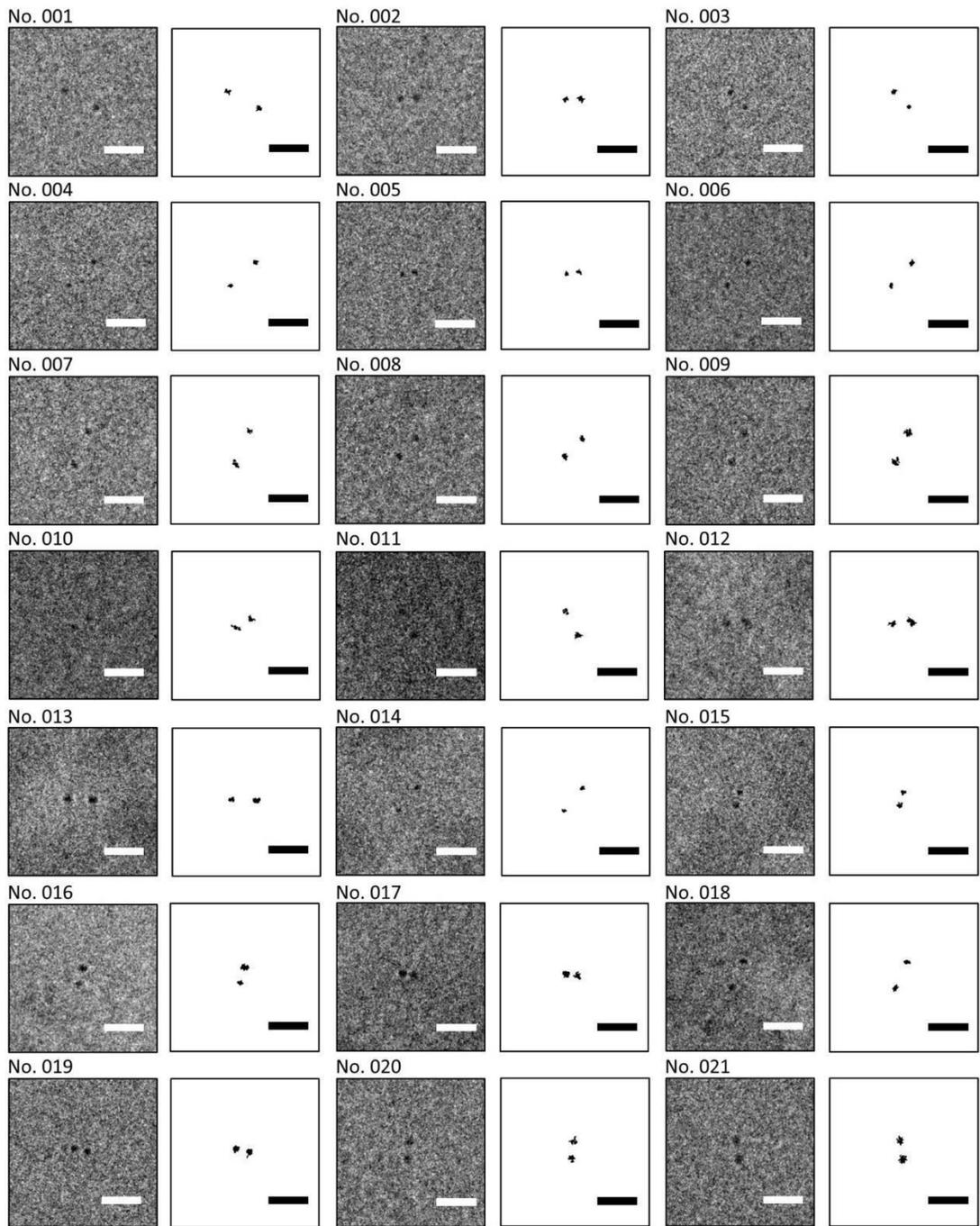
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



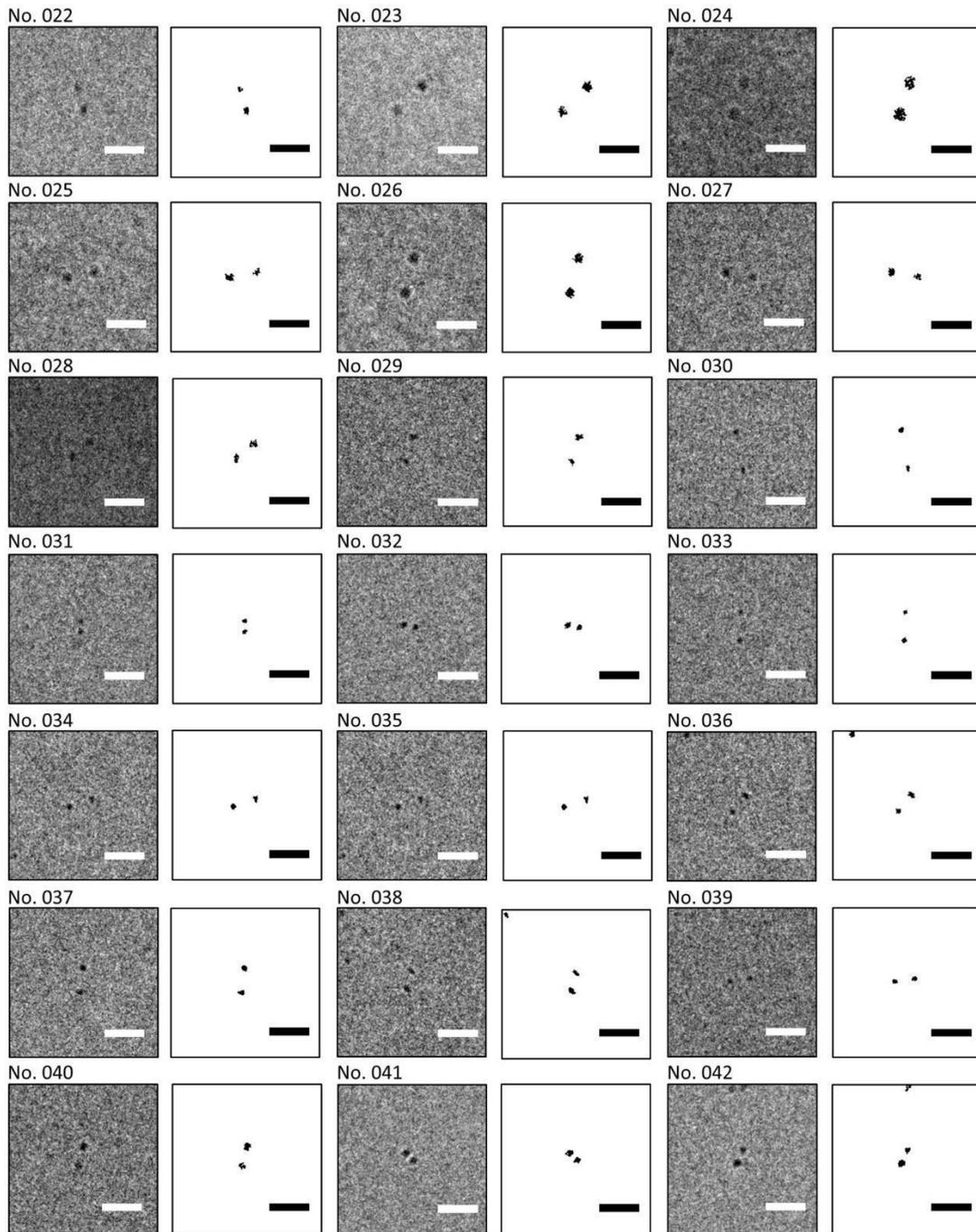
