(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 February 2025 (06.02.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/029985 A1

(51) International Patent Classification:
G01N 1/42 (2006.01) *C07K 14/435* (2006.01)

(21) International Application Number:
PCT/US2024/040485

(22) International Filing Date:
01 August 2024 (01.08.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/516,902 01 August 2023 (01.08.2023) US

(71) Applicant: **WISCONSIN ALUMNI RESEARCH FOUNDATION** [US/US]; 614 Walnut Street, Madison, Wisconsin 53726 (US).

(72) Inventor: **LIM, Ci Ji**; 6 E Newhaven Cir, Madison, Wisconsin 53717 (US).

(74) Agent: **VANHEYNINGEN, Tambryn** et al.; QUARLES & BRADY LLP, 33 East Main Street, Suite 900, Madison, Wisconsin 53703 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: COMPOSITIONS AND METHODS FOR PREPARATION OF A CRYOGENIC ELECTRON MICROSCOPY SAMPLE

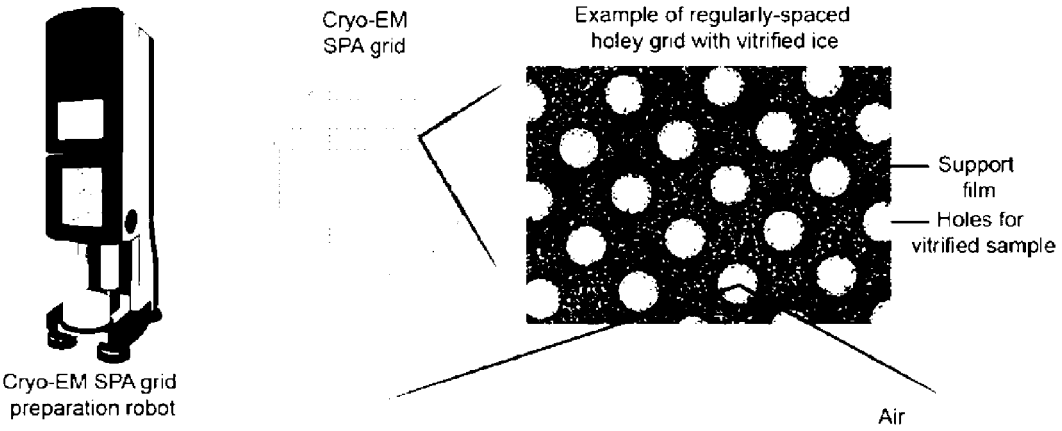

FIG. 1A

(57) Abstract: Compositions and methods for preparing a biological sample for cryo-electron microscopy (cryo-EM). The methods can include adding one or more protectant proteins to a biological sample during preparation of a cryo-EM sample. The compositions include one or more of the protectant proteins.

WO 2025/029985 A1

COMPOSITIONS AND METHODS FOR PREPARATION OF A CRYOGENIC ELECTRON MICROSCOPY SAMPLE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 63/516,902 filed on August 1, 2023, the contents of which are incorporated by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under GM131023 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] The contents of the electronic sequence listing (2024-07-31_960296.04537_WIPO_Sequence_listing_XML.xml; Size: 8,267 bytes; and Date of Creation: August 1, 2024) is herein incorporated by reference in its entirety.

BACKGROUND

[0004] Cryogenic electron microscopy (cryo-EM) has dramatically remodeled the structural biology research landscape. Large biological macromolecular assemblies that were once deemed structurally unsolvable via established methods (e.g., crystallography) can now be revealed in their entirety, with their functional conformations captured.

SUMMARY

[0005] In one aspect, a method for preparation of a cryogenic electron microscopy (cryo-EM) sample is provided. The method can include adding one or more protectant proteins to a biological sample during preparation of the cryo-EM sample, where the one or more protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3.

[0006] In another aspect, a method for performing cryogenic electron microscopy (cryo-EM) is provided. The method can include adding one or more first protectant proteins to a first portion of a biological sample to provide a first cryo-EM sample, where the one or more first protectant

proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3. The method can also include adding one or more second protectant proteins to a second portion of the biological sample to provide a second cryo-EM sample, where the one or more second protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3, and where the one or more second protectant proteins are different than the one or more first protectant proteins.

[0007] In yet another aspect, a composition for preparing a cryogenic electron microscopy (cryo-EM) sample is provided. The composition can include one or more protectant proteins, where the one or more protectant proteins comprise a truncated RvLEAM1 protein comprising an amino acid sequence at least 85% identical to SEQ ID NO: 2, a truncated dAFP-1 protein comprising an amino acid sequence at least 85% identical to SEQ ID NO: 3, or both.

[0008] In another aspect, a construct is provided. The construct can include a promoter operably connected to a polynucleotide encoding a polypeptide of SEQ ID NO: 2, SEQ ID NO: 3, a polypeptide having at least 90% identity to SEQ ID NO: 2 or a polypeptide having at least 90% identity to SEQ ID NO: 3.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1A is a schematic depiction of one example cryo-EM single particle analysis (SPA) grid preparation system.

[0010] FIG. 1B is a schematic depiction of one example of a commonly perceived sample distribution in a thin film of vitrified ice.

[0011] FIG. 1C is a schematic depiction of one example of a conventional sample distribution in a thin film of vitrified ice.

[0012] FIG. 1D is a schematic depiction of one example of sample distribution in a thin film of vitrified ice with protectant proteins at the air-water interface, according to aspects of the present disclosure.

[0013] FIG. 2A is a schematic depiction of a 3D confocal fluorescence microscopy system for evaluating a protectant protein.

[0014] FIG. 2B is a schematic depiction of Z-stacks from the system of FIG. 2A.

[0015] FIGS. 3A-3D. Cryo-EM Sample Grid Preparation and Potential Role of LEA Proteins on AWI Damage Mitigation. Schematic illustration of the process and outcomes of preparing cryo-EM SPA grids and the hypothesized effects of LEA proteins on sample structural integrity. FIG.

3A shows a Cryo-EM SPA grid preparation robot alongside a detailed view of a typical holey grid with vitrified ice, which is used for embedding protein samples. FIGS. 3B-3D depict different states of protein sample distribution within the vitrified ice. FIG. 3B depicts an ideal sample distribution is shown where proteins are evenly dispersed without any structural damage. FIG. 3C depicts the typical impact of air-water interface (AWI) damage on protein samples, leading to preferred orientation, complex dissociation, and protein denaturation. FIG. 3D depicts how LEA proteins can protect the sample by forming a barrier at the AWI, which significantly mitigates these damages and preserves protein integrity in vitrified ice.

[0016] FIGS. 4A-4F. High-Resolution Cryo-EM Structure Determination of Fragile Human Polymerase Alpha-Primase and Polycomb Repressive Complex 2 Using Nematode AavLEA1. Cryo-EM single-particle analysis of human polymerase α-primase complex (PP), FIGS. 4A-4C, and polycomb repressive complex 2 (PRC2), FIGS. 4D-4F, both with the addition of Nematode AavLEA1. FIGS. 4A and 4D display representative micrographs for the complexes both alone and with AavLEA1 added at a 1:40 ratio, highlighting improved sample preservation due to LEA protein addition. FIGS. 4B and 4E depict 2D class averages, illustrating defined and consistent particle shapes with visible protein features when AavLEA1 was used. Finally, FIGS. 4C and 4F show reconstructed cryo-EM maps of the complexes from the AavLEA1 datasets, presented in two orientations.

[0017] FIGS. 5A-5F. A Truncated Form of LEA Protein from Tardigrade is Also an Effective AWI Damage Protectant. Cryo-EM single-particle analysis of protein complexes with $RvLEAM_{short}$. FIGS. 5A and 5D display representative micrographs of the polymerase-primase complex (PP) and polycomb repressive complex 2 (PRC2), each treated with $RvLEAM_{short}$ at a molar ratio of 1:6. FIGS. 5B and 5E show 2D class averages, which demonstrate the structural homogeneity and sample quality of the $RvLEAM_{short}$ protected samples. FIGS. 5C and 5F present the reconstructed cryo-EM maps of PP and PRC2, depicted in two orientations, achieving resolutions of 4.5 Å and 3.7 Å respectively, which underscore the efficacy of $RvLEAM_{short}$ in facilitating high-resolution structure determination of these complexes.

[0018] FIGS. 6A-6C. LEA-Protected Sample Particles are Adsorbed to Vitrified Ice Surfaces. Particle orientation distribution and cryo-electron tomography (cryo-ET) analysis for samples protected by LEA proteins. FIG. 6A displays Mollweide projections that compare the particle distribution for the polymerase α-primase complex (PP) and polycomb repressive complex 2

(PRC2) with AavLEA1 (1:40) and $RvLEAM_{short}$ (1:6) added respectively. Corresponding sphericity values demonstrate the degree of isotropy achieved in the cryo-EM map under each condition. FIGS. 6B and 6C show cryo-ET cross-sectional analysis of the spatial distribution of PP and PRC2 particles within the grid holes, respectively. These plots highlight the location of particles relative to the edge of the holes and identify regions affected by ice contamination. The axes are expressed in pixels, with a scale of 4.4 Å per pixel.

[0019] FIGS. 7A and 7B. Chemical Crosslinking Enhances Orientation Distribution in Cryo-EM Samples Protected by LEA Proteins. FIG. 7A displays Mollweide projections of the particle orientation distribution for the polycomb repressive complex 2 (PRC2) treated with chemical crosslinking for 2 minutes and 10 minutes, demonstrating improved isotropy as evidenced by the increased sphericity values of 0.83 and 0.97, respectively. FIG. 7B shows the high-resolution cryo-EM maps of PRC2 crosslinked for 10 minutes, presented in two orientations to highlight the detailed structural features achieved, with a global resolution of 3.1 Å. These results illustrate the efficacy of glutaraldehyde crosslinking in enhancing particle orientation distribution when combined with LEA protein protection.

[0020] FIGS. 8A-8C. Comparative Evaluation of LEA Proteins and CHAPSO in Protecting Against AWI Damage. FIG. 8A showcases the reconstructed cryo-EM map of the polymerase α-primase complex, visualized in two orientations, achieving a global resolution of 3.4 Å. FIG. 8B and FIG. 8C detail the ResLog and per-particle spectra SNR (ppSSNR) analyses respectively, comparing the effects of AavLEA1 and CHAPSO addition on particle image data and map reconstruction quality. The ResLog analysis in FIG. 8B illustrates the spatial frequency improvements associated with each additive, plotted against batch size on a logarithmic scale, indicating that AavLEA1 outperforms CHAPSO at higher spatial frequencies. FIG. 8C displays the logarithm of ppSSNR, demonstrating that AavLEA1 maintains higher SNR values across the range of spatial frequencies.

[0021] FIGS. 9A-9C. Representative micrograph comparison of different molar ratios of PP and AavLEA1. Micrograph on left with zoomed-in section on right. FIG. 9A: 1 μM PP with no added AavLEA1. No discernable intact particles of PP. FIG. 9B: 1.5 μM PP with 12 μM AavLEA1 (1:8 molar ratio). Several particles visible that are the expected size for intact PP. FIG. 9C: 1.5 μM PP with 60 μM AavLEA1 (1:40 molar ratio). Several particles visible that are expected size for intact PP.

[0022] FIG. 10. Structural alignment of human DNA Polymerase alpha-primase models obtained from the 1:40 AavLEA1 cryo-EM map and a published X-ray crystal structure. Structural alignment of a model refined from the 1:40 AavLEA1 cryo-EM map with the apo-state model of DNA polymerase alpha-primase (PP) from a published X-ray crystal structure (PDB 5EXR). The cryo-EM derived structure is depicted in pink, while the X-ray structure is shown in blue. The calculated root-mean-square deviation (RMSD) between the two models is 1.2 Å, indicating a high degree of structural similarity.

[0023] FIG. 11. Assessment of glutaraldehyde crosslinked Polycomb repressive complex 2. SDS-PAGE analysis was used to determine the extent of glutaraldehyde crosslinked Polycomb repressive complex 2 (PRC2) samples. Two crosslinking incubation time was tested: 2 and 10 min. Remaining cryo-EM samples were used for the analysis hence LEA proteins were present in the samples.

[0024] FIG. 12. Structural alignment of human Polycomb repressive complex 2 models obtained from the 10 min crosslinked sample 1:6 $RvLEAM_{short}$ cryo-EM map and a published cryo-EM structure. Structural alignment of a model refined from the cryo-EM map of crosslinked PRC2 with 1:6 molar ratio of $RvLEAM_{short}$ with an RNA-bound PRC2 dimer model (PDB 8FYH), which was determined using cryo-EM. The PDB 8FYH model is colored blue, while the structure from our cryo-EM data is colored pink. The calculated RMSD between the two models is 0.89 Å, highlighting the high structural similarity between these structures.

[0025] FIG. 13. Structural alignment of human DNA Polymerase alpha-primase models obtained from the 1:40 AavLEA1 and 4 mM CHAPSO cryo-EM maps. A model refined from the cryo-EM map with a 1:40 ratio of Polymerase alpha-primase (PP) to AavLEA1 is aligned to a model obtained using 4 mM CHAPSO. The structure with AavLEA1 is depicted in pink, while the CHAPSO-derived structure is shown in orange. The calculated RMSD between the two models is 0.73 Å.

[0026] FIG. 14. Quality assessment of DNA polymerase alpha-primase particles from the AavLEA1 and CHAPSO datasets using Rosenthal-Henderson plots. Particles from the DNA polymerase alpha-primase cryo-EM maps, derived from the AavLEA1 (1:40) and CHAPSO datasets, were analyzed using the Rosenthal-Henderson plot to assess the quality of the cryo-EM data.

DETAILED DESCRIPTION

[0027] Various aspects of the present disclosure relate to compositions and/or methods for preparation of a cryo-EM sample. The compositions can include one or more protectant proteins and the methods can include addition of one or more protectant proteins to a biological sample during preparation of the cryo-EM sample.

[0028] In cryo-EM single-particle analysis (SPA), biological samples can be embedded in a thin layer of vitrified ice (50-100 nm thick), which allows inspection of individual particles of a sample under a transmission electron microscope (TEM). This process of preparing a cryo-EM SPA sample grid requires the biological samples to sit momentarily in a thin film of aqueous solution before being rapidly frozen to achieve vitrification. FIG. 1A depicts an example cryo-EM single particle analysis (SPA) grid preparation system.

[0029] Preparing the cryo-EM sample grid can expose the samples to the air-water interface (AWI) at both sides of the thin film, and how the sample interacts with the AWIs can determine the outcome of the cryo-EM SPA project. Research has shown that the commonly held perception that the sample is pristinely preserved in a thin layer of vitrified ice is likely misguided (FIG. 1B). In particular, air-water interfaces (AWI) are formed during sample preparation that are often destructive to the biological samples (FIG. 1C). Recent efforts have provided solutions to overcome this problem. However, such processes are either too expensive to adopt (e.g., requiring specific instruments) and/or depend on following protocols that are technically challenging and not easily reproducible. Current technology, e.g., Spotiton or its commercial version, Chameleon, attempts to outrace the diffusion of fragile samples to the AWI by rapid vitrification (100s of milliseconds range). However, this sophisticated instrument not only is expensive at about 500,000 US Dollars, but it also requires a trained operator with deep experience to achieve any tangible results. Another current technology for fragile cryo-EM samples attempts to avoid interaction with the AWI by adsorption to a monolayer carbon or graphene support film overlaid on a cryo-EM grid. But this approach suffers from high background signal from the film (in the case of amorphous carbon support film), imaging throughput, and reproducibility challenges. Together, these current solutions represent a high technical and/or cost barrier for mainstream cryo-EM researchers to overcome.

[0030] The most vulnerable samples to damage by the AWI are not small or truncated protein domains, but the multi-subunit macromolecular assemblies that are highly dynamic in their

composition and conformation. These large, fragile molecular complexes (e.g., having a molecular weight (MW) > 150 kDa) hold the highest value in biomolecular structure-function relationship understanding. There is a need to develop new accessible methods and compositions to prevent AWI damage to these samples, which in turn, enable scientists to ask the most important biological questions at the frontiers of fundamental and human disease structural biology research.

[0031] As discussed above, various aspects of the disclosure include compositions and methods for preparation of a cryo-EM sample, which can include one or more protectant proteins or their use thereof. The one or more protectant proteins can protect biological samples from AWI damage during cryo-EM sample grid preparation. In various aspects, the one or more protectant proteins are derived from proteins of various organisms that provide cellular protection under extreme conditions, e.g., freezing, desiccation, or the like. One example protectant protein can be derived from and/or based on late embryogenesis abundant (LEA) proteins. LEA proteins are a class of small, helical-structured, and highly hydrophilic proteins that have been shown to accumulate to high levels for protection against dehydration and other stress conditions. Because LEA proteins maintain the integrity and stability of cell membranes, LEA proteins increase the cell's tolerance to dehydration stress, and reduce the osmotic and freezing damage during freezing.

[0032] One example LEA protein is the AavLEA1 protein from nematode, *Aphelenchus avenae*. The AavLEA1 protein is 14 kDa in mass and is involved in the desiccation protection of *Aphelenchus avenae* in harsh dry environments. Unstructured in solution, AavLEA1 forms a helical secondary structure upon adhesion to AWI. This biphasic behavior allows AavLEA1 to exhibit a specific affinity for AWIs with high residence stability. Its unstructured conformation in solution minimizes the likelihood of any strong cumulative signal in cryo-EM images. Studies have shown AavLEA1 was able to protect test enzymes from protein aggregation and/or damage during freeze-thaw processes. This protection function is achieved by AavLEA1 forming protecting barriers at the AWIs of air bubbles formed during the free-thaw process. AavLEA1 can be expressed and purified to high yield (in milligrams) and purity (>95%) using standard bacterial expression and purification systems. In another non-limiting example, *Ramazzottius varieornatus* can survive in extreme environments, including desiccation, and utilizes the RvLEAM1 protein. The protect-ability of the RvLEAM1 protein is provided by accumulating small proteins at the AWI, similar to the LEA proteins. In yet another non-limiting example, *Dendroides canadensis* can survive subzero temperatures due to a family of antifreeze proteins (AFPs) that accumulate

during freezing conditions. The small *D. canadensis* AFP (DAFP) protein forms a beta-helix structure when adsorbed to the AWI.

[0033] The smaller size of the protectant proteins, e.g., exhibiting smaller MW relative to the biological sample, imparts certain benefits to the compositions and methods disclosed herein. For instance, in certain aspects, since the protectant proteins are smaller than the biological sample the protectant proteins are less likely to interfere with downstream cryo-EM image analysis and/or processing of the biological sample. In the same or alternative aspects, the small size of the protectant proteins relative to the size of the biological sample means that such protectant proteins have a faster rate of diffusion compared to the biological sample and thus may outrace the biological sample to the AWI, thereby forming a protective barrier at the AWI (see e.g., FIG. 1D). Additionally, the small size of the one or more protectant proteins makes such proteins amenable to expression, purification, and handling processes.

[0034] Compositions, Constructs, and Kits

[0035] In various aspects, the compositions disclosed herein include one or more protectant proteins. As discussed above, the one or more protectant proteins can be derived from one or more proteins from various organisms that provide cellular protection under extreme conditions, such as desiccation, freezing, and the like. In various aspects, the one or more protectant proteins can include one or more of: an RvLEAM1 protein or fragment thereof, a dAFP-1 protein or fragment thereof, or an AavLEA1 protein or fragment thereof.

[0036] The RvLEAM1 protein or fragment thereof can be a truncated RvLEAM1 protein. In such aspects, the truncated RvLEAM1 protein comprises or consists of an amino acid sequence at least 85%, 90%, 95%, 99%,or 100% identical to SEQ ID NO: 2.

[0037] The dAFP-1 protein or fragment thereof can be a truncated dAFP-1 protein. In such aspects, the truncated dAFP-1 protein comprises or consists of an amino acid sequence at least 85%, 90%, 95%, 99%, or 100% identical to SEQ ID NO: 3.

[0038] The AavLEA1 protein or fragment thereof comprises or consists of an amino acid sequence at least 85%, 90%, 95%, 99%, or 100% identical to SEQ ID NO: 1, in one or more aspects.

[0039] In various aspects, the one or more protectant proteins can include one or more purification tags for facilitating purification. The purification tags can include any convenient purification tag that is suitable for use in the compositions and methods disclosed herein. In

various aspects, the purification tag can include a 6xHis tag, a FLAG tag, or both. In certain aspects, the one or more protectant proteins can include an N-terminal 6xHis tag and comprise or consist of an amino acid sequence that is at least 85%, 90%, 95%, 99%, or 100% identical to SEQ ID NOs: 4, 5, or 6. SEQ ID NO: 4 includes the AavLEA1 amino acid sequence of SEQ ID NO: 1 but also includes an N-terminal 6xHis-tag; SEQ ID NO: 5 includes the truncated RvLEAM1 amino acid sequence of SEQ ID NO: 2 but also includes an N-terminal 6xHis-tag; and SEQ ID NO: 6 includes the truncated dAFP-1 amino acid sequence of SEQ ID NO: 3 but also includes an N-terminal 6xHis-tag. Table 1 below lists the sequences of SEQ ID NOs: 1-6. Table 2 below lists a polynucleotide sequence encoding AavLEA1. Those of skill in the art will understand that the polypeptides provided as SEQ ID Nos: 1-3 could be generated as longer sequences and cleaved to generate the sequences described and used herein or that the polypeptides may contain other additional amino acids to aid in purification (like the 6-His tag), identification or tracking or aid in synthesis of the polypeptide (addition of a methionine or other amino acids that may aid transcription and/or translation of the DNA or RNA encoding the polypeptide to aid in production). In some embodiments to polypeptide may comprise SEQ ID NO: 1-3 and include at most an additional 30 amino acids, 25 amino acids, 20 amino acids, 15 amino acids, 10 amino acids, 5 amino acids or any extra amino acids therebetween.

[0040] Table 1: Polypeptide Sequences for Example Protectant Proteins

SEQ ID NO:	**Polypeptide Sequence**
1	LVPRGSHMSSQQNQNRQGEQQEQGYMEAAKEKVVNAWESTKETLS STAQAAAEKTAEFRDSAGETIRDLTGQAQEKGQEFKERAGEKAEETK QRAGEKMDETKQRAGEMRENAGQKMEEYKQQGKGKAEELRDTAA EKLHQAGEKVKGRD
2	LEVLFQGPGKDLKNEASWKAKGVANQAAGAFERAKDTVKEGVHD MKRSGSRVFEQGQEEVEAGAQHAKAGYQSAKNAAQDTAATLKDKA GSAWNQAKHVVEDKGEDVVEAVKDTASKVWGKAKHVAEDVKENA
3	LEVLFQGPGGQCTGGSDCRSCTVSCTDCQNCPNARTACTRSSNCINA LTCTDSYDCHNAETCTRSTNCYKAKTCTGSTNCYEATACTDSTGCP
4	MGSSHHHHHHSSGLVPRGSHMSSQQNQNRQGEQQEQGYMEAAKEK VVNAWESTKETLSSTAQAAAEKTAEFRDSAGETIRDLTGQAQEKGQE

	FKERAGEKAEETKQRAGEKMDETKQRAGEMRENAGQKMEEYKQQ GKGKAEELRDTAAEKLHQAGEKVKGRD
5	MGSSHHHHHHSSGLEVLFQGPGKDLKNEASWKAKGVANQAAGAFE RAKDTVKEGVHDMKRSGSRVFEQGQEEVEAGAQHAKAGYQSAKNA AQDTAATLKDKAGSAWNQAKHVVEDKGEDVVEAVKDTASKVWGK AKHVAEDVKENA
6	MGSSHHHHHHSSGLEVLFQGPGGQCTGGSDCRSCTVSCTDCQNCPN ARTACTRSSNCINALTCTDSYDCHNAETCTRSTNCYKAKTCTGSTNC YEATACTDSTGCP

[0041] Table 2: Polynucleotide Sequence for AavLEA1

Polynucleotide Sequence for AavLEA1 (SEQ ID NO: 7)
atgggcagcagccatcatcatcatcacagcagcggcctggtgccgcgcggcagccatatgtcctctcagcagaaccagaaccgac agggtgagcagcaggagcagggctacatggaggcggccaaggagaaggtcgtcaacgcatgggagagcacgaaggaaaccctctc gagcacggctcaagcggccgccgagaagacggctgagtttcgcgattccgccggtgagaccatccgtgacctgaccggacaggcgca ggagaagggtcaggagttcaaggagcgcgctggcgagaaggcagaggagacgaagcagcgtgccggggagaagatggatgagac caagcagcgggctggcgaaatgcgcgagaacgcgggccagaagatggaggagtacaagcagcagggcaagggcaaggccgagg agcttcgcgacactgccgccgagaagctccaccaggctggcgagaaggtcaagggccgcgactaa

[0042] In various aspects, the compositions can alternatively or additionally include one or more of the polypeptides disclosed in Table 1, and/or have a sequence that is at least 85 %, 90 %, 95 %, or 99 % identical to one or more of the sequences in Table 1.

[0043] In various aspects, the one or more protectant proteins can have a mass of 20 kiloDaltons (kDa) or less, 18 kDa or less, 17 kDa or less, or 15 kDa or less.

[0044] In various aspects, the compositions disclosed herein can include multiple distinct compositions where each composition includes a distinct protectant protein. For instance, in certain aspects, a first composition may include a first protectant protein, while a second composition may include a second protectant protein, where the first protectant protein and the second protein are different from one another. In further aspects, a third composition may include a third protectant protein that is different than the first and second protectant proteins. In

alternative aspects, the compositions can include a combination of two or more of the protectant proteins described herein.

[0045] In certain aspects, the compositions disclosed herein can include an aqueous buffer. The aqueous buffer can be any type of buffer suitable for use in the compositions and methods disclosed herein. In the same or alternative aspects, the compositions disclosed herein can include additional components, chelating agents, preservatives, and the like. In one example aspect, the aqueous buffer can include HEPES-Na pH 7.5, 150 mM NaCl, and/or 1 mM TCEP/DTT.

[0046] As discussed above, certain aspects of the present disclosure also relate to constructs. In various aspects, the one or more constructs can include a promotor operably connected to a polynucleotide encoding one or more protectant proteins. In one or more aspects, the one or more constructs can include a promotor operably connected to a polynucleotide encoding a polypeptide having at least 85%, 90%, 95%, 99% or 100% identity to one or more of SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3.

[0047] In certain aspects, the polynucleotide sequence may further encode for a purification tag to allow for purification of the polypeptide. In various aspects, the purification tag can encode a 6xHis tag, a FLAG tag, or both. In certain aspects, the polynucleotide can encode for an N-terminal 6xHis tag and can encode for a polypeptide that is at least 85%, 90%, 95%, 99%, or 100% identical to one or more of SEQ ID NOs: 4, 5, or 6.

[0048] In one aspect, the polynucleotide can include a polynucleotide sequence that is at least 85%, 90%, 95%, 99%, or 100% identical to SEQ ID NO: 7.

[0049] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity over a specified region, e.g., of an entire nucleic acid or polypeptide sequence or individual portions or domains of a nucleic acid or polypeptide), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence, in the context of nucleic acids. By way of example, in aspects, the identity exists over a region that is about or at least about 5, 10, 15, 20, 50, or 100

amino acids in length, to about, less than about, or at least about 80, 85, 90, or 100 amino acids in length.

[0050] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0051] An example of algorithms suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., *Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul et al., J. Mol. Biol. 215:403-410 (1990), respectively. As will be appreciated by one of skill in the art, the software for performing BLAST analyses is publicly available through the website of the National Center for Biotechnology Information (NCBI). In embodiments, BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the nucleic acids and proteins. In embodiments, a BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. In embodiments, T is referred to as the neighborhood word score threshold (Altschul et al., supra). In embodiments, these initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. In embodiments, the word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. In embodiments, cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). In embodiments, for amino acid sequences, a scoring matrix is used to calculate the cumulative score. In embodiments, extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. In embodiments, the BLAST algorithm parameters W, T, and X determine the sensitivity and speed

of the alignment. In embodiments, the NCBI BLASTN or BLASTP program is used to align sequences. In embodiments, the BLASTN or BLASTP program uses the defaults used by the NCBI. In embodiments, the BLASTN program (for nucleotide sequences) uses as defaults: a word size (W) of 28; an expectation threshold (E) of 10; max matches in a query range set to 0; match/mismatch scores of 1, −2; linear gap costs; the filter for low complexity regions used; and mask for lookup table only used. In embodiments, the BLASTP program (for amino acid sequences) uses as defaults: a word size (W) of 3; an expectation threshold (E) of 10; max matches in a query range set to 0; the BLOSUM62 matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1992)); gap costs of existence: 11 and extension: 1; and conditional compositional score matrix adjustment.

[0052] One of skill will recognize that individual substitutions to a peptide, polypeptide, or protein sequence which alters a single amino acid is a conservatively modified variant where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles.

[0053] The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, Proteins (1984)).

[0054] In accordance with aspects of the present disclosure, kits for preparing cryo-EM samples are also described. In various aspects, the kits can comprise one or more of the compositions and/or or constructs described herein. Such kits may comprise one or more components, which may be contained in one or more containers such as boxes, cartons, tubes, vials, ampules, bags, and the like. In one example aspect, the kits may include multiple distinct compositions in separate containers. In such an example aspect, the kits may include a first container housing a first composition comprising a first protectant protein or construct encoding for the first protectant protein, a second container housing a second composition comprising a second protectant protein or a construct encoding for the second protectant protein, and a third

container housing a third composition comprising a third protectant protein or a construct encoding for the third protectant protein.

[0055] In certain aspects, the kits may further comprise one or more additional components, such as dilution buffers or reagents, reagents for expression and/or purification of the protectant proteins. In one or more aspects, the kits may also comprise a crosslinking agent. In one aspect, a non-limiting example of a crosslinking agent is glutaraldehyde. In various aspects, such kits may also comprise cryo-EM SPA grids and/or protocols or instructions for carrying out the methods of the invention.

[0056] Methods

[0057] As discussed above, methods are disclosed herein that relate to the preparation of a cryo-EM sample. In certain aspects, the methods can include adding one or more protectant proteins to a biological sample during preparation of the cryo-EM sample. The one or more protectant proteins can be any of the protectant proteins and/or compositions discussed herein.

[0058] In various aspects, the biological sample can be any type of biological sample suitable for use in cryo-EM and/or in cryo-EM SPA. In certain aspects, the biological sample comprises a biomolecule having a mass of 50 kDa or more, 100 kDa or more, 120 kDa or more, or 150 kDa or more. In the same or alternative aspects, the biological sample comprises a multi-subunit protein, or nucleic acid, or protein and nucleic acid complex.

[0059] The one or more protectant proteins can be added to the biological sample during preparation of the cryo-EM sample in any suitable manner. In various aspects, the one or more protectant proteins are present in an aqueous buffer and/or in liquid form when added to the biological sample. In the same or alternative aspects, the biological sample can be in liquid form when the one or more protectant proteins are added.

[0060] In various aspects, the molar ratio of the one or more protectant proteins relative to the biomolecule in the biological sample is at least 2:1, at least 4:1, at least 6:1, at least 10:1, at least 20:1, at least 30:1, at least 40:1, at least 50:1, at least 60:1, or at least 70:1; and/or 70:1 or less, 60:1 or less, 50:1 or less, 40:1 or less, 30:1 or less, or 20:1 or less; and/or no more than about 80 μM of protective proteins, no more than about 70 μM of protective proteins, no more than about 60 μM of protective proteins, or about 6 μM of protective proteins, or about 12 μM of protective proteins, or about 60 μM of protective proteins. In the same or alternative aspects, the molar ratio of the one or more protectant proteins relative to the biomolecule in the biological sample is

between 60:1 to 2:1, or between 40:1 to 6:1. In certain aspects, a suitable molar ratio of a protectant protein to the biomolecule in the biological sample may differ depending upon the specific biomolecule and/or specific protectant protein of interest.

[0061] As discussed above, the one or more protectant proteins are added to the biological sample during preparation of the cryo-EM sample. The preparation of the cryo-EM sample can include any suitable processes for preparing a biological sample for cryo-EM. In certain aspects, the preparation of the cryo-EM sample can include an initial preparation step, such as, adding the biological sample to a container and/or adding one or more reagents or buffers to the biological sample, which is performed prior to the freezing process.

[0062] In certain aspects, the biological sample can be exposed to a crosslinking agent prior to freezing. In one or more aspects, the biological sample, in the presence or absence of the protectant protein, can be exposed to a crosslinking agent prior to freezing. The crosslinking agent can be any suitable crosslinking agent, including but not limited to glutaraldehyde, in one aspect. In various aspects, the biological sample, in the presence or absence of the protectant protein, can be exposed to one or more crosslinking agents for at least one minute, at least two minutes, at least five minutes, or at least ten minutes.

[0063] In various aspects, the methods disclosed herein can also include freezing the cryo-EM sample to provide a vitreous cryo-EM sample. The cryo-EM sample can be frozen using any convenient cryo-EM freezing processes and/or systems. In certain aspects, the cryo-EM sample can be applied to a grid and frozen within the grid.

[0064] In various aspects, the methods disclosed herein can also include performing electron microscopy on the vitreous cryo-EM sample to obtain cryo-EM data. Optionally, two-dimensional analysis, three-dimensional analysis, or both, can be performed on the cryo-EM data. In various aspects, any convenient electron microscopy system and/or data analysis processes can be used and chosen by one of skill in the art for a particular purpose.

[0065] In various aspects, it may be advantageous to perform cryo-EM on a biological sample using different protectant proteins, as different protectant proteins may differently affect the orientation and/or other property of the biomolecule in the vitreous cryo-EM sample, thereby allowing for multiple and different EM data sets on a biomolecule that may aid in resolving the structure. The use of different protectant proteins may provide several distinct pathways for biological sample AWI protection, especially given the complexity and unclear nature of AWI

damage to cryo-EM samples and would allow cryo-EM SPA users to screen and concoct a cocktail that provides optimal AWI protection for their sample of interest to achieve the highest possible cryo-EM map resolution. For example, after forming a protective barrier at the AWI, each protectant protein may present a specific interaction surface towards the protected sample, resulting in a preferred orientation problem. A mixture of protectant proteins may alleviate this problem.

[0066] In various aspects, a method can include adding a first protectant protein to a first portion of a biological sample to provide a first cryo-EM sample. Further in such aspects, a second protectant protein can be added to a second portion of the biological sample to provide a second cryo-EM sample, where the second protectant protein is different than the first protectant protein in the first cryo-EM sample. In various aspects, the first and second cryo-EM sample can be frozen and electron microscopy can be performed on each vitreous cryo-EM sample separately. Optionally, two-dimensional analysis, three-dimensional analysis, or both can be performed on the cryo-EM data for the first and/or second vitreous cryo-EM sample. It should be understood that the example method can also and/or optionally include adding additional protectant proteins to other portions of the biological sample, e.g., a third protectant protein being added to a third portion of the biological sample to obtain additional cryo-EM data, and so on. In an alternative aspect, multiple protectant proteins can be added to a single biological sample, and/or different combinations of protectant proteins can be added to different portions of the biological sample.

[0067] Specific aspects of the present disclosure as illustrated in the following Examples are for illustrative purposes and are not limited to the materials, conditions, or process parameters set forth in these aspects.

Examples

Example 1: AavLEA1 protein

[0068] Example 1A: Tolerance level of AavLEA1 addition in a cryo-EM sample in terms of electron microscope imaging and sample grid vitrification quality

[0069] One of the key criteria for a successful high-resolution cryo-EM SPA project is the visibility of individual sample particles in the EM images. The particle visibility directly impacts the process of particle picking and subsequent particle image alignment steps for accurate analysis and reconstruction. Adding AavLEA1 to the sample buffer will inevitably increase the background

noise. Thus, the amount of signal contributed by AavLEA1 and how it scales with increasing AavLEA1 will be determined.

[0070] First, purified recombinant AavLEA1 of high purity and quantity will obtained. A bacterial expression plasmid of AavLEA1 with an N-terminal 6xHIS-tag is readily available from the public plasmid repository, Addgene. The protein will be expressed using an established protocol that is available. Beyond the published protocol, a homogenous population of AavLEA1 proteins will be collected by running the protein through a size-exclusion chromatography (SEC) column. This step removes any protein aggregations and contaminations and characterizes the quality of the purified protein. Maximum storage concentration of the purified AavLEA1 protein to achieve optimal storage conditions will be determined. These optimization assays will be determined in several biochemical buffers (for example, low vs. high salt conditions) to evaluate AavLEA1 compatibility in a wide range of buffer conditions.

[0071] Next, the quality of thin vitrified ice embedding purified apoferritin proteins in the presence of various concentrations of AavLEA1 will be tested. The mouse apoferritin is a gold-standard test protein specimen for cryo-EM SPA high-resolution reconstruction. Most importantly, apoferritin is not affected by AWI damage, thus allowing for a determination of the impact of AavLEA1 on a high-resolution cryo-EM SPA map reconstruction, independent of any alteration to a sample AWI damage. The quality of vitrified ice in the presence of AavLEA1 will be assessed by measuring the apoferritin particle contrast, beam-induced motion, and beam-induced protein damage. These analyses will then be compared across cryo-EM datasets (~500 cryo-EM movies) collected with the sample molar ratio of AavLEA1:apoferritin at 1:10, 1:1, and 10:1. Finally, the high-resolution SPA maps of apoferritin derived from these datasets to identify potential structural changes will be identified.

[0072] As buffer additives may change the viscosity of the aqueous sample and impact thin-film formation during sample grid preparation, the ice thickness of apoferritin cryo-EM grids in varying concentrations of AavLEA1 will be determined. The ice thickness of these cryo-EM samples are analyzed using the aperture limited scattering method. The consistency of aqueous spread is quantified by measuring the number of usable mesh squares per grid for productive cryo-EM imaging. These analyses will identify a set of parameters to establish a reproducible cryo-EM sample grid preparation protocol using AavLEA1.

[0073] Example 1B: Determine if AavLEA1 can improve particle orientation distribution of the influenza virus hemagglutinin trimer for high-resolution cryo-EM SPA reconstruction.

[0074] The influenza virus hemagglutinin (HA) is a homotrimeric glycoprotein that has been routinely used as a cryo-EM SPA test specimen to demonstrate the problem of preferred particle orientation. Failure to obtain adequate particle orientations of a sample leads to a highly anisotropic low-resolution cryo-EM map. Recent studies have revealed that the cause of the preferred orientation problem is the sample's interaction with AWI. Consequently, adding detergents or Spotiton fast vitrification, designed to reduce the sample interaction with AWI, can alleviate this problem.

[0075] If AavLEA1 confers AWI protection to samples during vitrification, improvements to the particle orientation distribution during cryo-EM SPA are expected. The HA trimer, which is commercially available, will be used as the test specimen to study the effects of AavLEA1 addition on particle orientation distribution. A standard cryo-EM SPA should yield only the top-down views of the protein, which can be assessed from its two-dimensional classification averages and angular distribution map after three-dimensional model reconstruction. Similar analyses will be performed with AavLEA1 added at molar ratios of AavLEA1:HA trimer of 1:10, 1:1, and 10:1. The wide AavLEA1 titration range will characterize how the HA trimer particle orientation changes upon AavLEA1 increment. The minimum amount of AavLEA1 needed to achieve isotropic particle orientation distribution will then be determined from these analyses. Once a high-resolution cryo-EM map of HA trimer is obtained from the AavLEA1 dataset, its structure with that without AavLEA1 addition will be compared to determine any perturbations to HA trimer conformation in the presence of AavLEA1.

[0076] Example 1C: Establish if the stoichiometric addition of AavLEA1 can preserve the structural integrity of the fragile human CST complex during cryo-EM sample grid vitrification.

[0077] The human CTC1-STN1-TEN1 (CST) is a heterotrimer protein complex that requires detergent to see any discernable particles in vitrified cryo-EM SPA grids. Without detergents forming a protective barrier at the AWI, the CST complex will dissociate and denature upon adsorption to the AWI. Hence the CST complex represents an excellent opportunity to test if AavLEA1 can protect fragile complexes such as CST from AWI damages.

[0078] The recombinant human CST complex will be expressed and purified as per established protocols. Cryo-EM SPA grids will then be made with the molar ratios of AavLEA1:CST of 1:10,

1:1, and 10:1, in addition to that without AavLEA1. The number of discernable particles per cryo-EM image will be automatically picked and quantified by a template-picking algorithm (via either Relion or cryoSPARC software) using two-dimensional templates generated from the published cryo-EM complex map (EMD-21563). The analysis will determine how the number of discernable particles per image correlates with the molar ratio of AavLEA1:CST. Comparison will also be made with a CST dataset using the detergent, CHAPSO (3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate), as the additive. It is possible that if AavLEA1 prevents fragile samples from denaturation and recovers discernable particles, a new but manageable problem of preferred particle orientation may arise, akin to AavLEA1 reducing the interaction between CST and AWI. Thus, the above Example will also extend to two-dimensional image classification analysis and three-dimensional model reconstruction to determine how AavLEA1 affects the CST particle orientation distribution upon discernable particle recovery.

[0079] Like all buffer additives, it is beneficial to ascertain if AavLEA1 has any impact on CST functions. Functional changes will be tested by comparing CST's DNA-binding affinity and its interaction with DNA polymerase alphaprimase complex with/without AavLEA1. These experiments will be performed with AvLEA1 concentrations similar to those used in the cryo-EM experiments. The above describes how one can perform these controls using established assays pertaining to their samples. This is a beneficial advisory step before users jump into a comprehensive cryo-EM SPA pipeline using AavLEA1. However, if AavLEA1 affects their sample's structure or/and functions, alternative protectant proteins can be used.

Example 2: RvLEAM and DAFP-1 proteins

[0080] Ideally, the AavLEA1 protein (described in Example 1) would be a universal solution to all problematic and fragile cryo-EM SPA samples. However, as in all research methodology development, this proposed methodology will benefit from a diverse portfolio of stress-response, sacrificial proteins. Establishing a diverse class of sacrificial proteins would provide application flexibility for a broad spectrum of sample types, primarily when a particular sacrificial protein is found to interfere with the cryo-EM sample's structure and function. The identified sacrificial proteins may provide several distinct pathways for sample AWI protection, especially given the complexity and unclear nature of AWI damage to cryo-EM samples. Having a variety of small sacrificial proteins would allow cryo-EM SPA users to screen and concoct a cocktail that provides

optimal AWI protection for their sample of interest to achieve the highest possible cryo-EM map resolution. For example, after forming a protective barrier at the AWI, each sacrificial protein may present a specific interaction surface towards the protected sample, resulting in a preferred orientation problem. A mixture of sacrificial proteins would alleviate this problem. In addition, this new technology can be combined with existing methods to tackle the most challenging cryo-EM SPA samples. In other words, this proposed methodology represents not an incremental advancement over current technology but an entirely novel approach that users will primarily use for cryo-EM structural biology research.

[0081] Example 2A: Using 3D confocal fluorescent microscopy to characterize the air-water interface affinity of RvLEAM and DAFP-1.

[0082] The RvLEAM and DAFP-1 proteins will be expressed and purified in the standard bacterial system, similar to the approach described for AavLEA1 in Example 1. The recombinant proteins will be assessed for purity and quality using protein gel analysis and size-exclusion chromatography. Their ability to interact with AWI will be analyzed by 3D confocal fluorescence microscopy (using fluorescently-labeled protein candidates), their coverage at AWI and the associated dynamics can be determined, as previously demonstrated (FIG. 2A). 3D confocal fluorescence microscopy has a Z-axis resolution of ~ 500 nm (depending on objective) and thus sufficient to visualize a layer of fluorescence molecules that has accumulated at the AWI - a full Z-scan will be about 150-200 μm, which gives plenty of spatial resolution to discern a fluorescence layer that has formed at the AWI. Microscope imaging experiments will be performed at high throughput by using multiwell microplates. To fluorescently label small protein candidates, a short eight amino acids peptide sequence, ybbR-tag, will be added to their N-terminus, which would allow for covalently conjugating a wide range of commercially available fluorophores to the proteins. The small size of the ybbR tag will minimize any potential perturbation to the proteins' AWI interaction properties. The proteins' coverage at the AWI will be quantified using the 3D Z-stack confocal fluorescence imaging method and determine the minimum protein concentration needed to achieve complete AWI coverage (FIG. 2B). AavLEA1, RvLEAM, and DAFP-1 will be tested, and their results will be compared. Next, their stability at AWI will be characterized by fluorescence recovery after photobleaching (FRAP) experiment (FIG. 2B). If the proteins are stably bound to the AWI, there will be no (or slow) recovery of fluorescent signals upon

photobleaching. Their recovery dynamics will then be quantified and compared between the protein candidates.