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-E104C cyt-c)



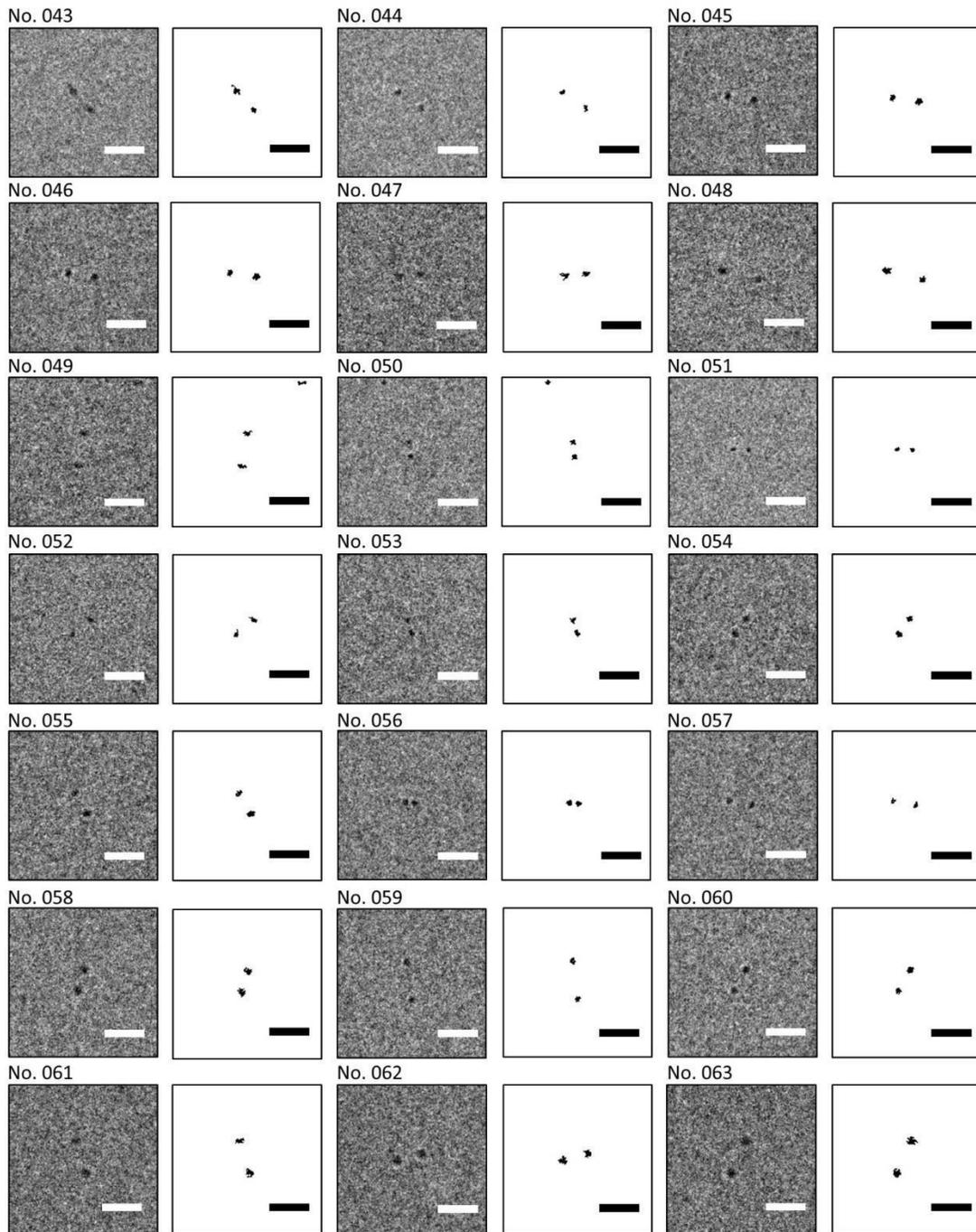
**Figure S10.** Cryo-EM images obtained from nanogold labeled G1C-E104C cyt-c. All the cropped images and corresponding background-removed images are presented. Scale bars are 10 nm.