[0083] Example 2B: Determine if a combination of protectant proteins enhances coverage of their air-water interface protection.

[0084] The confocal fluorescence microscopy approach is extended to multicolor fluorescence imaging to simultaneously study multiple proteins' collective AWI protection effect (FIG. 2B). This will inform how the proteins compete for AWI coverage. A collection of combinations will be identified that give optimal AWI coverage within the shortest time.

[0085] Next, using the same technique, the proteins' AWI sample protection function will be directly tested by studying the assembly of fluorescently labeled sample proteins at the AWI in the presence of the protectant proteins. For a test protein sample that adsorbs to the AWI, the mouse apoferritin protein complex (~ 480 kDa) that is fluorescently labeled (via the ybbR tag will be used). Apoferritin coverage at the AWI in the presence of varying concentrations of sacrificial proteins will be investigated. If there is AWI protection, the apoferritin fluorescent signal should remain in the solution and not at the AWI. Time-lapse experiments will be performed to determine if the proteins can prevent apoferritin from exchanging at the AWI. This will determine how long the fragile samples can remain protected during the entire cryo-EM SPA sample preparation process. Also, these experiments do not require the proteins to be fluorescently labeled, which can then serve as an additional control for the influence of ybbR tagging on the proteins' AWI affinity.

[0086] To validate the results from the in-solution fluorescent microscopy analysis, the proteins' AWI sample protection performance in cryo-EM SPA sample grid preparation will be tested. The cross-sectional particle distribution of apoferritin in the presence of the proteins in thin vitrified ice will be analyzed using cryo-electron tomography (cryo-ET). The 480 kDa mass of apoferritin provides a sufficient signal to visualize individual apoferritin particles in a cryo-ET experiment. In the absence of any additives, they adsorb to the AWI without structural damage.

[0087] In the presence of the sacrificial proteins, most, if not all, of the apoferritin particles are expected to be in the middle of the vitrified ice instead of at the AWI, which will be the case for the control.

[0088] References

[0089] 1. Wu M, Lander GC. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. Curr Opin Struct Biol. 2020;64:9-16. Epub 2020/07/01. doi: 10.1016/j.sbi.2020.05.007. PubMed PMID: 32599507; PMCID: PMC7666008.

[0090] 2. Henderson R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys. 1995;28(2):171-93. Epub 1995/05/01. doi: 10.1017/s003358350000305x. PubMed PMID: 7568675.

[0091] 3. Han BG, Watson Z, Kang H, Pulk A, Downing KH, Cate J, Glaeser RM. Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. J Struct Biol. 2016;195(2):238-44. Epub 2016/06/21. doi: 10.1016/j.jsb.2016.06.009. PubMed PMID: 27320699; PMCID: PMC4943657.

[0092] 4. Klebl DP, Gravett MSC, Kontziampasis D, Wright DJ, Bon RS, Monteiro DCF, Trebbin M, Sobott F, White HD, Darrow MC, Thompson RF, Muench SP. Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. Structure. 2020;28(11):1238-48 e4. Epub 2020/08/20. doi: 10.1016/j.str.2020.07.018. PubMed PMID: 32814033; PMCID: PMC7652391.

[0093] 5. Glaeser RM. Proteins, Interfaces, and Cryo-Em Grids. Curr Opin Colloid Interface Sci. 2018;34:1-8. Epub 2018/06/06. doi: 10.1016/j.cocis.2017.12.009. PubMed PMID: 29867291; PMCID: PMC5983355.

[0094] 6. Noble AJ, Wei H, Dandey VP, Zhang Z, Tan YZ, Potter CS, Carragher B. Reducing effects of particle adsorption to the air-water interface in cryo-EM. Nat Methods. 2018;15(10):793-5. Epub 2018/09/27. doi: 10.1038/s41592-018-0139-3. PubMed PMID: 30250056; PMCID: PMC6168394.

[0095] 7. Chen J, Noble AJ, Kang JY, Darst SA. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. J Struct Biol X. 2019;1. Epub 2019/01/01. doi: 10.1016/j.yjsbx.2019.100005. PubMed PMID: 32285040; PMCID: PMC7153306.

[0096] 8. Dandey VP, Wei H, Zhang Z, Tan YZ, Acharya P, Eng ET, Rice WJ, Kahn PA, Potter CS, Carragher B. Spotiton: New features and applications. J Struct Biol. 2018;202(2):161-9. Epub 2018/01/26. doi: 10.1016/j.jsb.2018.01.002. PubMed PMID: 29366716; PMCID: PMC6317895.

[0097] 9. Pantelic RS, Meyer JC, Kaiser U, Baumeister W, Plitzko JM. Graphene oxide: a substrate for optimizing preparations of frozen-hydrated samples. J Struct Biol. 2010;170(1):152-6. Epub 2009/12/29. doi: 10.1016/j.jsb.2009.12.020. PubMed PMID: 20035878.

[0098] 10. Doerr A. Graphene-on-gold grids for cryo-EM. Nat Methods. 2019;16(7):578. Epub 2019/06/30. doi: 10.1038/s41592-019-0488-6. PubMed PMID: 31249417.

[0099] 11. Passmore LA, Russo CJ. Specimen Preparation for High-Resolution Cryo-EM. Methods Enzymol. 2016;579:51-86. Epub 2016/08/31. doi: 10.1016/bs.mie.2016.04.011. PubMed PMID: 27572723; PMCID: PMC5140023.

[0100] 12. Glaeser RM, Hall RJ. Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects. Biophys J. 2011;100(10):2331-7. Epub 2011/05/18. doi: 10.1016/j.bpj.2011.04.018. PubMed PMID: 21575566; PMCID: PMC3093552.

[0101] 13. Goyal K, Tisi L, Basran A, Browne J, Burnell A, Zurdo J, Tunnacliffe A. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. J Biol Chem. 2003;278(15):12977-84. Epub 2003/02/06. doi: 10.1074/jbc.M212007200. PubMed PMID: 12569097.

[0102] 14. Browne J, Tunnacliffe A, Burnell A. Anhydrobiosis: plant desiccation gene found in a nematode. Nature. 2002;416(6876):38. Epub 2002/03/08. doi: 10.1038/416038a. PubMed PMID: 11882885.

[0103] 15. Yuen F, Watson M, Barker R, Grillo I, Heenan RK, Tunnacliffe A, Routh AF. Preferential adsorption to air-water interfaces: a novel cryoprotective mechanism for LEA proteins. Biochem J. 2019;476(7):1121-35. Epub 2019/03/23. doi: 10.1042/BCJ20180901. PubMed PMID: 30898848; PMCID: PMC6458962.

[0104] 16. Kampjut D, Steiner J, Sazanov LA. Cryo-EM grid optimization for membrane proteins. iScience. 2021;24(3):102139. Epub 2021/03/06. doi: 10.1016/j.isci.2021.102139. PubMed PMID: 33665558; PMCID: PMC7900225.

[0105] 17. Sigworth FJ. Principles of cryo-EM single-particle image processing. Microscopy (Oxf). 2016;65(1):57-67. Epub 2015/12/26. doi: 10.1093/jmicro/dfv370. PubMed PMID: 26705325; PMCID: PMC4749045.

[0106] 18. Wu M, Lander GC, Herzik MA, Jr. Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200keV. J Struct Biol X. 2020;4:100020. Epub

2020/07/11. doi: 10.1016/j.yjsbx.2020.100020. PubMed PMID: 32647824; PMCID: PMC7337053.

[0107] 19. Rice WJ, Cheng A, Noble AJ, Eng ET, Kim LY, Carragher B, Potter CS. Routine determination of ice thickness for cryo-EM grids. J Struct Biol. 2018;204(1):38-44. Epub 2018/07/08. doi: 10.1016/j.jsb.2018.06.007. PubMed PMID: 29981485; PMCID: PMC6119488.

[0108] 20. Tan YZ, Baldwin PR, Davis JH, Williamson JR, Potter CS, Carragher B, Lyumkis D. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nat Methods. 2017;14(8):793-6. Epub 2017/07/04. doi: 10.1038/nmeth.4347. PubMed PMID: 28671674; PMCID: PMC5533649.

[0109] 21. Lim C, Barbour AT, Zaug AJ, Goodrich KJ, Mckay AE, Wuttke DS, Cech TR. The structure of human CST reveals a decameric assembly bound to telomeric DNA. Science. 2020;368(6495):1081-5. doi: 10.1126/science.aaz9649. PubMed PMID: WOS:000539738400038.

[0110] 22. Scheres SH. Semi-automated selection of cryo-EM particles in RELION-1.3. J Struct Biol. 2015;189(2):114-22. Epub 2014/12/09. doi: 10.1016/j.jsb.2014.11.010. PubMed PMID: 25486611; PMCID: PMC4318617.

[0111] 23. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14(3):290-6. Epub 2017/02/07. doi: 10.1038/nmeth.4169. PubMed PMID: 28165473.

[0112] 24. Boothby TC, Tapia H, Brozena AH, Piszkiewicz S, Smith AE, Giovannini I, Rebecchi L, Pielak GJ, Koshland D, Goldstein B. Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. Mol Cell. 2017;65(6):975-84 e5. Epub 2017/03/18. doi: 10.1016/j.molcel.2017.02.018. PubMed PMID: 28306513; PMCID: PMC5987194.

[0113] 25. Yamaguchi A, Tanaka S, Yamaguchi S, Kuwahara H, Takamura C, Imajoh-Ohmi S, Horikawa DD, Toyoda A, Katayama T, Arakawa K, Fujiyama A, Kubo T, Kunieda T. Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. PLoS One. 2012;7(8):e44209. Epub 2012/09/01. doi: 10.1371/journal.pone.0044209. PubMed PMID: 22937162; PMCID: PMC3429414.

[0114] 26. Hand SC, Menze MA, Toner M, Boswell L, Moore D. LEA proteins during water stress: not just for plants anymore. Annu Rev Physiol. 2011;73:115-34. Epub 2010/11/03. doi: 10.1146/annurev-physiol-012110-142203. PubMed PMID: 21034219.

[0115] 27. Duman JG, Li N, Verleye D, Goetz FW, Wu DW, Andorfer CA, Benjamin T, Parmelee DC. Molecular characterization and sequencing of antifreeze proteins from larvae of the beetle Dendroides canadensis. J Comp Physiol B. 1998;168(3):225-32. Epub 1998/05/20. doi: 10.1007/s003600050140. PubMed PMID: 9591363.

[0116] 28. Meister K, Ebbinghaus S, Xu Y, Duman JG, DeVries A, Gruebele M, Leitner DM, Havenith M. Long-range protein-water dynamics in hyperactive insect antifreeze proteins. Proc Natl Acad Sci U S A. 2013;110(5):1617-22. Epub 2013/01/02. doi: 10.1073/pnas.1214911110. PubMed PMID: 23277543; PMCID: PMC3562781.

[0117] 29. Tanaka S, Tanaka J, Miwa Y, Horikawa DD, Katayama T, Arakawa K, Toyoda A, Kubo T, Kunieda T. Novel mitochondria-targeted heat-soluble proteins identified in the anhydrobiotic Tardigrade improve osmotic tolerance of human cells. PLoS One. 2015;10(2):e0118272. Epub 2015/02/13. doi: 10.1371/journal.pone.0118272. PubMed PMID: 25675104; PMCID: PMC4326354.

[0118] 30. Liao Z, Lampe JW, Ayyaswamy PS, Eckmann DM, Dmochowski IJ. Protein assembly at the air-water interface studied by fluorescence microscopy. Langmuir. 2011;27(21):12775-81. Epub 2011/09/29. doi: 10.1021/la203053g. PubMed PMID: 21942221; PMCID: PMC3212854.

[0119] 31. Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc Natl Acad Sci U S A. 2005;102(44):15815-20. Epub 2005/10/21. doi: 10.1073/pnas.0507705102. PubMed PMID: 16236721; PMCID: PMC1276090.

[0120] Example 3: Small LEA proteins as an effective air-water interface protectant for fragile samples during cryo-EM grid plunge freezing

[0121] Sample loss due to air-water interface (AWI) interactions during cryo-electron microscopy (cryo-EM) sample grid plunge freezing is a major bottleneck for structural biology research. As reported herein, desiccation stress-response Late Embryogenesis Abundant (LEA) proteins from nematodes and tardigrades, used as sample additives, can protect fragile multisubunit molecular samples from complete loss due to AWI damage during plunge freezing. High-resolution cryo-EM maps achieved using LEA protein additives are comparable to, if not better than, those obtained using more sophisticated AWI protection methods. It was found that the protected samples using the LEA proteins exhibited a certain degree of particle preferred

orientation and the extent of bias is sample dependent. Cryogenic electron tomography analysis indicates that the protected sample particles are localized at an interface, which led to propose that LEA proteins protect fragile samples from AWI damage by forming a LEA protein barrier at the AWI. Interaction with the LEA-water interface (LWI) likely explains the observed particle preferred orientation. In conclusion, LEA proteins provide a simple, cost-effective, and rapidly adoptable approach for cryo-EM structural biologists to mitigate sample damage caused by AWI interaction during plunge freezing.

[0122] The 'resolution revolution' of cryogenic-electron microscopy (cryo-EM) marks a significant shift in the field of structural biology1. However, the continued growth of cryo-EM single particle analysis structural biology faces a critical problem at the step of sample grid preparation — sample damage [2–9]. The standard method for cryo-EM single-particle analysis sample grid preparation is plunge freezing, where small volumes of sample are applied onto a pretreated cryo-EM grid, blotted, and rapidly plunge-frozen into a cryogen such as liquid ethane (FIG. 3A). This process forms a very thin layer (~10-100 nm) [5,10] of sample-embedded vitreous ice for transmission electron microscopy imaging [11,12] (FIG. 3B). After the blotting and before the rapid freezing step, the sample exists as a thin aqueous film with a high surface-to-volume ratio. During this transitional phase (typically in seconds), the proteins in the aqueous solution collide with the air-water interface (AWI) thousands of times before the sample hits the cryogen for vitrification [5,9].

[0123] Adsorption to the AWI can destroy the protein's structural integrity, causing it to denature, or disintegrate [2,5,7,13] (FIG. 3C). In a scenario where the protein maintains its structure after AWI adsorption, the sample interaction with the AWI may be biased, leading to preferred sample orientations and an anisotropic 3D reconstruction of the sample [2,3,14,15]. To mitigate the AWI problem, researchers have developed a variety of solutions. These include the addition of mild detergents to the sample [16,17], the use of chemical crosslinking [18–21], and adsorbing or tethering samples to surfaces to prevent sample contact with the AWI [4,22–30]. More sophisticated techniques involve rapidly freezing the sample before the proteins can diffuse to the AWIs [30–35]. However, these methods are often sample-specific, technically challenging, or cost-prohibitive.

[0124] AavLEA1 protects fragile cryo-EM samples from AWI damage during grid plunge freezing.

[0125] This example sought to develop a new AWI mitigation strategy that is highly accessible, cost-effective, and readily deployable in all cryo-EM facilities and research laboratories. To this end, the biological world was explored, hypothesizing that cellular response mechanisms to desiccation stress could offer an evolved solution to protein protection from damage caused by AWI interaction. Late Embryogenesis Abundant (LEA) proteins were identified [36–38], specifically a group 3 LEA protein, AavLEA1 from the roundworm (nematode) *Aphelenchus avenae* [39] to offer such a solution. Studies have indicated that AavLEA1, which has a molecular weight of ~16 kilodalton (kDa), prevents proteins from aggregation due to desiccation [39]. This function is due to AavLEA1 forming a protective barrier at the AWI, which prevents protein samples from damage caused by direct interaction with AWI [38]. It is proposed that AavLEA1 can be used as a sample additive to protect protein samples from damage caused by AWI interaction during cryo-EM grid plunge freezing (FIG. 3D). The relatively small molecular mass of AavLEA1 (approximately 16 kDa) compared to a typical sample of interest (several hundred kDa) should allow AavLEA1 to reach the AWI faster than the larger samples.

[0126] To evaluate the AWI protective effects of the LEA proteins, two protein complexes that are AWI-sensitive as model systems were used: human DNA polymerase-α-primase (PP)40–42 and Polycomb repressive complex-2 (PRC2) [20,43,44]. In this study, both complexes were tested in their apo-state, and all data collection was performed using a 200 keV electron microscope with a direct electron detector (For details, please see the Methods section below). Reported cryo-EM structures of apo-state PP complexes were solved using chemically crosslinked samples [42], suggesting that the PP particles fell apart due to AWI damage during plunge freezing. Another arguably more challenging structure to solve for cryo-EM single-particle analysis is that of PRC2. Reported PRC2 cryo-EM structures were solved by highly specialized strategies using either chemically crosslinked samples on thin carbon-support grids [20,45] or biotinylated samples tethered onto Streptavidin-crystal grids [43,44].

[0127] It was found that the addition of AavLEA1 to PP and PRC2 prior to freezing eliminate the need for the abovementioned complicated methods, allowing us to solve their structures at comparable or higher cryo-EM map resolutions (FIGS. 4A-4F). The optimal range of addition was found to be between 1:8-1:40 molar ratio of the sample to the LEA proteins (FIGS. 9A-9C). For example, the addition of AavLEA1 at a 1:40 PP:AavLEA1 molar ratio before plunge freezing led to a much more positive outcome than the PP sample grid without any additives; we saw

discernable and homogeneously sized particles (FIG. 4A). Multiple high-resolution two-dimensional (2D) class averages of PP were obtained from these particles (FIG. 4B). Subsequent image processing led to a 3.0 Å global resolution cryo-EM map of the apo-state PP (FIG. 4C). This is the highest resolution cryo-EM map reported for the human apo-state PP; the previous X-ray structure was solved at 3.6 Å [46] and a chemically crosslinked cryo-EM structured determined at a 3.8 Å global resolution [42]. Most importantly, we demonstrated that AavLEA1 can be used to mitigate sample damage from AWI interaction during plunge freezing, resulting in high-resolution cryo-EM maps from the protected particles.

[0128] Initial concerns with the AavLEA1 approach are 1) elevated background signal in the micrograph from the AavLEA1 addition and 2) AavLEA1 distorting the conformation of the sample of interest. Given that a 3.0 Å resolution cryo-EM map for human PP apo-state complex was obtained in the presence of excess AavLEA1 (sample:AavLEA1 of 1:40), it can be deduced that background AavLEA1 did not affect image alignment and hinder high-resolution cryo-EM map reconstruction. Extra map density was also not found that would indicate AavLEA1 binding to PP nor a significant change in the PP conformation; a R.M.S.D of 1.2 Å was calculated when a refined model derived from the 3.0 Å cryo-EM map was aligned to a published crystal structure of an apo-state human PP (PDB: 5EXR) [46] (FIG. 10). In short, AavLEA1 addition did not alter the PP apo-state conformation.

[0129] Similar results for PRC2 with the addition of AavLEA1 (1:40 PRC2:AavLEA1 molar ratio) were observed — Adding AavLEA1 resulted in discernable and homogeneously sized particles (FIG. 4D). Without additives, smaller sized particles in the micrographs were observed, indicating broken PRC2 complexes which is consistent as previously reported [47]. Subsequent cryo-EM image processing led to a 3.8 Å global resolution cryo-EM map of PRC2 (FIG. 4E and 4F). This result further exemplifies the efficacy of AavLEA1 in facilitating the determination of high-resolution cryo-EM structures of protein samples sensitive to the air-water interface.

[0130] A truncated form of LEA protein from tardigrade is also an effective AWI damage protectant.

[0131] Encouraged by the success of using nematode AavLEA1 as a sample AWI damage protectant, it was pondered whether LEA proteins from other organisms could provide similar results. Tardigrades (*Ramazzottius varieornatus*), also known as water bears, are extremely tolerant to desiccation stress [48] and have a group 3 LEA protein, RvLEAM [49]. Because

RvLEAM has nine LEA-like motif and thus about 30 kDa in mass, we truncated it to approximately 15 kDa to elevate its diffusion coefficient and match the size of AavLEA1 for ease of comparison (refer to the Methods section below for details). The truncated RvLEAM is hereafter termed $RvLEAM_{short}$.

[0132] Similar to AavLEA1, adding RvLEAMshort to PP or PRC2 sample at a molar ratio of 1:6 sample:RvLEAMshort before plunge freezing led to the appearance of monodisperse and discernable particles (FIGS. 5A-5F). From datasets collected over a single night, reconstructed cryo-EM maps of human apo-state PP and PRC2 at a reported global resolution of 4.5 and 3.7 Å respectively were successfully constructed (FIGS. 5C and 5F). Given that this is the second LEA protein tested that successfully rescued the cryo-EM structure determination of two challenging AWI-sensitive protein complexes, it is believed group 3 LEA proteins, as a cryo-EM single-particle analysis sample additive, offer a promising and powerful new method for structural biologists to mitigate sample damage caused by AWI during plunge freezing.

[0133] LEA protected samples are distributed at vitrified ice surfaces.

[0134] While the prediction that LEA proteins function as AWI damage protectants is correct, the model for the protection mechanism is not accurate. It was speculated that once the LEA proteins form a protective layer at the AWI, which were termed the LEA-water interface (LWI), the sample particles would remain in the aqueous layer. When imaged in vitreous ice after plunge freezing, it was expected the projected views of the frozen particles to be randomly distributed (FIGS. 3A-3D). However, a major underlying assumption was that there was no interaction between the sample and the LWI. If such interactions existed, they would manifest as a bias in the particle orientation (view projection) [6,17]. The particle orientation distribution analysis for the cryo-EM maps derived from the LEA protein datasets suggests this was the case (FIG. 6A). Calculations from the 3DFSC server [14] indicated that the PP maps show a good isotropic distribution, approaching perfect isotropy with a sphericity of 0.98 for those with AavLEA1 and 0.83 with $RvLEAM_{short}$. Conversely, the PRC2 maps exhibited poorer particle orientation distribution, regardless of the LEA protein used; sphericity values were 0.77 and 0.79 for maps with AavLEA1 and $RvLEAM_{short}$, respectively. Overall, these results suggest that the samples interact with the LWI to varying degrees, dependent on the specific sample.

[0135] If the samples are indeed interacting with the interface, the particles should be distributed as plane(s). To test this, fiducial-less cryo-electron tomography (cryo-ET) analysis was

utilized to visualize the particle distribution of PP with AavLEA1 (1:40 molar ratio) and PRC2 with $RvLEAM_{short}$ (1:6 molar ratio) in cryo-EM single-particle analysis grids (FIGS. 6A and 6B). In both cases, it was observed that most, if not all, of the particles were distributed within one or two planes in the vitrified ice within a hole of the cryo-EM grid, which confirms that the samples are adsorbed to a surface when LEA proteins are present.

[0136] Due to the small sizes of LEA proteins, we were unable to visualize them in our cryo-EM images, preventing us from determining their empirical coverage at the AWI. Consequently, the possibility of incomplete LEA protein coverage cannot be ruled out, and that the particle orientation bias could have resulted, in part or entirely, from interactions with exposed pockets of the AWI. If this is indeed the case, it is reassuring to note that these interactions with the AWI, in the presence of LEA proteins, did not significantly compromise the samples' structural integrity, as demonstrated by the reconstructed high-resolution cryo-EM maps of the tested samples (FIGS. 4A-4F, and FIGS. 5A-5F).

[0137] Biochemical strategies to improve particle orientation distribution of LEA datasets.

[0138] Given that interactions between the sample and LEA proteins at the LWI can contribute to the observed particle orientation bias, we aim to explore strategies that can influence these interactions to achieve a desirable particle orientation outcome. Both AavLEA1 and $RvLEAM_{short}$ are predicted to form amphiphilic alpha-helical structures, with the hydrophobic face oriented towards the air at the AWI and the hydrophilic side (negatively charged) facing the aqueous solution [49,50]. Thus, the negatively charged, solution-facing side of the LWI could create a "sticky" surface that leads to the particle orientation bias observed in the LEA datasets.

[0139] Past studies have demonstrated that electrostatic charge on a surface, whether AWI or carbon film, can influence sample particle orientation [51, 52]. Using PRC2 and $RvLEAM_{short}$ as the model condition, divalent cation, MgCl2, was added to the sample buffer to attempt to neutralize the negatively charged LWI surface, with the hope that this strategy will change the PRC2 particle orientation distribution. However, data was obtained suggesting that this approach did not improve the particle orientation distribution and instead made it worse with more limited views.

[0140] Next, the effect of mild glutaraldehyde crosslinking on PRC2 prior to freezing with $RvLEAM_{short}$ was investigated to determine if this could improve particle orientation distribution. The rationale is the glutaraldehyde reaction would neutralize the positively charged groups on the

surface of PRC2 and leads to particle orientation distribution changes, as has been demonstrated by past studies [18, 19, 52]. Chemical crosslinking also has the benefit of stabilizing the complex for improved structural homogeneity [18,19]. Cryo-EM datasets were collected of crosslinked PRC2 with $RvLEAM_{short}$ at a 1:6 sample:LEA molar ratio; the PRC2 samples was crosslinked with glutaraldehyde for 2 or 10 min (FIG. 11).

[0141] The analysis indicated that the chemical crosslinking improved the orientation distribution of PRC2 particles (FIG. 7A). The map sphericity increased from 0.79 in non-crosslinked samples to 0.83 after 2 minutes of crosslinking. Data was obtained showing that extending the crosslinking to 10 minutes further enhanced this value to 0.97. This incremental improvement suggests that longer crosslinking durations may be slightly more effective. The 10 min crosslinked sample dataset enabled us to achieve a 3.1 Å global resolution cryo-EM map of the PRC2 complex (FIG. 7B). This map represents the highest resolution yet reported for the human PRC2 apo-state complex. Given that the map was reconstructed from a dataset collected on a 200 kV microscope, using a 300 kV microscope could potentially yield even higher resolution. The PRC2 model, built using this cryo-EM map, has a structure similar to that obtained through the streptavidin-biotin affinity EM grid approach [25, 44], with the R.M.S.D. between the two aligned structures being 0.89 Å (FIG. 12).

[0142] A control 10 min crosslinked PRC2 cryo-EM dataset without $RvLEAM_{short}$ was also collected to determine the effect of adding LEA proteins crosslinked samples. While discernable and somewhat homogeneous particles were observed, subsequent single-particle analysis revealed that most particles were broken subcomplexes of PRC2, a result consistent with a previous study [47]. Nonetheless, a 4.3 Å cryo-EM map of an intact PRC2 complex was obtained, although the efficiency was modest; from 3,274 movies, approximately 1.7 million particles were picked, but only about ~3% contributed to the final cryo-EM map reconstruction. In contrast, with the addition of $RvLEAM_{short}$, approximately 14% of initially picked particles (~2.5 million particles) were utilized in the final reconstruction. Furthermore, particles from the crosslinked control dataset suffered from biased particle orientation, with a 3DFSC calculated sphericity value of 0.66. Data was obtained showing that the inclusion of $RvLEAM_{short}$ improved the sphericity value to 0.97. Overall, the above comparison suggests the addition of $RvLEAM_{short}$ enhanced the stability and particle orientation distribution of crosslinked PRC2 complexes.

[0143] Comparative analysis of LEA proteins and CHAPSO as AWI damage protectants.

[0144] CHAPSO, a zwitterionic detergent, can protect protein samples from AWI damage during plunge freezing [17, 53]. In at least certain scenarios, it has been an effective strategy for determining the high-resolution cryo-EM structures of PP-related complexes [40, 54]. Hence, LEA protein's efficiency in obtaining a high-resolution cryo-EM structure was compared against that of CHAPSO. To this end a 3.4 Å cryo-EM map of apo-state PP using CHAPSO was determined (FIG. 8A). By visual inspection, no discernable differences between the two cryo-EM maps or modeled structures was found (compared to the 40:1 AavLEA1:PP cryo-EM map, FIG. 13). However, three separate empirical analyses, Reslog55 (FIG. 8B), per-particle spectral signal-to-noise (ppSSNR) [56] (FIG. 8C), and Rosenthal-Henderson [57] (FIG. 14) plots, all pointed to the AavLEA1 dataset having the better particle image quality. It is possible that thicker ice in the CHAPSO dataset significantly impacted particle image quality, which may explain the observed differences. Nonetheless, these results demonstrate that AavLEA1, when used as a sample additive, can achieve cryo-EM map quality comparable to, or even better than, that obtained with detergents.

[0145] A major advantage of using AavLEA1 over CHAPSO, or other detergents, is its effectiveness when cryo-EM samples are limited or unstable at high concentrations. CHAPSO and similar detergents typically require high sample concentrations [51, 58] — over 4 mg/mL or approximately 13 μM for a 300 kDa protein sample — which can be challenging for precious samples. In contrast, with AavLEA1, only about 1-1.5 μM of sample protein was utilized, which is more than ten times less than what is required for CHAPSO or detergents in general. The application of LEA proteins can lead to an anisotropy distribution of particle orientation that is sample-dependent (see result section: LEA protected samples are distributed at vitrified ice surfaces and FIGS. 6B and 6C), and this issue does not arise with the use of CHAPSO. The particles in the CHAPSO dataset have an evenly distributed particle orientation. Consequently, the calculated sphericity value of the PP cryo-EM map from the CHAPSO dataset is 0.99.

[0146] Discussion

[0147] In summary, it was shown that two group 3 LEA proteins, AavLEA1 from nematodes and RvLEAM (truncated) from tardigrades, can confer sample protection from damage caused by sample interaction with AWI during cryo-EM sample grid plunge freezing. The two model multisubunit protein complexes samples, human PP and PRC2 apo-state complexes, are challenging targets for cryo-EM single-particle analysis structure determination. By simply adding

AavLEA1 or RvLEAMshort to the samples before plunge freezing, the cryo-EM structures of both PP and PRC2 were able to be determined at comparable, or better, resolution than those previously reported using more challenging and complicated anti-AWI damage strategies [20, 42–44, 47]. It is important to note that these fragile samples did not yield any discernable monodisperse particles using standard plunge freezing methods. Therefore, these results underscore the effectiveness of LEA proteins in revitalizing cryo-EM projects that would otherwise be deemed unfeasible.

[0148] The LEA proteins require lower sample concentration compared to other AWI mitigation solutions, such as Spotiton [31–33] or detergents [51, 58, 59], offering a significant advantage when working with limited samples or those prone to aggregation at high concentrations. Another common solution for accommodating low sample concentrations is the use of cryo-EM grids with support films. However, the quality of these specialized EM grids can vary due to production batch reproducibility, and they can be technically challenging to produce in research laboratories. Unlike these methods, the application of LEA proteins does not require specialized grids and is compatible with standard cryo-EM holey grids. While the utilization of LEA proteins can lead to preferred particle orientation problems, and the degree of anisotropy is sample-dependent, this issue is not insurmountable. It can be mitigated by employing chemical crosslinking [18] or tilted-stage data collection strategies [14].

[0149] It is believed that LEA proteins represent a promising avenue for structural biologists to revisit cryo-EM projects previously hindered by AWI issues, particularly those that have exhausted conventional AWI mitigation strategies. These proteins can be produced in large quantities using standard bacterial expression systems and purification schemes, providing a sustainable and cost-effective alternative to methods that rely on more expensive and perishable materials and reagents. Most importantly, the accessibility and economic benefits of LEA proteins enable any structural biology laboratory or cryo-EM facility to readily adopt this method, potentially revolutionizing their approach to cryo-EM structure determination.

[0150] References

[0151] 1. Kühlbrandt, W. The Resolution Revolution. Science (1979) 343, 1443–1444 (2014).

[0152] 2. Han, B.-G., Avila-Sakar, A., Remis, J. & Glaeser, R. M. Challenges in making ideal cryo-EM samples. Curr Opin Struct Biol 81, 102646 (2023).

[0153] 3. Xu, Y. & Dang, S. Recent Technical Advances in Sample Preparation for Single-Particle Cryo-EM. Front Mol Biosci 9, 892459 (2022).

[0154] 4. D'Imprima, E. et al. Protein denaturation at the air-water interface and how to prevent it. Elife 8, e42747 (2019).

[0155] 5. Noble, A. J. et al. Reducing effects of particle adsorption to the air–water interface in cryo-EM. Nat Methods 15, 793–795 (2018).

[0156] 6. Noble, A. J. et al. Routine single particle CryoEM sample and grid characterization by tomography. Elife 7, e34257 (2018).

[0157] 7. Glaeser, R. M. Proteins, interfaces, and cryo-EM grids. Curr Opin Colloid Interface Sci 34, 1–8 (2018).

[0158] 8. Glaeser, R. M. & Han, B.-G. Opinion: hazards faced by macromolecules when confined to thin aqueous films. Biophys Rep 3, 1–7 (2017).

[0159] 9. Taylor, K. A. & Glaeser, R. M. Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. J Struct Biol 163, 214–223 (2008).

[0160] 10. Kim, L. Y. et al. Benchmarking cryo-EM single particle analysis workflow. Front Mol Biosci 5, (2018).

[0161] 11. Adrian, M., Dubochet, J., Lepault, J. & McDowall, A. W. Cryo-electron microscopy of viruses. Nature 308, 32–36 (1984).

[0162] 12. Dubochet, J., Chang, J.-J., Freeman, R., Lepault, J. & McDowall, A. W. Frozen aqueous suspensions. Ultramicroscopy 10, 55–61 (1982).

[0163] 13. Glaeser, R. M. Preparing Better Samples for Cryo-Electron Microscopy: Biochemical Challenges Do Not End with Isolation and Purification. Annual Review of Biochemistry vol. 90 Preprint at https://doi.org/10.1146/annurev-biochem-072020-020231 (2021).

[0164] 14. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nat Methods 14, 793–796 (2017).

[0165] 15. Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. Nat Commun 8, 629 (2017).

[0166] 16. Liu, N. & Wang, H.-W. Better Cryo-EM Specimen Preparation: How to Deal with the Air–Water Interface? J Mol Biol 167926 (2022) doi:10.1016/j.jmb.2022.167926.

[0167] 17. Chen, J., Noble, A. J., Kang, J. Y. & Darst, S. A. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. J Struct Biol X 1, (2019).

[0168] 18. Stark, H. GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM. in Methods in Enzymology vol. 481 109–126 (Elsevier, 2010).

[0169] 19. Kastner, B. et al. GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods 5, 53–55 (2008).

[0170] 20. Kasinath, V. et al. Structures of human PRC2 with its cofactors AEBP2 and JARID2. Science (1979) 359, 940–944 (2018).

[0171] 21. Adamus, K., Le, S. N., Elmlund, H., Boudes, M. & Elmlund, D. AgarFix: Simple and accessible stabilization of challenging single-particle cryo-EM specimens through crosslinking in a matrix of agar. J Struct Biol 207, (2019).

[0172] 22. Wang, F. et al. General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. Proceedings of the National Academy of Sciences 117, 24269–24273 (2020).

[0173] 23. Wang, F. et al. Amino and PEG-amino graphene oxide grids enrich and protect samples for high-resolution single particle cryo-electron microscopy. J Struct Biol 209, 107437 (2020).

[0174] 24. Liu, N. et al. Bioactive Functionalized Monolayer Graphene for High-Resolution Cryo-Electron Microscopy. J Am Chem Soc 141, 4016–4025 (2019).

[0175] 25. Glaeser, R. M. & Han, B.-G. Streptavidin Affinity Grids for Single-particle Cryo-EM. Microscopy and Microanalysis 25, 990–991 (2019).

[0176] 26. Palovcak, E. et al. A simple and robust procedure for preparing graphene-oxide cryo-EM grids. J Struct Biol 204, 80–84 (2018).

[0177] 27. Lu, Y. et al. Functionalized graphene grids with various charges for single-particle cryo-EM. Nat Commun 13, 6718 (2022).

[0178] 28. Zheng, L. et al. Uniform thin ice on ultraflat graphene for high-resolution cryo-EM. Nat Methods 20, 123–130 (2023).

[0179] 29. Han, Y. et al. High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. Proceedings of the National Academy of Sciences 117, 1009–1014 (2020).

[0180] 30. Esfahani, B. G. et al. SPOT-RASTR—a cryo-EM specimen preparation technique that overcomes problems with preferred orientation and the air/water interface. Biophys J 123, (2024).

[0181] 31. Jain, T., Sheehan, P., Crum, J., Carragher, B. & Potter, C. S. Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. J Struct Biol 179, 68–75 (2012).

[0182] 32. Darrow, M. C., Moore, J. P., Walker, R. J., Doering, K. & King, R. S. Chameleon: Next Generation Sample Preparation for CryoEM based on Spotiton. Microscopy and Microanalysis 25, 994–995 (2019).

[0183] 33. Dandey, V. P. et al. Time-resolved cryo-EM using Spotiton. Nat Methods 17, 897–900 (2020).

[0184] 34. Rubinstein, J. L. et al. Shake-it-off: a simple ultrasonic cryo-EM specimen-preparation device. Acta Crystallogr D Struct Biol 75, 1063–1070 (2019).

[0185] 35. Tan, Y. Z. & Rubinstein, J. L. Through-grid wicking enables high-speed cryoEM specimen preparation. Acta Crystallogr D Struct Biol 76, 1092–1103 (2020).

[0186] 36. Hernández-Sánchez, I. E. et al. LEAfing through literature: late embryogenesis abundant proteins coming of age - achievements and perspectives. Journal of Experimental Botany vol. 73 Preprint at https://doi.org/10.1093/jxb/erac293 (2022).

[0187] 37. Hibshman, J. D. & Goldstein, B. LEA motifs promote desiccation tolerance in vivo. BMC Biol 19, 263 (2021).

[0188] 38. Hibshman, J. D., Clegg, J. S. & Goldstein, B. Mechanisms of Desiccation Tolerance: Themes and Variations in Brine Shrimp, Roundworms, and Tardigrades. Front Physiol 11, 592016 (2020).

[0189] 39. Goyal, K., Walton, L. J. & Tunnacliffe, A. LEA proteins prevent protein aggregation due to water stress. Biochemical Journal 388, (2005).

[0190] 40. He, Q. et al. Structures of human primosome elongation complexes. Nat Struct Mol Biol 30, 579–583 (2023).

[0191] 41. Yin, Z., Kilkenny, M. L., Ker, D. S. & Pellegrini, L. CryoEM insights into RNA primer synthesis by the human primosome. FEBS Journal 291, (2024).

[0192] 42. Kilkenny, M. L. et al. Structural basis for the interaction of SARS-CoV-2 virulence factor nsp1 with DNA polymerase α–primase. Protein Science 31, (2022).

[0193] 43. Kasinath, V. et al. JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. Science (1979) 371, eabc3393 (2021).

[0194] 44. Song, J. et al. Structural basis for inactivation of PRC2 by G-quadruplex RNA. Science (1979) 381, 1331–1337 (2023).

[0195] 45. Grau, D. et al. Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction. Nat Commun 12, 714 (2021).

[0196] 46. Baranovskiy, A. G. et al. Mechanism of Concerted RNA-DNA Primer Synthesis by the Human Primosome. Journal of Biological Chemistry 291, 10006–10020 (2016).

[0197] 47. Lipscomb, D. M. Electron microscopy methods to overcome the challenges of structural heterogeneity and preferred orientations in small (sub-500 kDa) macromolecular complexes. (2017).

[0198] 48. Møbjerg, N. et al. Survival in extreme environments - on the current knowledge of adaptations in tardigrades. Acta physiologica (Oxford, England) vol. 202 Preprint at https://doi.org/10.1111/j.1748-1716.2011.02252.x (2011).

[0199] 49. Tanaka, S. et al. Novel Mitochondria-Targeted Heat-Soluble Proteins Identified in the Anhydrobiotic Tardigrade Improve Osmotic Tolerance of Human Cells. PLoS One 10, e0118272 (2015).

[0200] 50. Li, D. & He, X. Desiccation Dependent Structure and Stability of an Anhydrobiotic Nematode Late Embryogenesis Abundant (LEA) Protein. in ASME 2009 Summer Bioengineering Conference, Parts A and B 213–214 (American Society of Mechanical Engineers, 2009). doi:10.1115/SBC2009-206862.

[0201] 51. Li, B., Zhu, D., Shi, H. & Zhang, X. Effect of charge on protein preferred orientation at the air–water interface in cryo-electron microscopy. J Struct Biol 213, 107783 (2021).

[0202] 52. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. Science (1979) 362, (2018).

[0203] 53. Chen, S., Li, J., Vinothkumar, K. R. & Henderson, R. Interaction of human erythrocyte catalase with air – water interface in cryoEM. Microscopy 71, i51–i59 (2022).

[0204] 54. He, Q. et al. Structures of the human CST-Polα–primase complex bound to telomere templates. Nature 608, 826–832 (2022).

[0205] 55. Stagg, S. M., Noble, A. J., Spilman, M. & Chapman, M. S. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. J Struct Biol 185, 418–426 (2014).

[0206] 56. Baldwin, P. R. & Lyumkis, D. Non-uniformity of projection distributions attenuates resolution in Cryo-EM. Prog Biophys Mol Biol 150, 160–183 (2020).

[0207] 57. Rosenthal, P. B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. J Mol Biol 333, 721–745 (2003).

[0208] 58. Snijder, J. et al. Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. J Struct Biol 198, 38–42 (2017).

[0209] 59. Kampjut, D., Steiner, J. & Sazanov, L. A. Cryo-EM grid optimization for membrane proteins. iScience 24, 102139 (2021).

[0210] 60. Tripathi, R., Benz, N., Culleton, B., Trouvé, P. & Férec, C. Biophysical Characterisation of Calumenin as a Charged F508del-CFTR Folding Modulator. PLoS One 9, e104970 (2014).

[0211] 61. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152, 36–51 (2005).

[0212] 62. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296 (2017).

[0213] 63. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221 (2010).

[0214] 64. Meng, E. C. et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Science 32, e4792 (2023).

[0215] 65. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. Nat Methods 17, (2020).

[0216] 66. Chan, L. M. et al. High-resolution single-particle imaging at 100-200 keV with the Gatan Alpine direct electron detector. Preprint at https://doi.org/10.1101/2024.02.14.580363 (2024).

[0217] 67. Baldwin, P. R. & Lyumkis, D. Tools for visualizing and analyzing Fourier space sampling in Cryo-EM. Prog Biophys Mol Biol 160, (2021).

[0218] 68. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. J Struct Biol 197, (2017).

[0219] 69. Grant, T., Rohou, A. & Grigorieff, N. CisTEM, user-friendly software for single-particle image processing. Elife 7, (2018).

[0220] 70. Zheng, S. et al. AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J Struct Biol X 6, (2022).

[0221] 71. Liu, Y. T. et al. Isotropic reconstruction for electron tomography with deep learning. Nat Commun 13, (2022).

[0222] 72. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. J Struct Biol 116, (1996).

[0223] 73. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nature Methods vol. 9 Preprint at https://doi.org/10.1038/nmeth.2089 (2012).

[0224] 74. Castaño-Díez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. J Struct Biol 178, (2012).

[0225] Methods for Example 3

[0226] Expression and purification of AavLEA1 and $RvLEAM_{short}$

[0227] The expression plasmid for HIS-tagged AavLEA1 was sourced from the Addgene plasmid repository [pET15b-AavLEA1, a gift from Claude Férec (Addgene plasmid # 53093)] [60]. The expression plasmid for HIS-tagged RvLEAMshort was constructed by inserting a truncated cDNA from pEThT-RvLEAM [pEThT-RvLEAM was a gift from Takekazu Kunieda (Addgene plasmid # 90033)] [49] into a pET15b vector. RvLEAMshort encodes residues 58-181 of RvLEAM (A0A0E4AVP3.1). Both recombinant AavLEA1 and RvLEAMshort were expressed in Escherichia coli BL21 (DE3) cells. A single bacterial colony containing the transformed plasmid was cultured overnight in 2 mL of Luria Bertani broth (LB) with 100 μg mL^{-1} carbenicillin at 37°C. This starter culture was then used to inoculate 1 L of LB supplemented with the same antibiotic. At an optical density (A600) of 0.6, gene expression was induced using 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 hours at 12°C with shaking at 230 rpm. HIS-

tagged AavLEA1 used in this Example is SEQ ID NO: 4 from Table 1. HIS-tagged RvLEAMshort used in this Example is SEQ ID NO: 5 from Table 1.