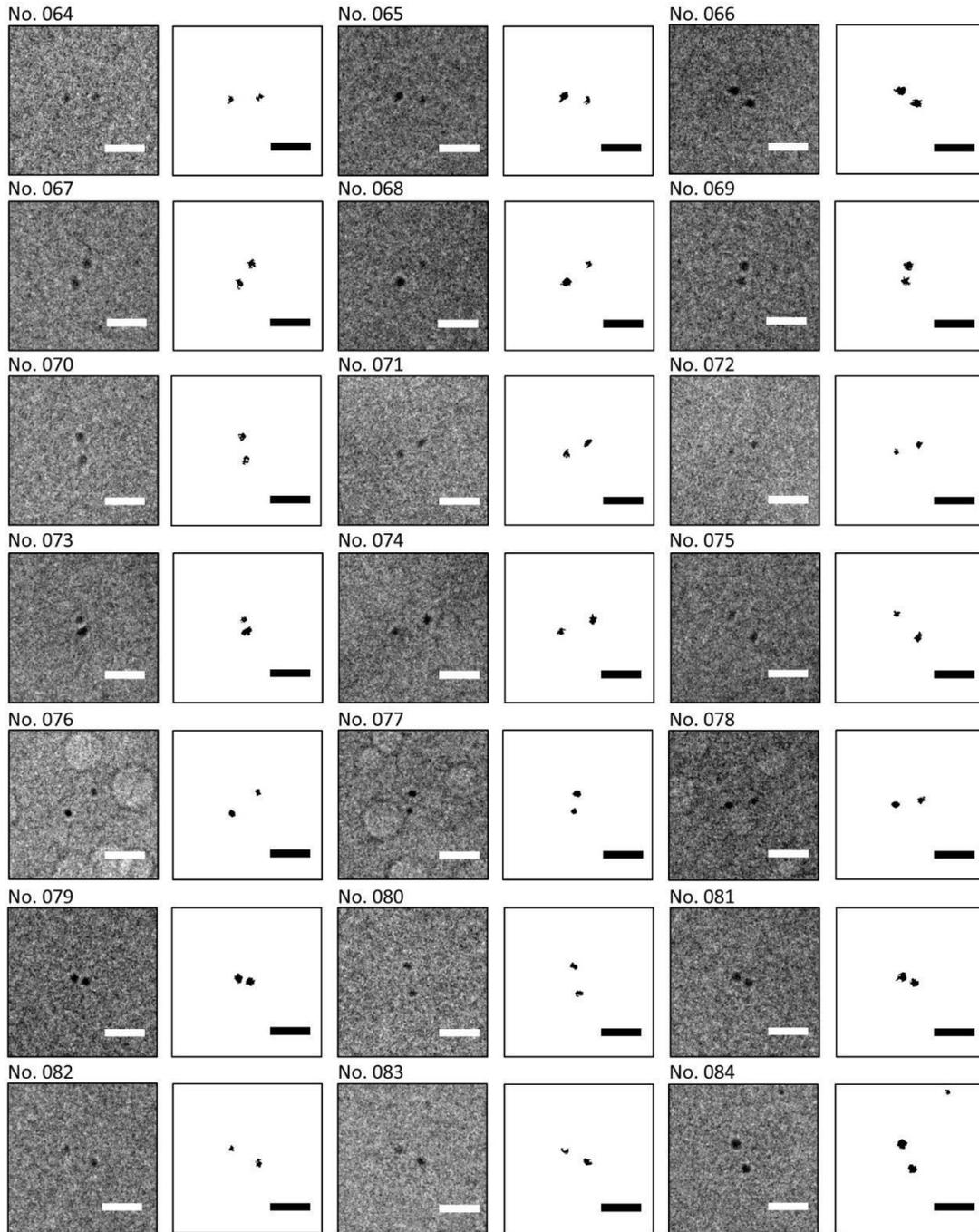
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



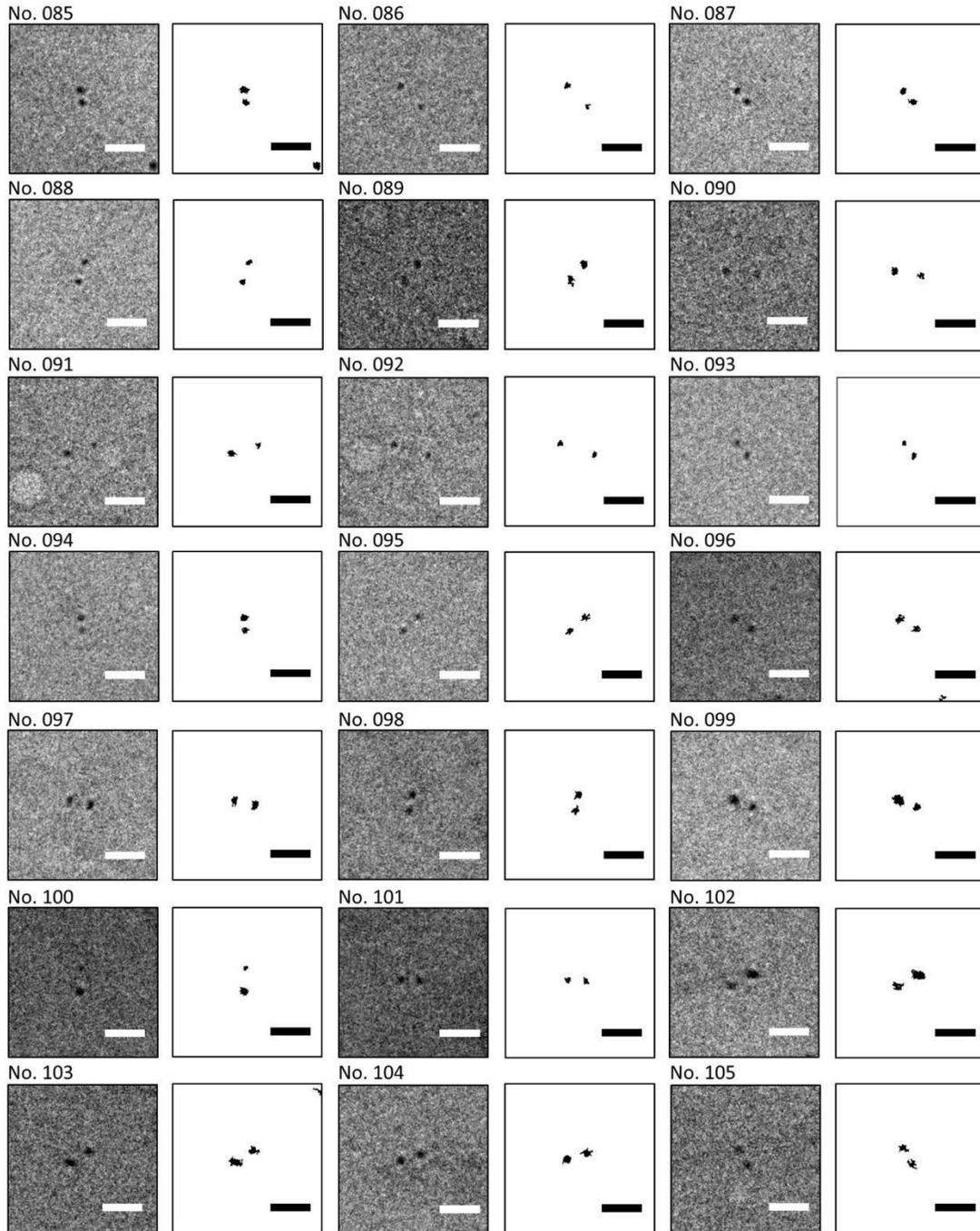
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



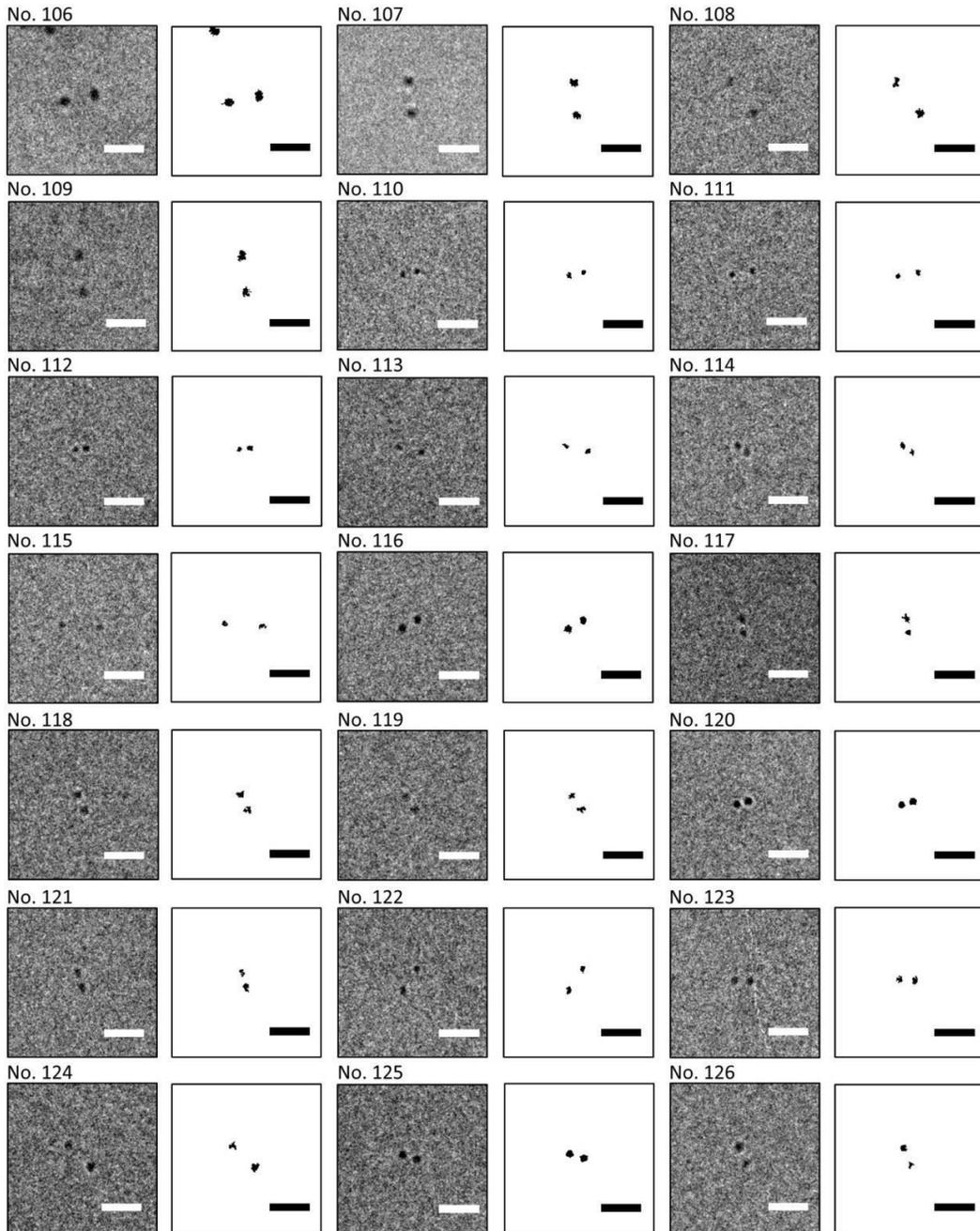
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



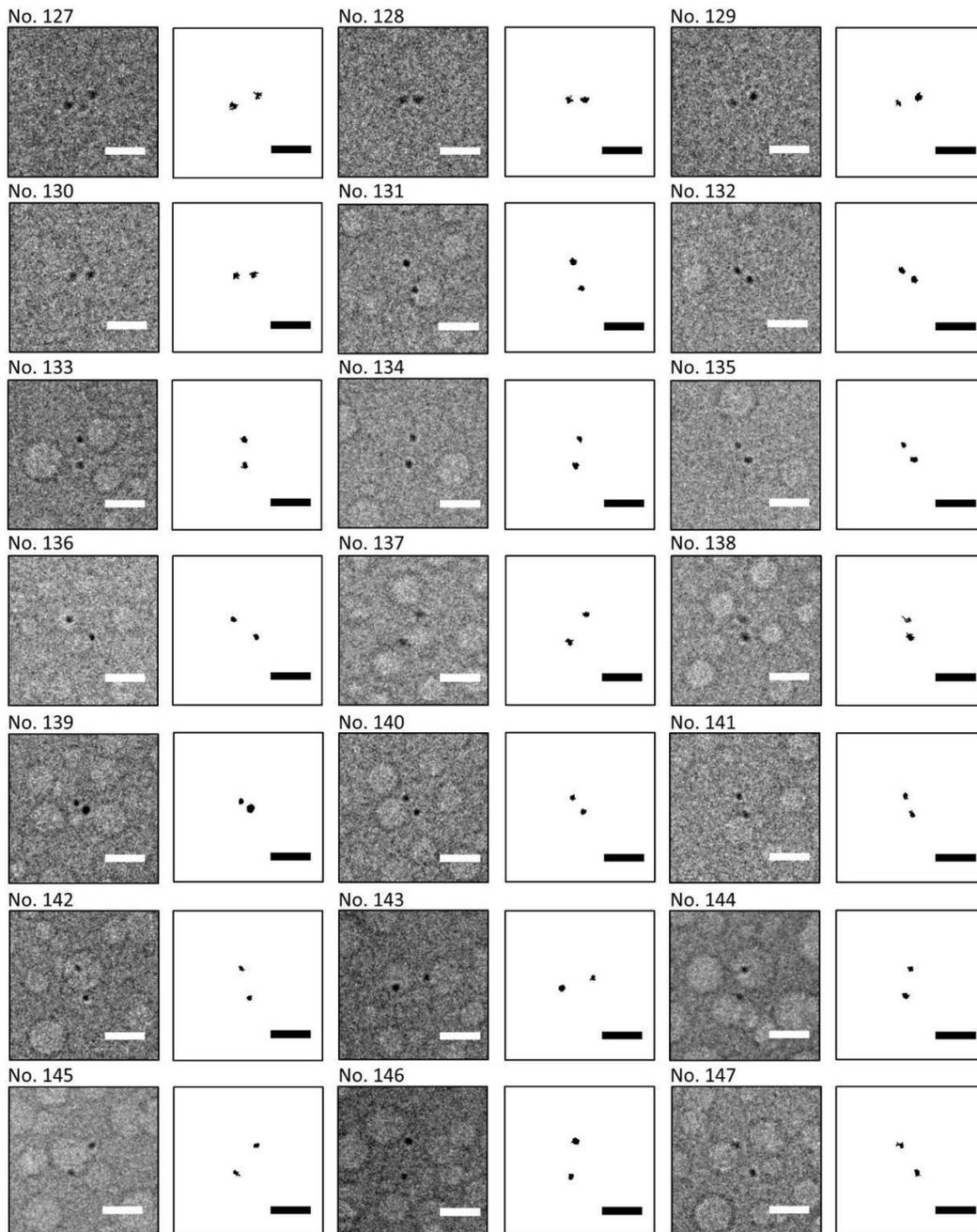