[0228] Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole, 1 mM DTT or TCEP, 1 mM PMSF), and lysed via sonication. The cell debris was then removed by centrifugation. The clarified lysate was incubated with pre-equilibrated nickel-NTA resin (Qiagen, Germany) and stirred for 1 hour at 4°C. The protein-bound resin was washed three times with 50 mL of lysis buffer. Proteins were eluted with 10 mL of elution buffer (wash buffer supplemented with 250 mM imidazole) using a gravity flow column. The proteins were then concentrated to ~500 µL using a 3 kDa MWCO spin column and further purified on a Superdex 75 10/300 size-exclusion chromatography (SEC) column (Cytiva, USA) pre-equilibrated with SEC buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP, 10% glycerol). Eluted fractions were analyzed by SDS-PAGE. Chosen fractions were pooled, concentrated, snap-frozen in 5-10 µL aliquots, and stored at −80°C until use. The protein concentration of the aliquots was determined using the Beer-Lambert equation, with absorbance measurements obtained from a NanoDrop spectrophotometer (Thermo Fisher, USA) and extinction coefficients calculated based on their protein sequences.

[0229] Production of recombinant human DNA Polymerase alpha-Primase and Polycomb Repressive Complex 2

[0230] Recombinant human DNA polymerase alpha-primase was expressed in insect cells using baculovirus infection and purified as previously reported54. Purified recombinant human polycomb repressive complex 2 (PRC2) protein complexes were generously provided by Dr. Tom Cech at the University of Colorado Boulder. The expression and purification protocols for PRC2 have been described previously [44].

[0231] PRC2 cryo-EM sample glutaraldehyde crosslinking

[0232] Approximately 2 µM of PRC2 was incubated with a 0.1% (v/v) final concentration of glutaraldehyde for 2- and 10-minute intervals. After each interval, an aliquot was removed and quenched with 80 mM Tris-HCl to stop the reaction. SDS-PAGE was then used to assess the crosslinking efficiency of the PRC2 samples at each incubation time point.

[0233] All samples were thawed just prior to cryo-EM grid preparation. Holey carbon cryo-EM grids, either Quantifoil R 1.2/1.3 300 mesh Au or C-flat R 1.2/1.3 300 mesh Au, were glow discharged using a PELCO EasiGlow glow-discharge unit (15 mA for 30 seconds with a 10-

second hold). These treated grids were used within 30 minutes. Protein samples were diluted to the working concentration immediately before application to the grid. Where indicated, the sample (~3.5 µL) was supplemented with LEA proteins, CHAPSO, or MgCl2 just before being applied to the glow-discharged grid. Typically, the high concentration LEA protein stock was first diluted to an intermediate working solution using the sample buffer, and then mixed with the sample (e.g., 2 µL of LEA with 2 µL of sample). The grid was then blotted for 4 to 6 seconds at 4°C and 95% humidity before being plunge frozen into liquid ethane using a Vitrobot Mark IV (Thermo Fisher, USA).

[0234] For all conditions except those specified, 1 – 1.5 µM of PP or PRC2 were used with the indicated molar ratio of AavLEA1 or $RvLEAM_{short}$. For conditions using PP and 4 mM CHAPSO, 13.8 µM of PP was utilized.

[0235] Cryo-EM data collection

[0236] All data collections and screenings were conducted on a Talos Arctica 200 kV TEM (Thermo Fisher Scientific, USA) equipped with a Gatan BioQuantum K3 direct electron detector (Gatan, USA). Data screening and acquisition were managed using either EPU (Thermo Fisher, USA) or SerialEM [61]. All cryo-EM datasets were collected at a pixel size of 1.064 Å/pixel, with a total dose of 50 e^{-} $Å^{-2}$ distributed across 40 frames. The CDS counting mode was utilized along with a 20 eV energy filter slit. The defocus range was set between −1 and −2.5 µm in 0.25 µm intervals.

[0237] Cryo-EM data processing

[0238] For all datasets, image processing was carried out using cryoSPARC [62]. In brief, movies were subjected to patch motion correction, and the aligned micrographs had their contrast transfer function (CTF) estimated. The CTF values were utilized to select a subset of micrographs deemed suitable for high-resolution single-particle analysis. Detailed procedures for subsequent image processing steps specific to each dataset are outlined below.

[0239] 1.5 µM PP, 12 µM AavLEA1 (1:8) dataset

[0240] A total of 2,764 movies were collected. After micrograph curation 2,112 movies remained and 1,724,984 particles were extracted at 4x binning (4.3 Å per pixel). After 2D classification, 785,653 particles proceeded to ab-initio reconstruction and sorted into four separated reference-free 3D classes. Particles underwent another round of ab-initio modeling and separated into two classes. The intact complex was sorted into one of the two classes (239,029

particles, 75%) and non-uniform refinement of this class with per-particle CTF refinement resulted in a global resolution (reported at Fourier shell correlation of 0.143) of 3.6 Å.

[0241] 1.5 μM PP, 60 μM AavLEA1 (1:40) dataset

[0242] A total of 4,403 movies were collected. After micrograph curation, 500 movies were initially used with a total of 340,430 particles extracted at 4x binning. From 2D classification, 201,337 particles were selected and re-extracted at original pixel size. Particles then proceeded to ab-initio reconstruction and were sorted into four separated reference-free 3D classes. Two of the four classes resulted in intact particles, which were verified through non-uniform refinement of the combined classes using 165,472 particles, resulting in a 3.8 Å structure. Particles were extracted from the remaining 3,771 movies and binned 4x, resulting in 2,668,602 particles. These particles were then sorted into 2D classes, and the selected 1,438,074 particles underwent ab-initio modeling. Selected particles underwent another round of ab-initio modeling with two classes. One of the two classes showed intact particles, and those 1,009,026 particles were sent to non-uniform refinement yielding a 3.4 Å global resolution. CryoSparc global and local CTF refinement jobs were run followed by further filtering, resulting in 988,417 particles. These particles were then extracted at the original pixel size. All resulting particles underwent non-uniform refinement, reference motion correction, and heterogeneous refinement. 856,205 particles were used for a final non-uniform refinement, with a final global resolution of 3.0 Å.

[0243] 1 μM PP, 6 μM $RvLEAM_{short}$ (1:6) dataset

[0244] A total of 1,308 movies were collected. Following micrograph curation, 335,907 particles were extracted from 1,076 micrographs and binned 4x. 125,231 particles were extracted with 2D classification. These particles then proceeded to ab-initio reconstruction and split into 3D classes. The selected 74,198 particles were re-extracted at the original pixel size from 1,071 micrographs. Particles underwent another round of 2D classification and Ab-Initio 3D reconstruction. Selected particles underwent non-uniform refinement and had a final global resolution of 4.5 Å. The final resolution of this dataset is lower than our other PP datasets which is likely because this data collection contains only 1,308 movies while other PP datasets had 2,700 or more movies.

[0245] 13.8 μM PP, 4 mM CHAPSO dataset

[0246] A total of 4,567 movies were collected. Following micrograph curation, 1,832,455 particles were extracted from 4,510 micrographs and binned 4x. Particles underwent 2D

classification where and ab-initio reconstruction. Selected particles were used to re-extract at the original pixel size (744,824 particles). Extracted particles were then sorted into 2D classes underwent non-uniform refinement with per-particle CTF refinement and reference motion correction, with a final global resolution of 3.4 Å with 674,793 particles.

[0247] 1.5 μM PRC2, 60 μM AavLEA1 (1:40) dataset

[0248] A total of 2,843 movies were collected. Following micrograph curation, 687,121 particles were extracted from 1,642 micrographs and binned 4x. Particles underwent 2D classification and ab-initio reconstruction. Selected classes were re-extracted at original pixel size, yielding 150,359 particles. The re-extracted particles underwent non-uniform refinement with per-particle CTF refinement and resulted in a global resolution of 3.8 Å.

[0249] 1 μM PRC2, 6.7 μM $RvLEAM_{short}$ (~1:6) dataset

[0250] A total of 3,896 movies were collected. Following micrograph curation, 1,766,944 particles were extracted from 3,637 micrographs and binned 4x. Particles were sorted into 2D classes followed by 3D classes via ab-Initio reconstruction. The selected 547,610 particles were re-extracted and subjected to ab-initio 3D reconstruction. The final 206,807 particles underwent non-uniform refinement with per-particle CTF refinement, resulting in a global resolution of 3.7 Å.

[0251] 1 μM PRC2, 6.7 μM $RvLEAM_{short}$ (~1:6) 10 mM MgCl2 dataset

[0252] A total of 4,333 movies were collected. Following micrograph curation, 1,755,138 particles were extracted from 3,759 micrographs and binned 4x. Particles underwent two rounds of 2D classification From here, 408,373 particles were selected and proceeded to ab-initio 3D reconstruction. The resulting 406,148 particles were re-extracted from 3,746 movies at the original pixel size and underwent another round of ab-initio reconstruction. The final 102,181 particles proceeded to non-uniform refinement with a final global resolution was 4.2 Å.

[0253] 1 μM PRC2 cross-linked 2 minutes, 6.7 μM $RvLEAM_{short}$ (~1:6) dataset

[0254] A total of 2,981 movies were collected. Following micrograph curation, 982,264 particles were extracted from 2,189 movies and binned 4x. Particles were sorted into 2D classes followed by 3D classes via ab-initio reconstruction. Selected particles were re-extracted at the original pixel size and underwent ab-initio reconstruction again. The resulting 534,068 particles were subjected to non-uniform refinement with per-particle CTF refinement, resulting in a global resolution of 3.5 Å.

[0255] 1 μM PRC2 cross-linked 10 minutes, 6.7 μM $RvLEAM_{short}$ (~1:6) dataset

[0256] A total of 3,432 movies were collected. Following micrograph curation, 2,084,816 particles were extracted from 3,295 movies and binned 4x. Particles underwent 2D classification followed by ab-initio 3D reconstruction. Selected particles were re-extracted at the original pixel size. Particles underwent non-uniform refinement, resulting in a global resolution of 3.3 Å. Global CTF refinement, reference-based motion correction and heterogeneous refinement jobs were run. 366,459 particles were used in this final round of non-uniform refinement, resulting in a global resolution of 3.1 Å.

[0257] 1 μM PRC2 cross-linked 10 minutes dataset

[0258] A total of 3,274 movies were collected. Following micrograph curation, 1,231,903 particles were extracted from 3,274 movies and binned 4x. Particles underwent 2D classification followed by ab-initio 3D reconstruction. Selected particles were re-extracted at the original pixel size and underwent a second batch of ab-initio 3D reconstruction, reference-based motion correction non-uniform refinement. The final 51,494 particles resulted in a global resolution of 4.3 Å.

[0259] Cryo-EM structure modeling and refinement

[0260] The published apo-state models of Polymerase alpha-primase (PP, PDB: 5EXR) [46] and Polycomb Repressive Complex 2 (PRC2, PDB: 8FYH) [44] were used as initial models for real-space refinement against their respective cryo-EM maps using Phenix63. Structural alignments between the published models and the refined models were performed using the MatchMaker module in ChimeraX [64]. Refinement statistics and validation reports were obtained for PP and PRC2, respectively. Q-score [65] analysis was conducted for each refined model and reported in the abovementioned tables.

[0261] Particle image quality and orientation distribution analysis

[0262] ResLog plots were generated using the ResLog job in cryoSPARC [62]. Rosenthal-Henderson plots were derived from the data utilized in the ResLog plot, following the method described by Rosenthal and Henderson [57]. The per-particle spectral signal-to-noise ratio (ppSSNR) plots were produced using the FSC_noisesub data from the ResLog Analysis job in cryoSPARC. Particles were segmented into stacks of 30,000, 60,000, 90,000, 120,000, and 150,000 for analysis. The ppSSNR calculations are performed as described by Chan et al [66]. Cryo-EM map sphericity values are calculated using the 3DFSC server [14]. The conical FSC Area

Ratio (cFAR) and Sampling Compensation Factor (SCF) [56, 67] were computed using CryoSPARC [62].

[0263] Tilt-series collection

[0264] Tilt series were collected on a Titan Krios (Thermo Fisher Scientific, USA) operating at 300 kV, equipped with a K3 summit direct electron detector and a Quantum energy filter (Gatan, USA), controlled by SerialEM [61]. Images were collected with an exposure of 8 e-/pixel/s on the detector, with the camera operating in CDS mode, with a calibrated pixel size of 1.1 Å per pixel. Tilt series were collected using a dose-symmetric tilt scheme from −45° to 45° with a tilt increment of 3° and nominal defocus between 2 and 4 μm [68]. Each tilt angle was collected as a 5-frame movie, with an exposure of 5 e–/Å2 per tilt, and therefore a total exposure of 150 e–/Å2 per tilt series.

[0265] Tilt-series data processing

[0266] Each movie was whole frame aligned using the Unblur package in cisTEM [69]. Tilt series were aligned and reconstructed at a binning factor 4 using AreTomo [70]. To further enhance the contrast of protein particles for better localization, the reconstructed tomograms were denoised using IsoNet [71]. These movies were created using 3dmod from the IMOD package [72] and ImageJ [73].

[0267] Particle localization and ice surface estimation

[0268] Particles in the tomograms were picked manually with Dynamo [74]. The surfaces of vitrified ice were determined via three markers: crystalline ice contaminations above the ice, carbon edges and the protein layers which are assumed to be close to the surface.

[0269] Methods References

[0270] 60. Tripathi, R., Benz, N., Culleton, B., Trouvé, P. & Férec, C. Biophysical Characterisation of Calumenin as a Charged F508del-CFTR Folding Modulator. PLoS One 9, e104970 (2014).

[0271] 61. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152, 36–51 (2005).

[0272] 62. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296 (2017).

[0273] 63. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221 (2010).

[0274] 64. Meng, E. C. et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Science 32, e4792 (2023).

[0275] 65. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. Nat Methods 17, (2020).

[0276] 66. Chan, L. M. et al. High-resolution single-particle imaging at 100-200 keV with the Gatan Alpine direct electron detector. Preprint at https://doi.org/10.1101/2024.02.14.580363 (2024).

[0277] 67. Baldwin, P. R. & Lyumkis, D. Tools for visualizing and analyzing Fourier space sampling in Cryo-EM. Prog Biophys Mol Biol 160, (2021).

[0278] 68. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. J Struct Biol 197, (2017).

[0279] 69. Grant, T., Rohou, A. & Grigorieff, N. CisTEM, user-friendly software for single-particle image processing. Elife 7, (2018).

[0280] 70. Zheng, S. et al. AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J Struct Biol X 6, (2022).

[0281] 71. Liu, Y. T. et al. Isotropic reconstruction for electron tomography with deep learning. Nat Commun 13, (2022).

[0282] 72. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. J Struct Biol 116, (1996).

[0283] 73. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nature Methods vol. 9 Preprint at https://doi.org/10.1038/nmeth.2089 (2012).

[0284] 74. Castaño-Díez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. J Struct Biol 178, (2012).

[0285] Cryo-EM maps and Coordinate Files

[0286] The described cryo-EM maps and coordinate files have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank (PDB) under the accession codes: EMD-43619

(Polymerase alpha-primase (PP)–AavLEA1, 1:8 molar ratio), PDB-ID 8VY3 with EMD-43628 (PP–AavLEA1, 1:40 molar ratio), EMD-43626 (PP–RvLEAM$_{short}$, 1:6 molar ratio), PDB-ID 9C8V with EMD-43627 (PP–CHAPSO 4 mM), EMD-43620 (Polycomb repressive complex 2 (PRC2)–AavLEA1, 1:40 molar ratio), EMD-43621 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio), EMD-43622 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio, 10 mM $MgCl_2$), EMD-43623 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio, 2 min crosslink), PDB-ID 9C8U with EMD-43625 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio, 10 min crosslink), and EMD-45273 (PRC2, 10 min crosslink). Each of these cryo-EM maps and coordinate files deposited in the Electron Microscopy Data Bank and the Protein Data Bank (PDB) are hereby incorporated by reference in their entirety.

[0287] Raw Data Sets (movies) and Gain References

[0288] Raw datasets (movies) and their respective gain references were deposited in the Electron Microscopy Public Image Archive (EMPIAR): EMPIAR-11963 (PP–AavLEA1, 1:8 molar ratio), EMPIAR-11964 (PP–AavLEA1, 1:40 molar ratio), EMPIAR-11965 (PP–RvLEAM$_{short}$, 1:6 molar ratio), EMPIAR-11966 (PP–CHAPSO 4 mM), EMPIAR-11975 (PRC2–AavLEA1, 1:40 molar ratio), EMPIAR-11976 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio), EMPIAR-11978 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio, 2 min crosslink), EMPIAR-11979 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio, 10 min crosslink), and EMPIAR-12125 (PRC2, 10 min crosslink). The PRC2–RvLEAM$_{short}$, 1:6 molar ratio with 10 mM $MgCl_2$ dataset is uploaded as aligned micrographs, EMPIAR-12140. Each of these raw datasets and their respective gain references deposited in the Electron Microscopy Public Image Archive (EMPIAR) are hereby incorporated by reference in their entirety.

[0289] Reconstructed Tomograms

[0290] Reconstructed tomograms are included in EMD-43619 (PP–AavLEA1, 1:8 molar ratio) and EMD-43621 (PRC2–RvLEAM$_{short}$, 1:6 molar ratio).

[0291] The terms "protectant protein" and "sacrificial protein" are used interchangeably in this disclosure.

[0292] It is contemplated that any embodied method, composition, or kit described herein can be implemented with respect to any other method, composition, or kit described herein. The methods described herein may include additional steps or may exclude one or more steps, and the steps may be performed in any order.

[0293] Where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0294] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications mentioned herein are incorporated by reference for the purpose of describing and disclosing devices, formulations and methodologies that may be used in connection with the presently described invention.

[0295] While the compositions, methods, and disclosure is susceptible to various modifications and alternative forms, specific aspects thereof have been shown by way of example in the drawings and are herein described in detail. It should be understood, however, that the description herein of specific aspects is not intended to limit the invention to the particular forms disclosed. To the contrary, it should be appreciated that many equivalents, alternatives, variations, and modifications, aside from those expressly stated, are possible and fall within the spirit and scope of the invention as defined by the appended claims.

[0296] All references listed in this application are incorporated in whole by reference for all purposes, as long as they do not conflict with the invention. While specific embodiments and examples of the disclosed subject matter have been discussed herein, these examples are illustrative and not restrictive. Many variations will become apparent to those skilled in the art upon review of this specification and the claims below.

CLAIMS

What is claimed is:

1. A method for preparation of a cryogenic electron microscopy (cryo-EM) sample, comprising:

adding one or more protectant proteins to a biological sample during preparation of the cryo-EM sample, wherein the one or more protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3.

2. The method of claim 1, wherein the biological sample comprises a biomolecule having a mass of 50 kiloDaltons (kDa) or more, 100 kDa or more, or 120 kDa or more.

3. The method of claim 1 or 2, wherein the biological sample comprises a multi-subunit protein.

4. The method of claim 2 or 3, wherein a molar ratio of the one or more protectant proteins relative to the biomolecule in the cryo-EM sample is between 60:1 and 2:1.

5. The method of any one of claims 1-4, wherein the one or more protectant proteins comprises an amino acid sequence at least 90%, 95%, or 99% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3.

6. The method of any one of claims 1-5, wherein the one or more protectant proteins have a mass of 20 kDa or less, 18 kDa or less, 17 kDa or less, or 15 kDa or less.

7. The method of any one of claims 1-6, wherein the biological sample is in liquid form.

8. The method of any one of claims 1-7, wherein the one or more protectant proteins are present in an aqueous buffer.

9. The method of any one of claims 1-8, further comprising: freezing the cryo-EM sample to provide a vitreous cryo-EM sample; performing electron microscopy on the vitreous cryo-EM sample to obtain cryo-EM data; and optionally performing two-dimensional analysis, three-dimensional analysis, or both, on the cryo-EM data.

10. A method for performing cryogenic electron microscopy (cryo-EM), comprising:

adding one or more first protectant proteins to a first portion of a biological sample to provide a first cryo-EM sample, wherein the one or more first protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3; and

adding one or more second protectant proteins to a second portion of the biological sample to provide a second cryo-EM sample, wherein the one or more second protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3, and wherein the one or more second protectant proteins are different than the one or more first protectant proteins.

11. The method of claim 10, wherein the first portion, the second portion, or both, of the biological sample comprise a biomolecule having a mass of 50 kilo Daltons (kDa) or more, 100 kDa or more, or 120 kDa or more.

12. The method of claim 10 or 11, wherein the first portion, the second portion, or both, of biological sample comprise a multi-subunit protein.

13. The method of claim 11 or 12, wherein a molar ratio of the one or more first protectant proteins relative to the biomolecule in the first cryo-EM sample is between 60:1 and 2:1, and/or wherein a molar ratio of the one or more second protectant proteins relative to the biomolecule in the second cryo-EM sample is between 60:1 and 2:1.

14. The method of any one of claims 10-13, wherein the one or more first protectant proteins comprises an amino acid sequence at least 90%, 95%, or 99% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3, and/or wherein the one or more second protectant proteins comprises an amino acid sequence at least 90%, 95%, or 99% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3.

15. The method of any one of claims 10-14, wherein the one or more first protectant proteins have a mass of 20 kDa or less, 18 kDa or less, 17 kDa or less, or 15 kDa or less, and/or wherein the one or more second protectant proteins have a mass of 20 kDa or less, 18 kDa or less, 17 kDa or less, or 15 kDa or less.

16. The method of any one of claims 10-15, wherein the first portion of the biological sample, the second portion of the biological sample, or both, are in liquid form.