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



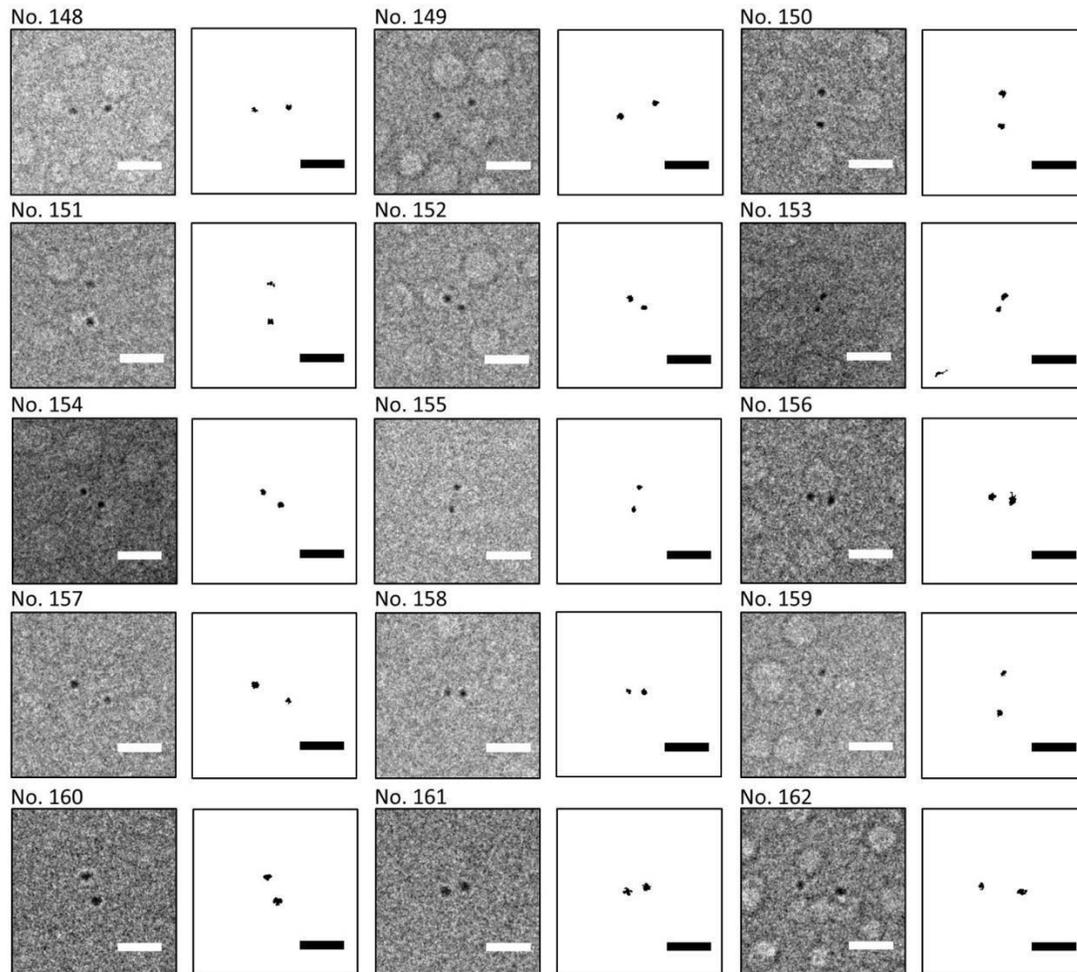
(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



(Continued from the previous page, Cryo-EM images of nanogold labeled G1C-G45C cyt-c in the unfolded state)



**Figure S11.** Cryo-EM images obtained from nanogold labeled G1C-G45C cyt-c in the unfolded state. All the cropped images and corresponding background-removed images are presented. Scale bars are 10 nm.