17. The method of any one of claims 10-16, wherein the one or more first protectant proteins, the one or more second protectant proteins, or both, are present in an aqueous buffer.

18. The method of any one of claims 10-17, further comprising: freezing the first cryo-EM sample to provide a first vitreous cryo-EM sample; performing electron microscopy on the first vitreous cryo-EM sample to obtain a first set of cryo-EM data; and optionally performing two-dimensional analysis, three-dimensional analysis, or both, on the first set of cryo-EM data.

19. The method of claim 18, further comprising: freezing the second cryo-EM sample to provide a second vitreous cryo-EM sample; performing electron microscopy on the second vitreous cryo-EM sample to obtain a second set of cryo-EM data; and optionally performing two-dimensional analysis, three-dimensional analysis, or both, on the second set of cryo-EM data.

20. A composition for preparing a cryogenic electron microscopy (cryo-EM) sample, comprising:

one or more protectant proteins, wherein the one or more protectant proteins comprise a truncated RvLEAM1 protein comprising an amino acid sequence at least 85% identical to SEQ ID NO: 2, a truncated dAFP-1 protein comprising an amino acid sequence at least 85% identical to SEQ ID NO: 3, or both.

21. The composition of claim 20, further comprising an aqueous buffer.

22. The composition of claim 20 or 21, wherein each of the one or more protectant proteins have a mass of 20 kDa or less, 18 kDa or less, 17 kDa or less, or 15 kDa or less.

23. The composition of any one of claims 20-22, wherein the composition is in liquid form.

24. The composition of any one of claims 20-23, wherein the truncated rvLEAM1 protein comprises an amino acid sequence at least 90%, 95%, or 99% identical to SEQ ID NO: 2, and/or wherein the truncated dAFP-1 protein comprises an amino acid sequence at least 90%, 95%, or 99% identical to SEQ ID NO: 3.

25. The composition of any one of claims 20-24, wherein the truncated rvLEAM1 protein further comprises one or more purification tags, and/or wherein the truncated dAFP-1 protein further comprises one or more purification tags.

26. A construct comprising a promoter operably connected to a polynucleotide encoding a polypeptide of SEQ ID NO: 2, SEQ ID NO: 3, a polypeptide having at least 90% identity to SEQ ID NO: 2 or a polypeptide having at least 90% identity to SEQ ID NO: 3.

27. The construct of claim 26, wherein the polynucleotide may further comprise a peptide tag to allow for purification.

28. The construct of claim 27, wherein the peptide tag is a 6xHis tag or a FLAG tag.

29. The method of any one of claims 1-9, further comprising exposing the cryo-EM sample to one or more crosslinking agents.

30. The method of any one of claims 10-19, further comprising exposing the first cryo-EM sample and/or the second cryo-EM sample to one or more crosslinking agents.

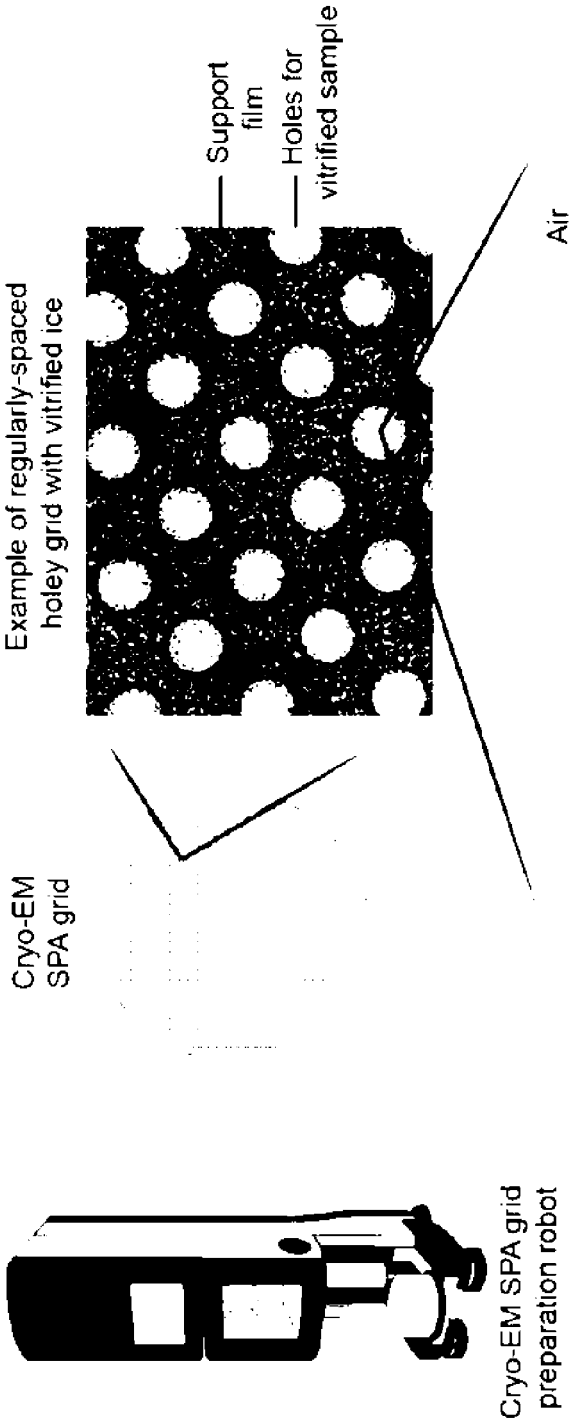
Cryo-EM SPA grid preparation robot
Cryo-EM SPA grid
Example of regularly-spaced holey grid with vitrified ice
Support film
Holes for vitrified sample
Air

FIG. 1A


Water
Vitrified sample
Support film
Ideal
(or commonly perceived)
sample distribution
in thin film of vitrified ice

FIG. 1B

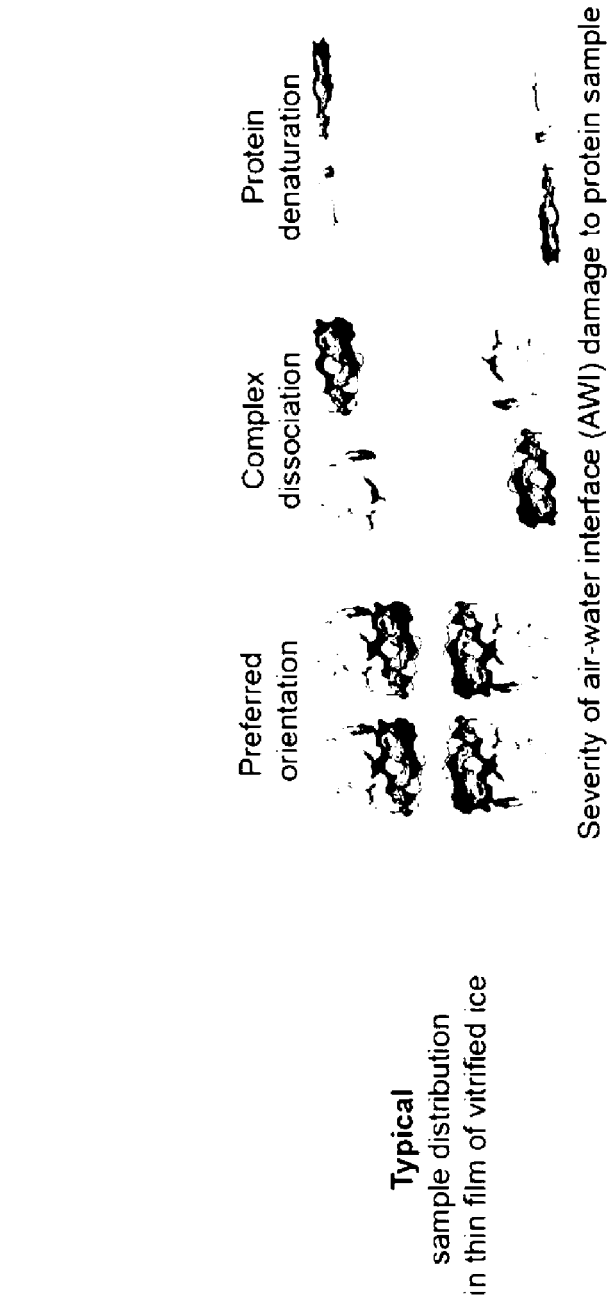
Typical
sample distribution
in thin film of vitrified ice
Preferred
orientation
Complex
dissociation
Protein
denaturation
Severity of air-water interface (AWI) damage to protein sample

FIG. 1C

Rescued
sample distribution
in thin film of vitrified ice with
small sacrificial proteins

Small sacrificial proteins forming a barrier at AWI

FIG. 1D

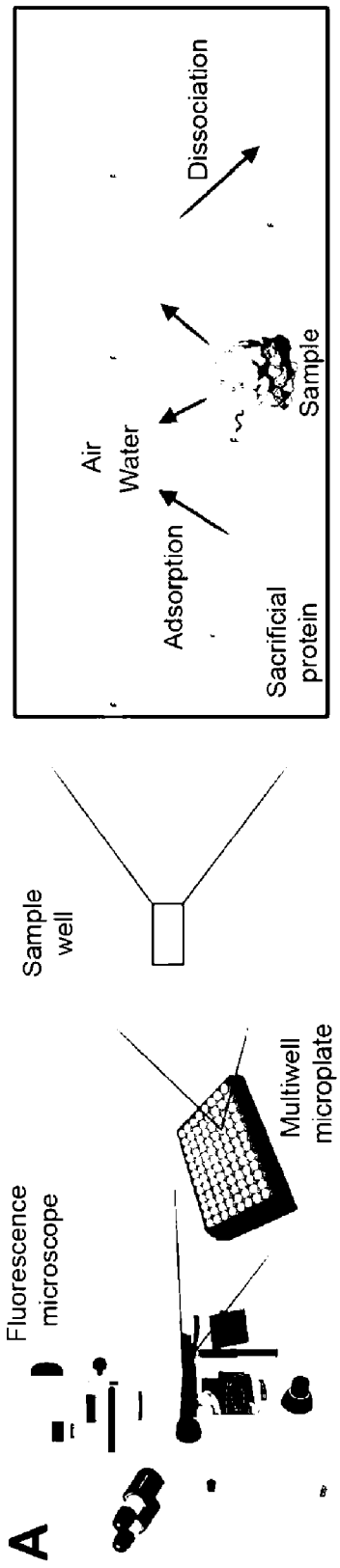
A
Fluorescence microscope
Multiwell microplate
Sample well
Sacrificial protein
Adsorption
Air
Water
Sample
Dissociation

FIG. 2A

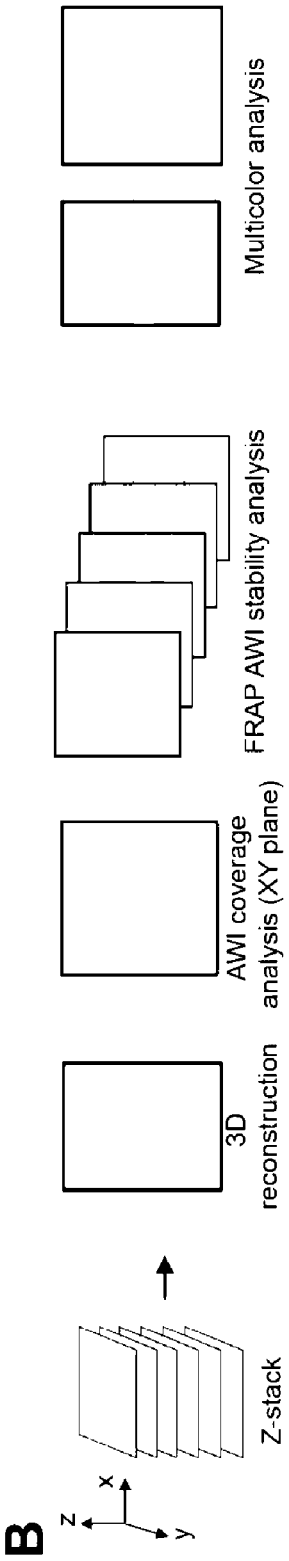
B
x
z
y
Z-stack
3D reconstruction
AWI coverage analysis (XY plane)
FRAP AWI stability analysis
Multicolor analysis

FIG. 2B

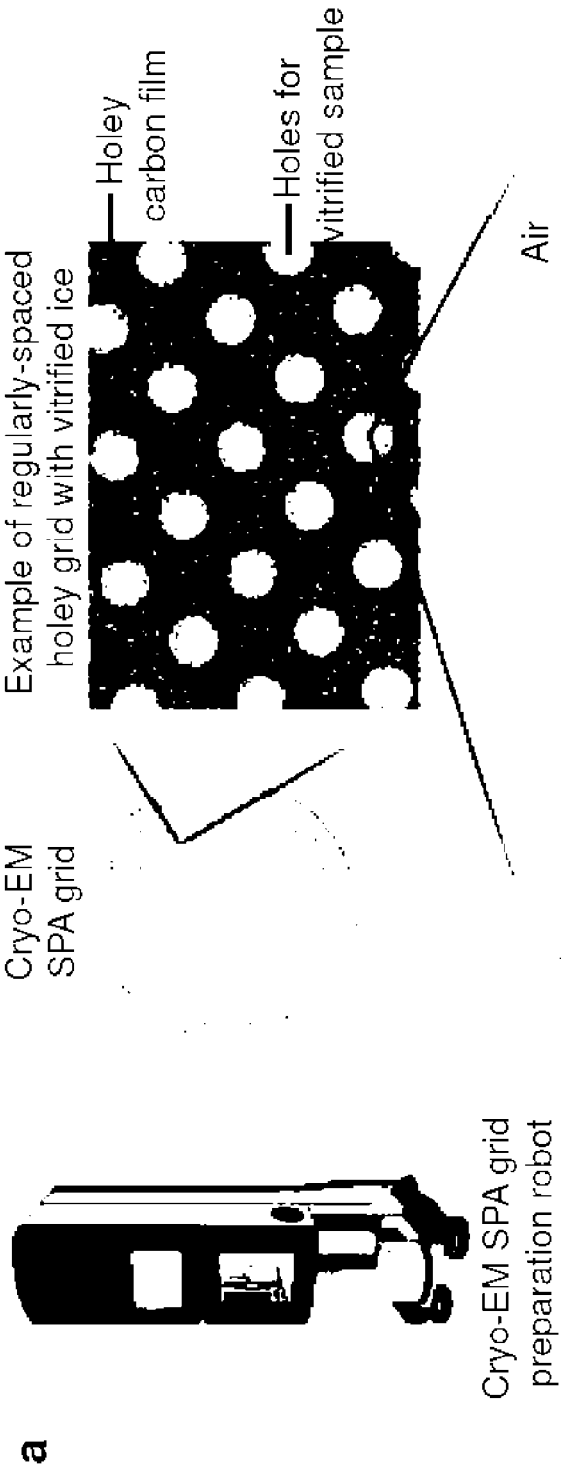
a
Cryo-EM SPA grid preparation robot
Cryo-EM SPA grid
Example of regularly-spaced holey grid with vitrified ice
Holey carbon film
Holes for vitrified sample
Air

FIG. 3A


b
Ideal
sample distribution
in thin film of vitrified ice
Air
Water
Vitrified sample
Carbon film

FIG. 3B

c
Typical
sample distribution
in thin film of vitrified ice
Preferred orientation
Complex dissociation
Protein denaturation
Severity of air-water interface (AWI) damage to protein sample

FIG. 3C

d **Rescued**
sample distribution
in thin film of vitrified ice
with **LEA proteins**

FIG. 3D

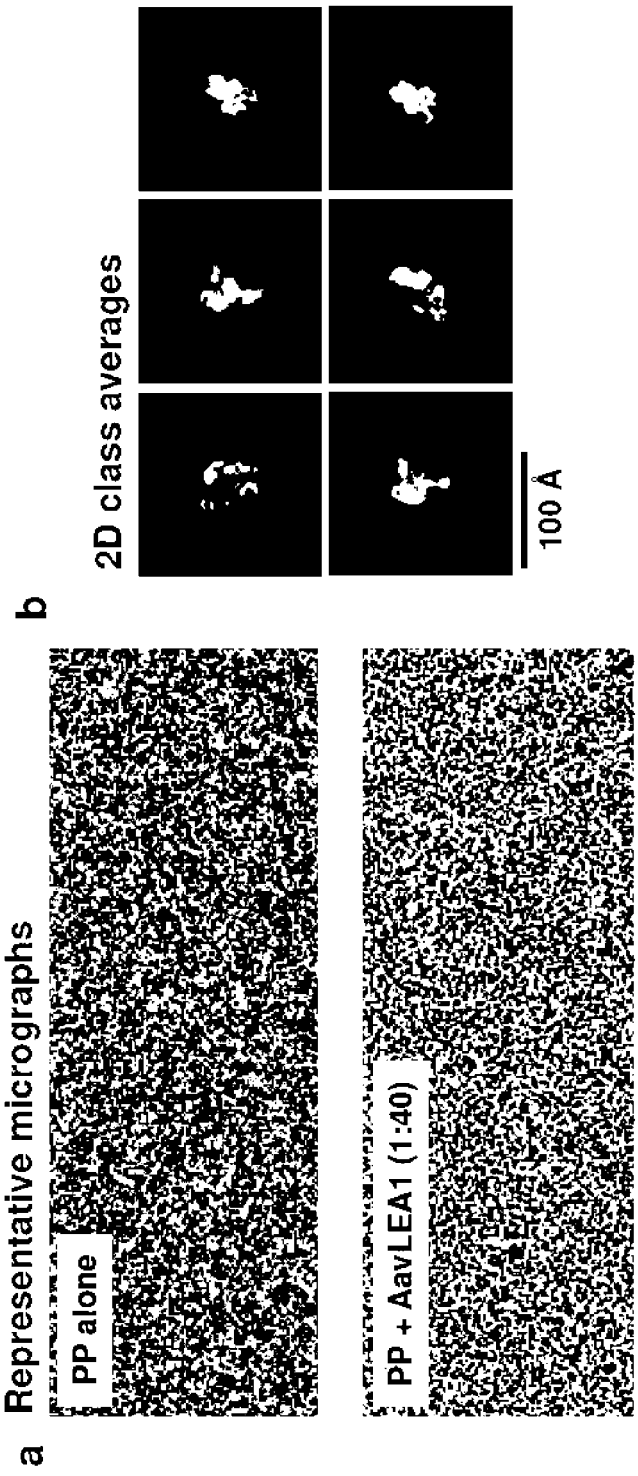
a
Representative micrographs
PP alone
PP + AavLEA1 (1:40)
b
2D class averages
100 Å

FIG. 4A

FIG. 4B

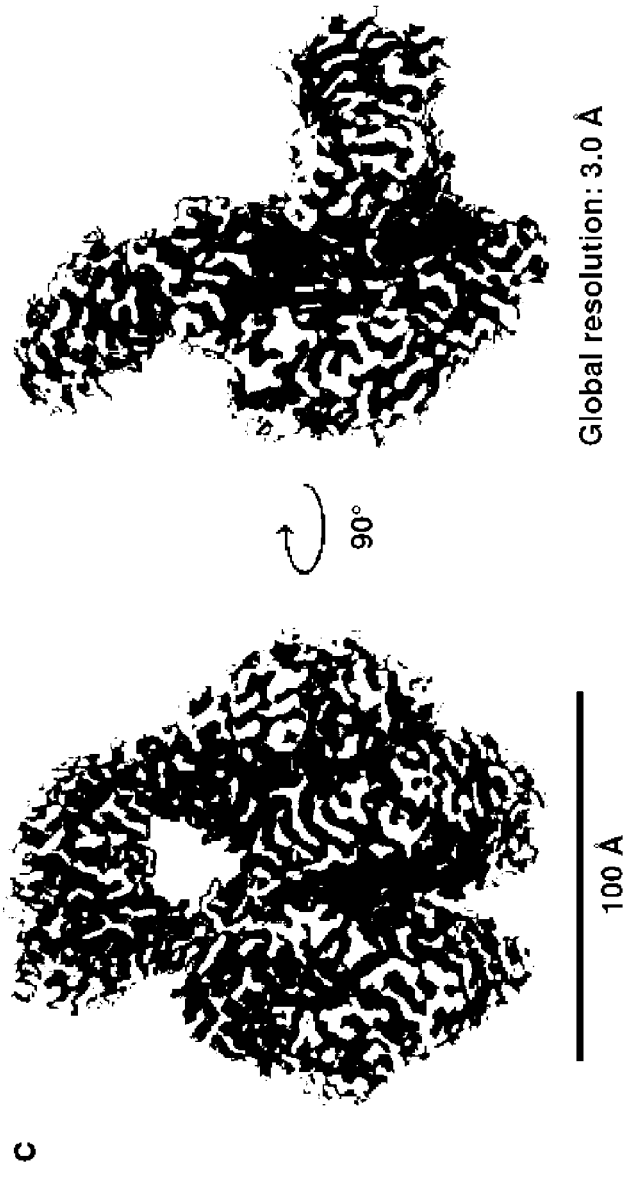

FIG. 4C

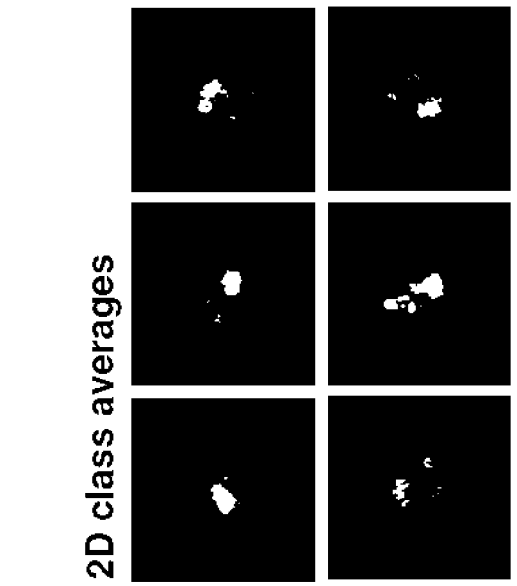
e
2D class averages
100 Å

FIG. 4E

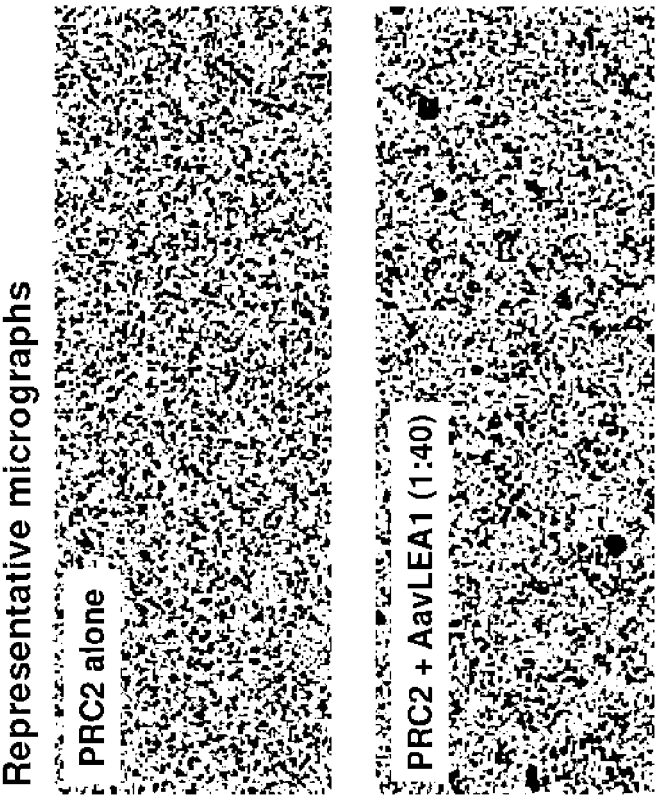
d
Representative micrographs
PRC2 alone
PRC2 + AavLEA1 (1:40)

FIG. 4D

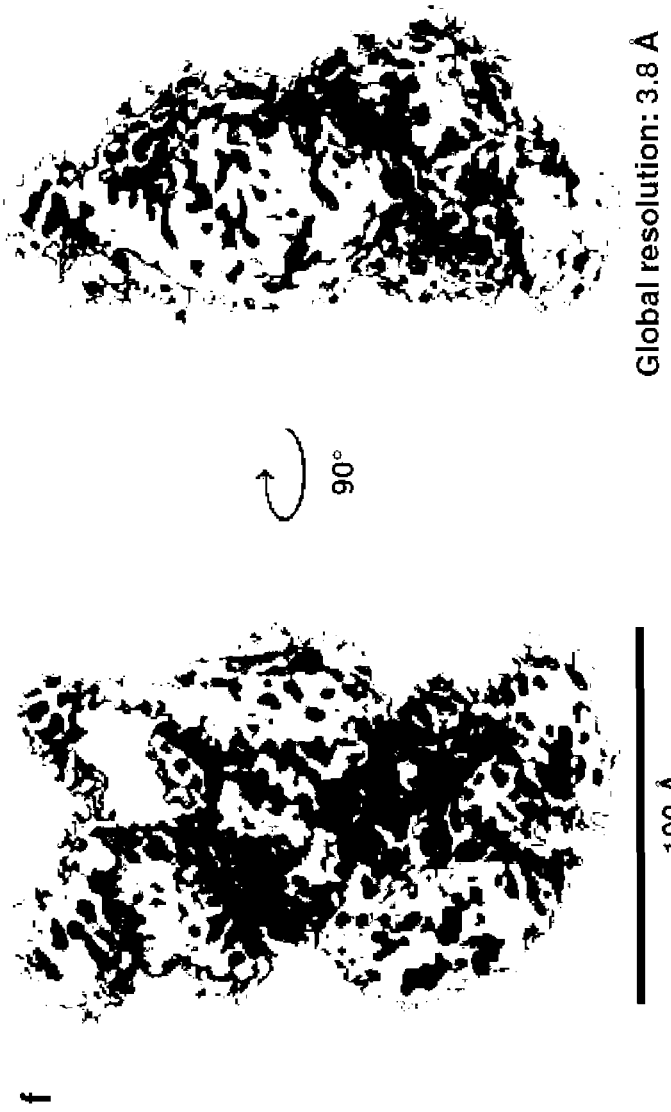

FIG. 4F

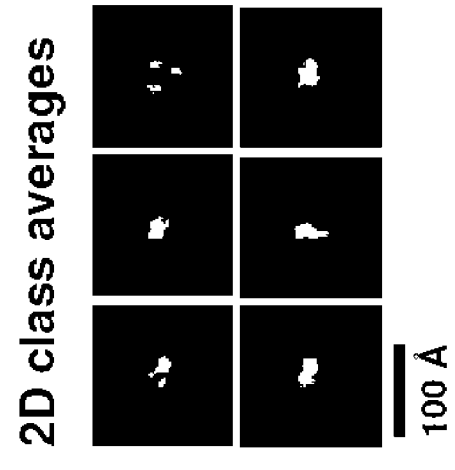
b 2D class averages
100 Å

FIG. 5B

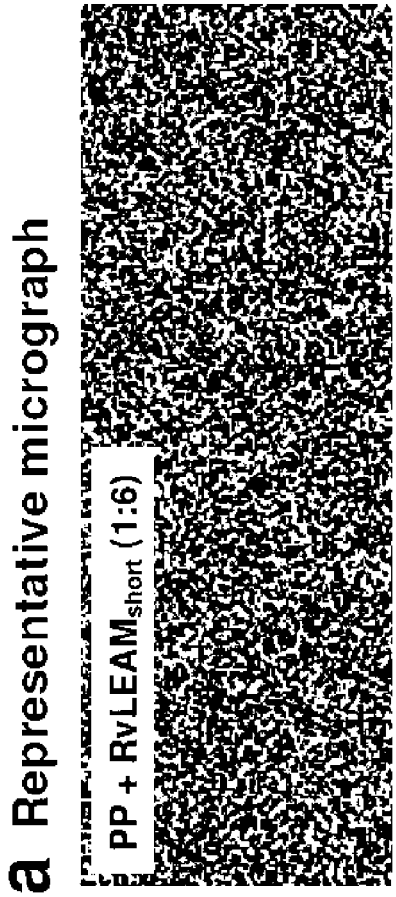
a Representative micrograph
PP + RvLEAMshort (1:6)

FIG. 5A

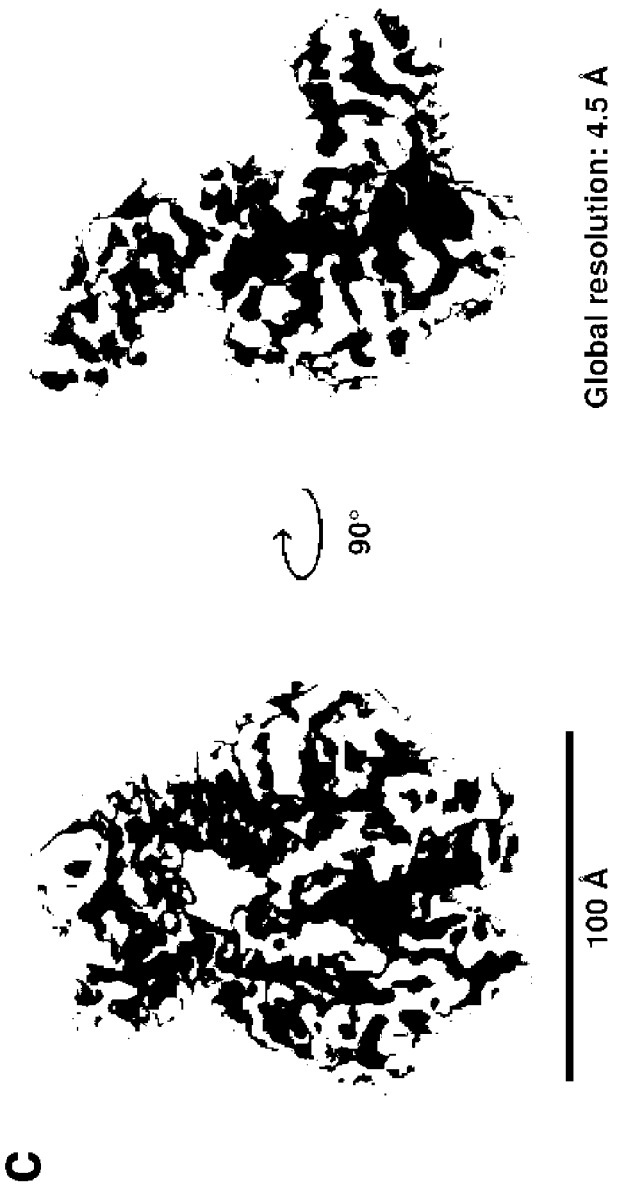

FIG. 5C

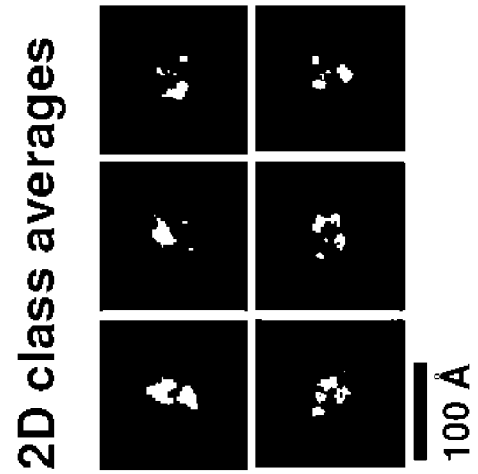
e 2D class averages
100 Å

FIG. 5E

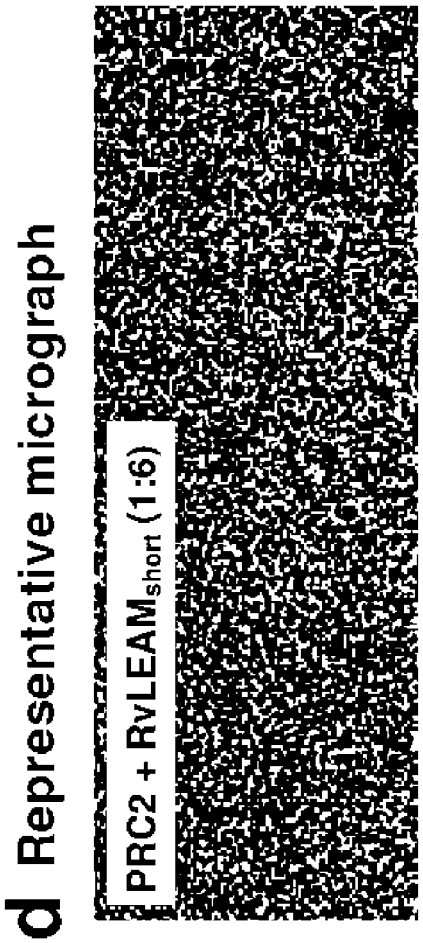
d Representative micrograph
PRC2 + RvLEAM$_{short}$ (1:6)

FIG. 5D

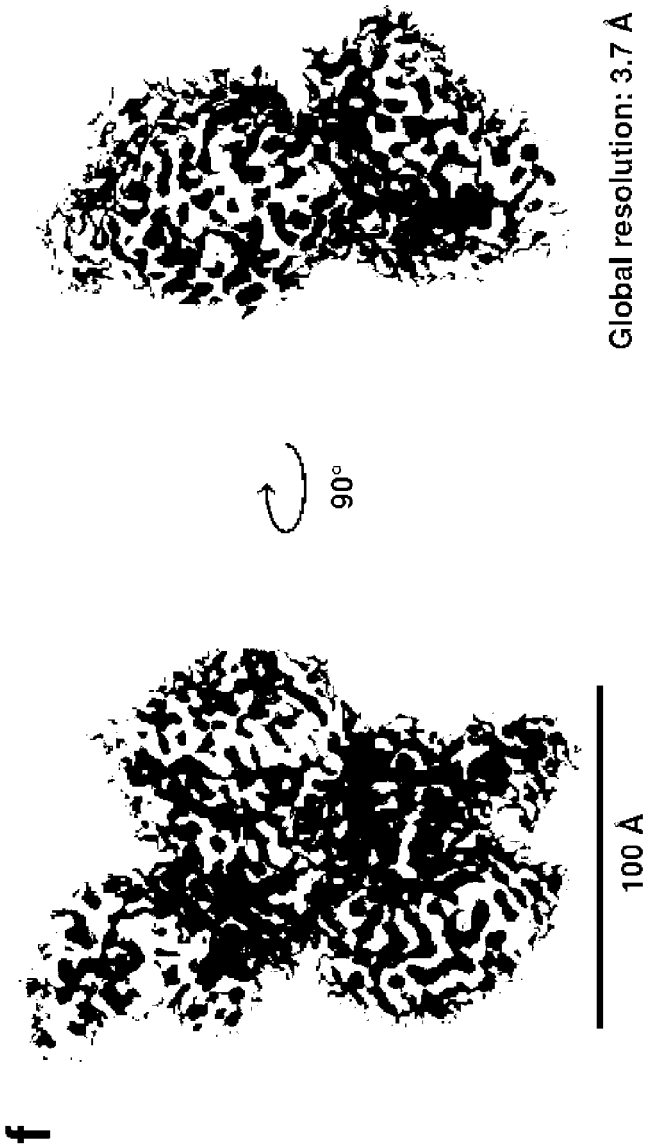

FIG. 5F

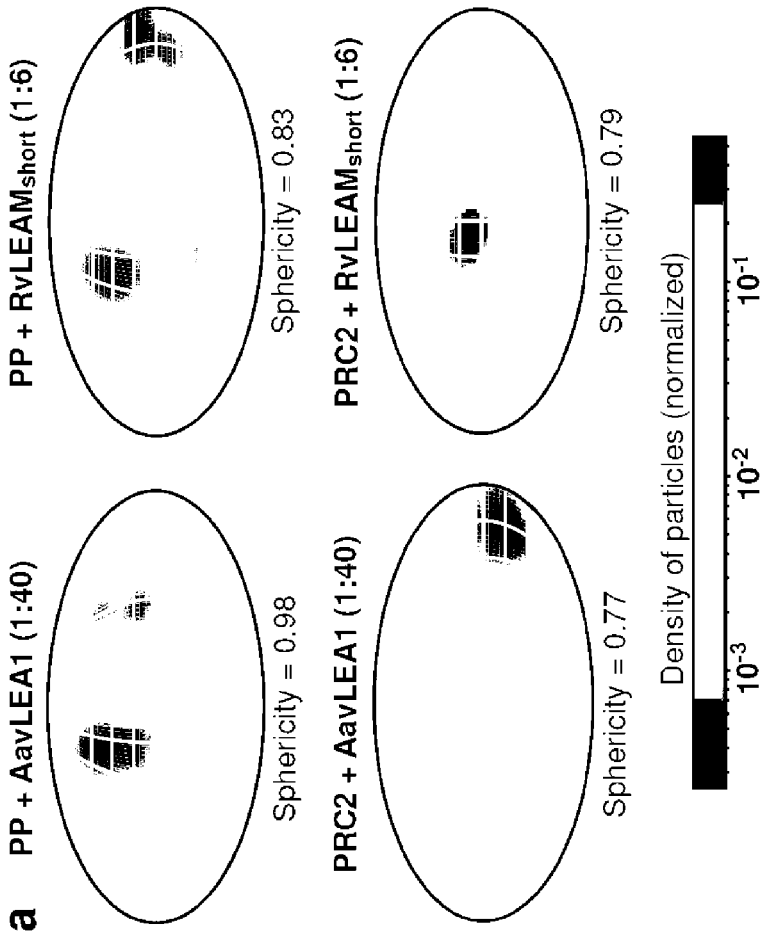
a
PP + AavLEA1 (1:40)
Sphericity = 0.98
PP + RvLEAM$_{short}$ (1:6)
Sphericity = 0.83
PRC2 + AavLEA1 (1:40)
Sphericity = 0.77
PRC2 + RvLEAM$_{short}$ (1:6)
Sphericity = 0.79
Density of particles (normalized)
10^{-3}
10^{-2}
10^{-1}

FIG. 6A

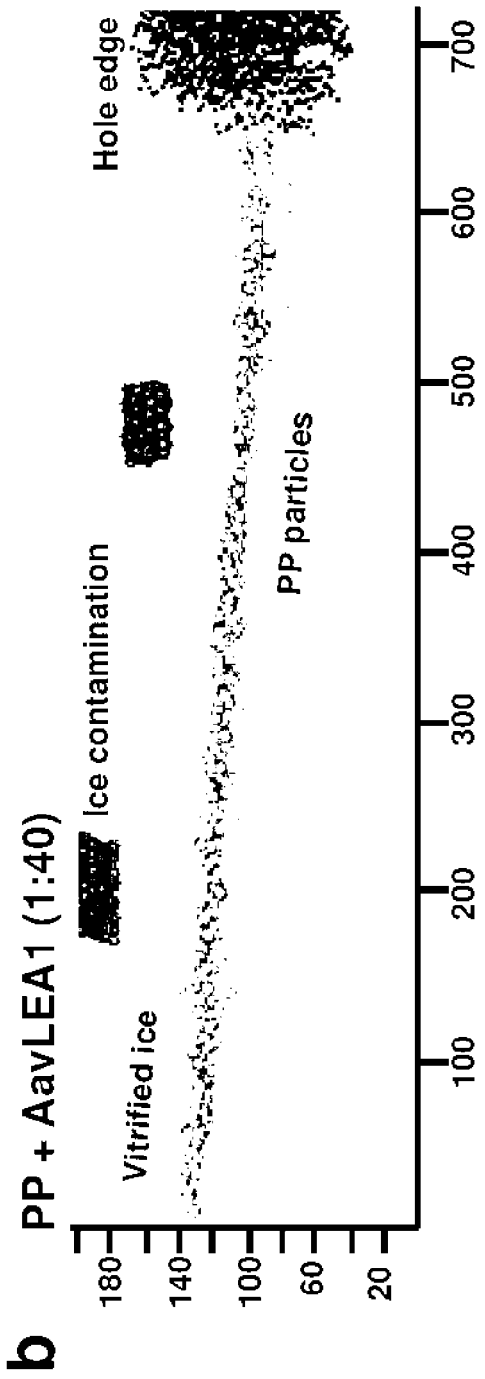
b
PP + AavLEA1 (1:40)
Vitrified ice
Ice contamination
PP particles
Hole edge
180
140
100
60
20
100
200
300
400
500
600
700

FIG. 6B

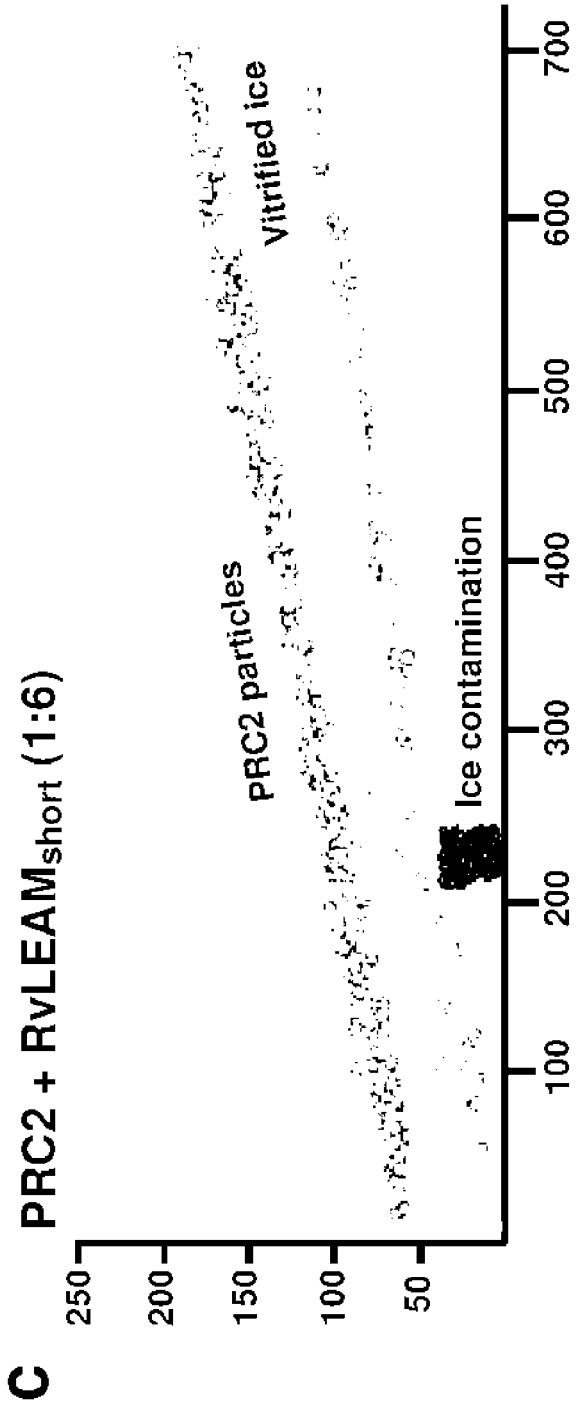
C
PRC2 + RvLEAMshort (1:6)
250
200
150
100
50
100
200
300
400
500
600
700
Ice contamination
PRC2 particles
Vitrified ice

FIG. 6C

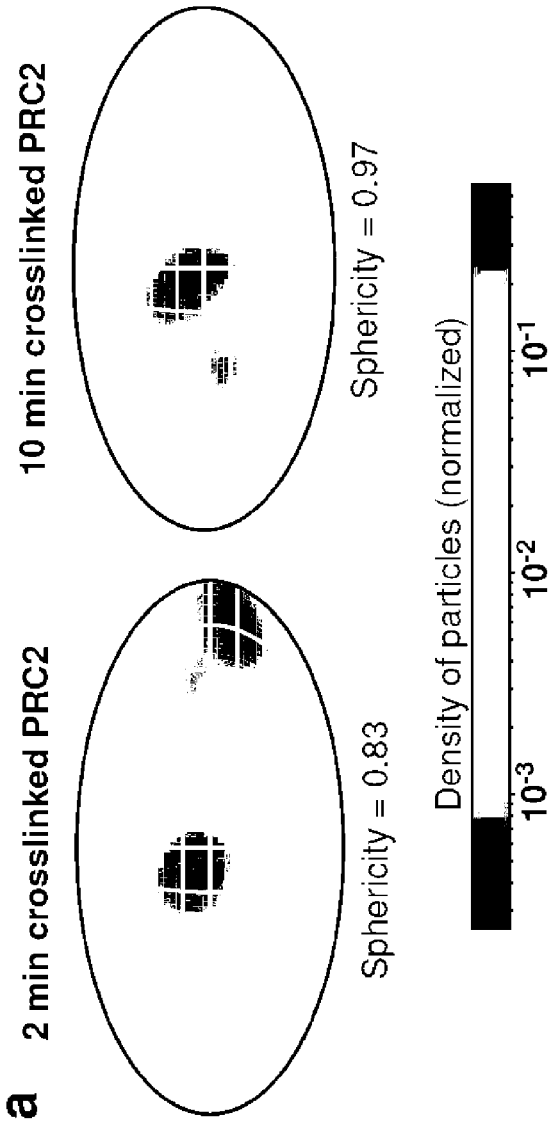
a
2 min crosslinked PRC2
Sphericity = 0.83
10 min crosslinked PRC2
Sphericity = 0.97
Density of particles (normalized)
10-3
10-2
10-1

FIG. 7A

FIG. 7B

FIG. 8A

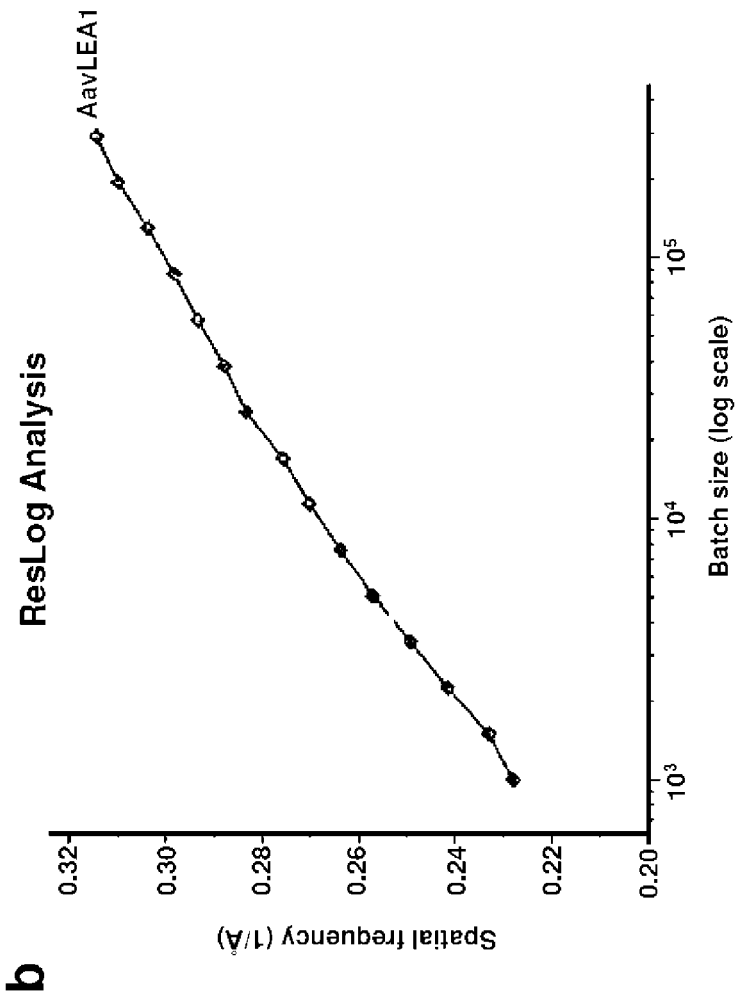
b
ResLog Analysis
AavLEA1
Spatial frequency (1/Å)
0.32
0.30
0.28
0.26
0.24
0.22
0.20
10^3
10^4
10^5
Batch size (log scale)

FIG. 8B

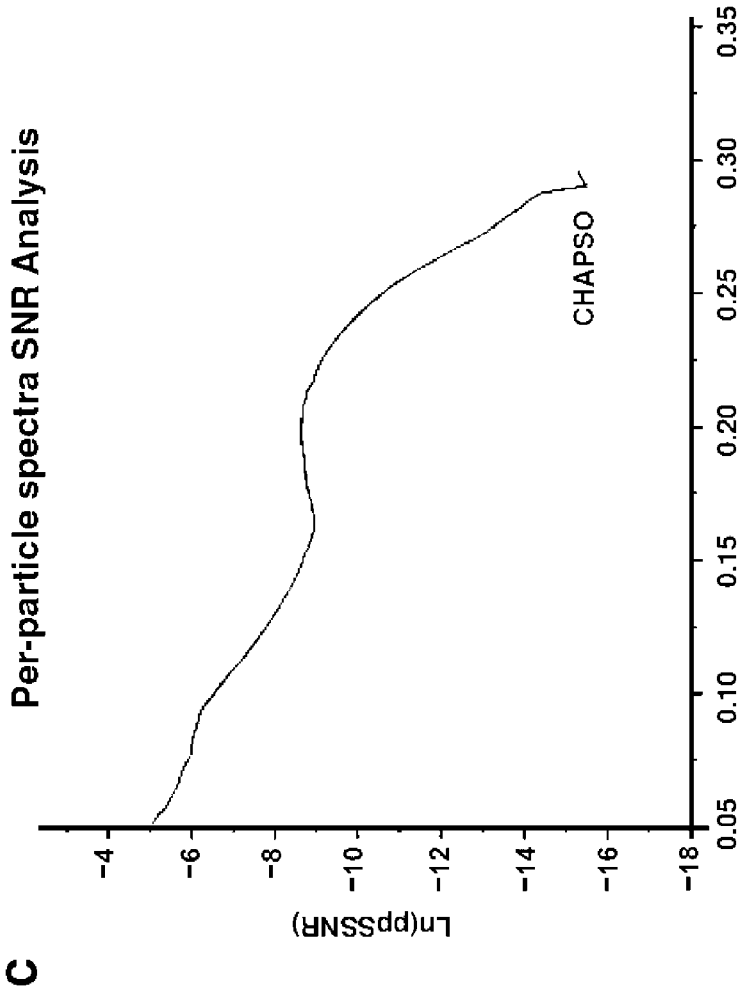
C
Per-particle spectra SNR Analysis
Ln(ppSSNR)
−4
−6
−8
−10
−12
−14
−16
−18
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Spatial frequency (1/Å)
CHAPSO

FIG. 8C

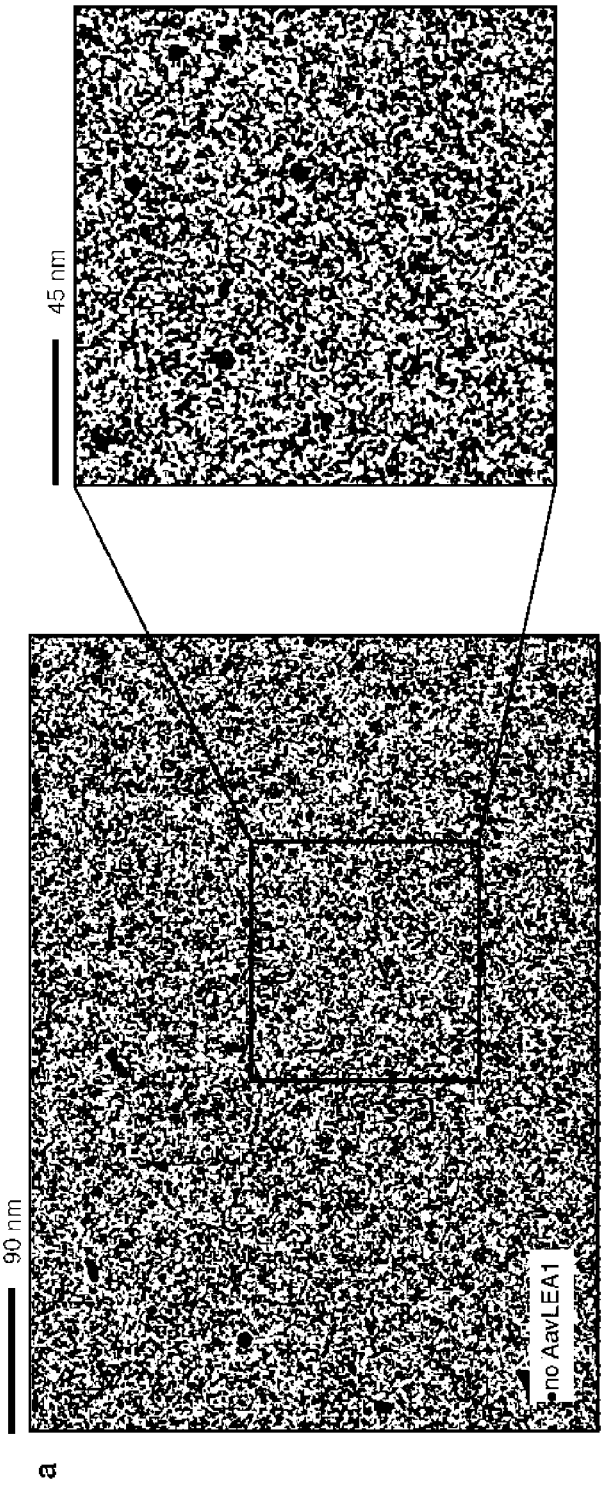

FIG. 9A

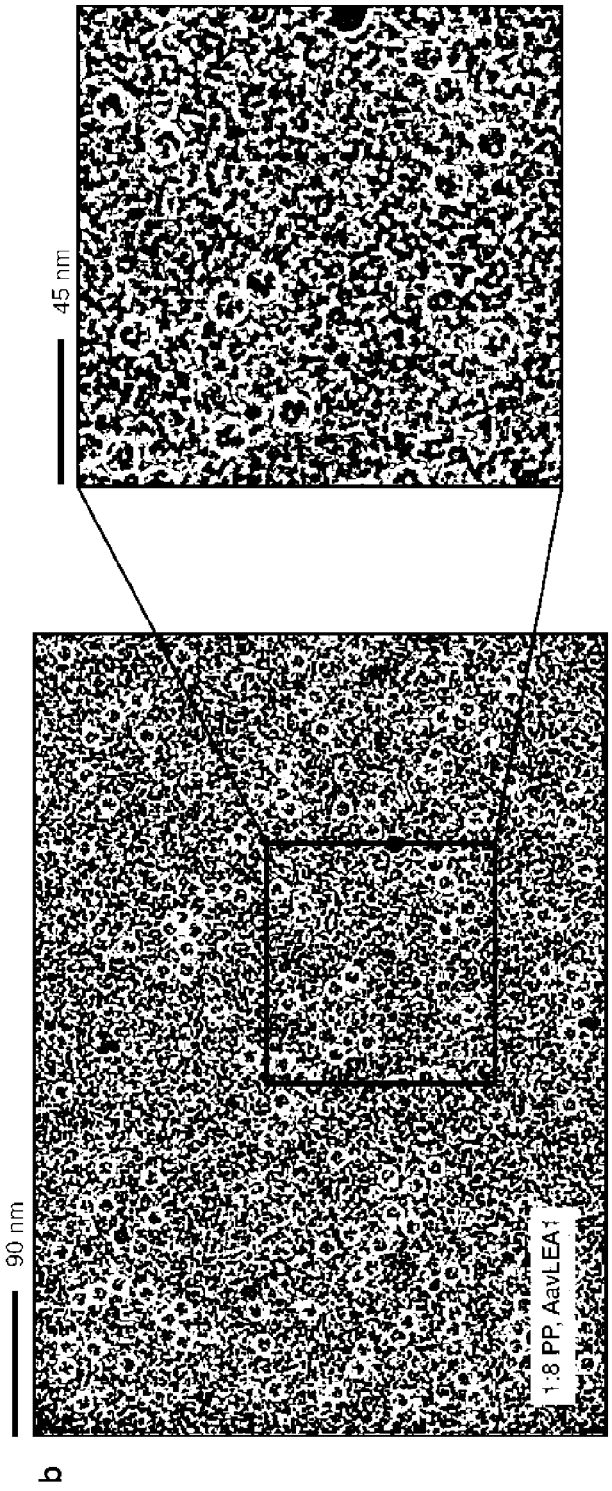

FIG. 9B

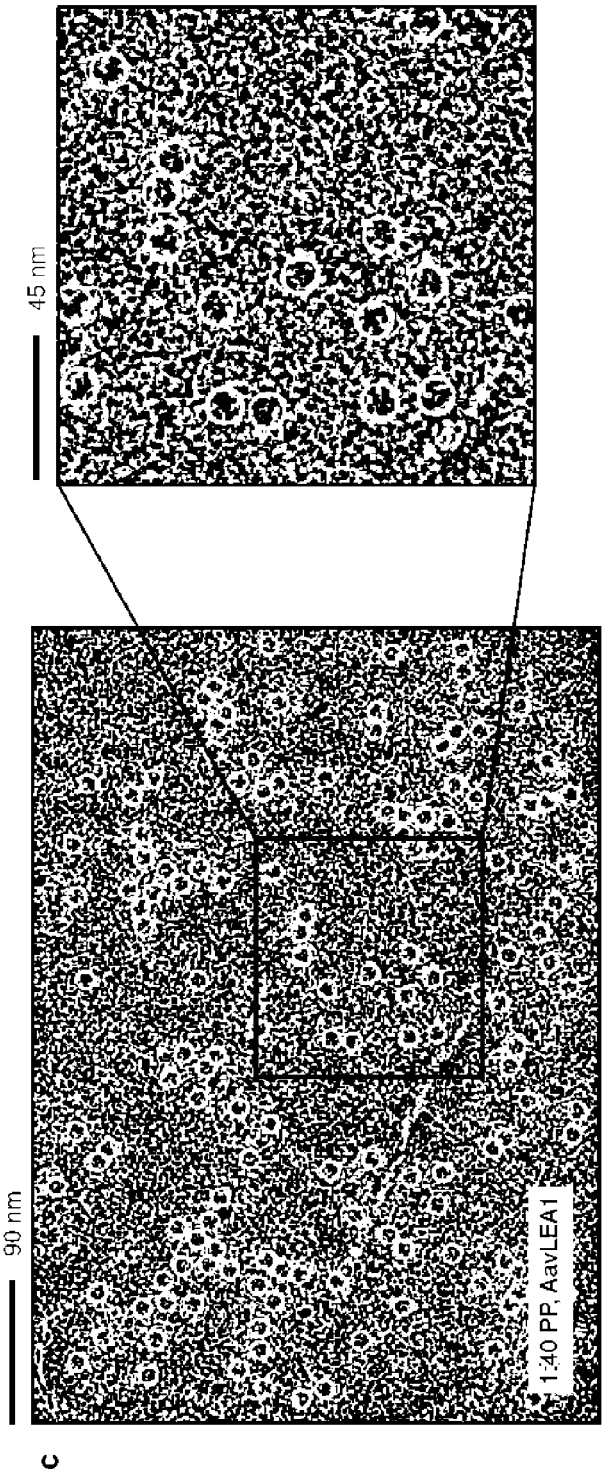

FIG. 9C

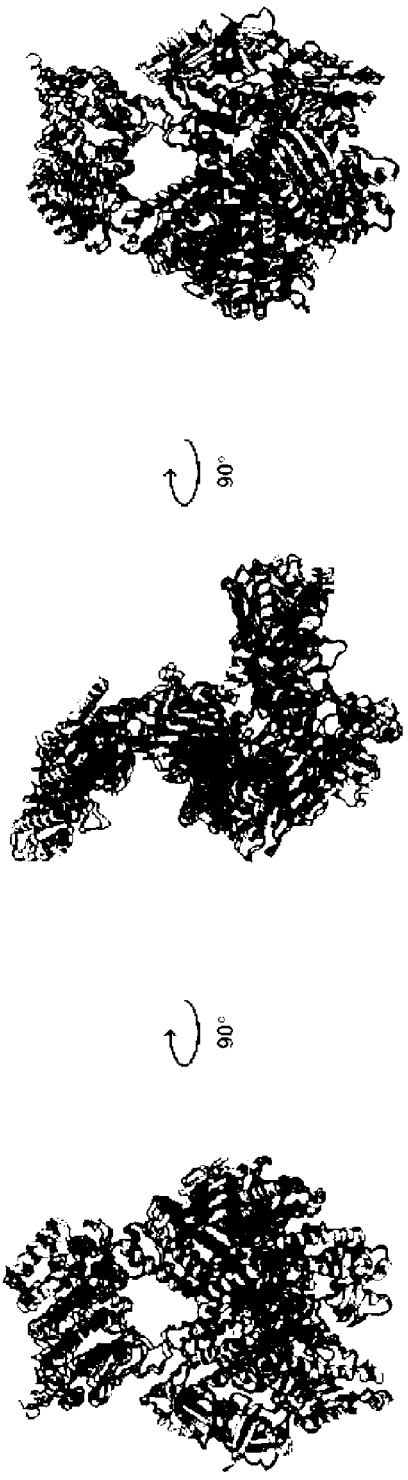

FIG. 10

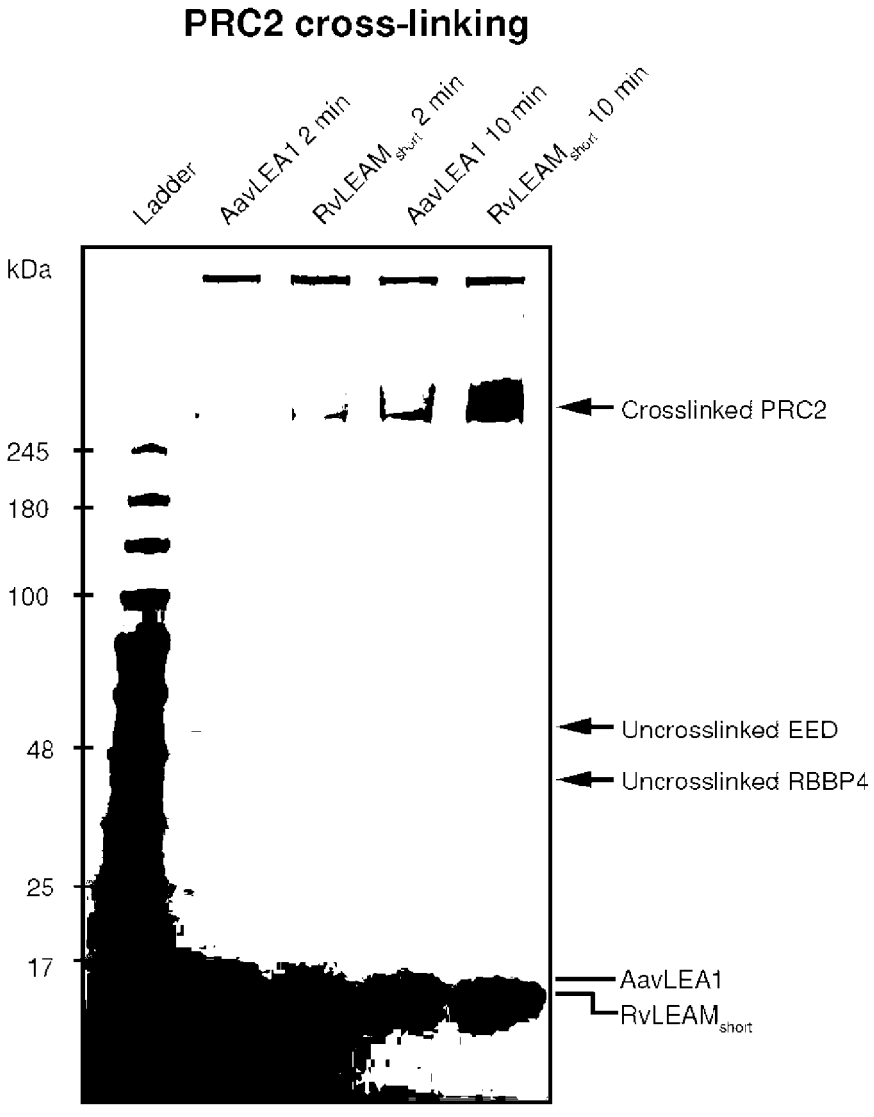
PRC2 cross-linking
Ladder
AavLEA1 2 min
RvLEAM short 2 min
AavLEA1 10 min
RvLEAM short 10 min
kDa
245
180
100
48
25
17
Crosslinked PRC2
Uncrosslinked EED
Uncrosslinked RBBP4
AavLEA1
RvLEAM short

FIG. 11

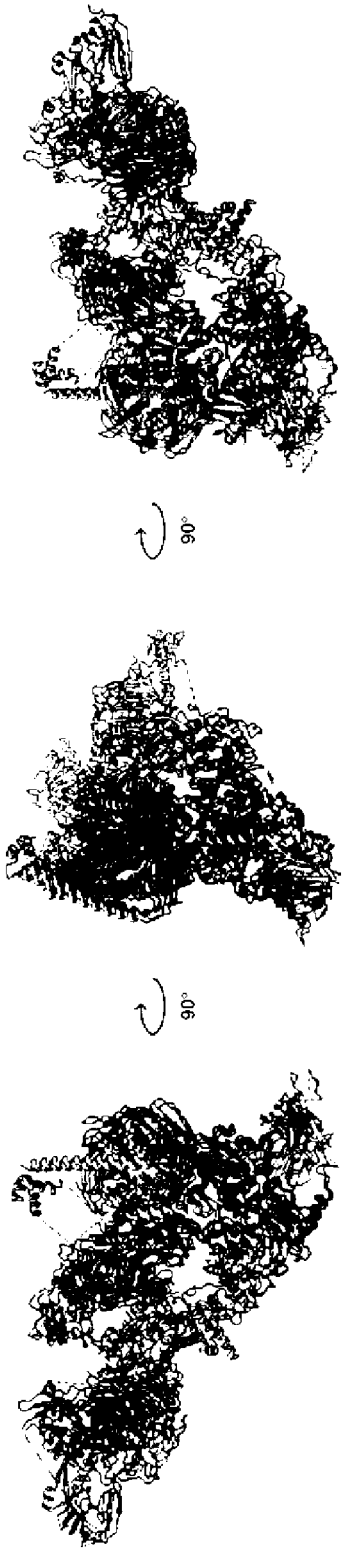

FIG. 12

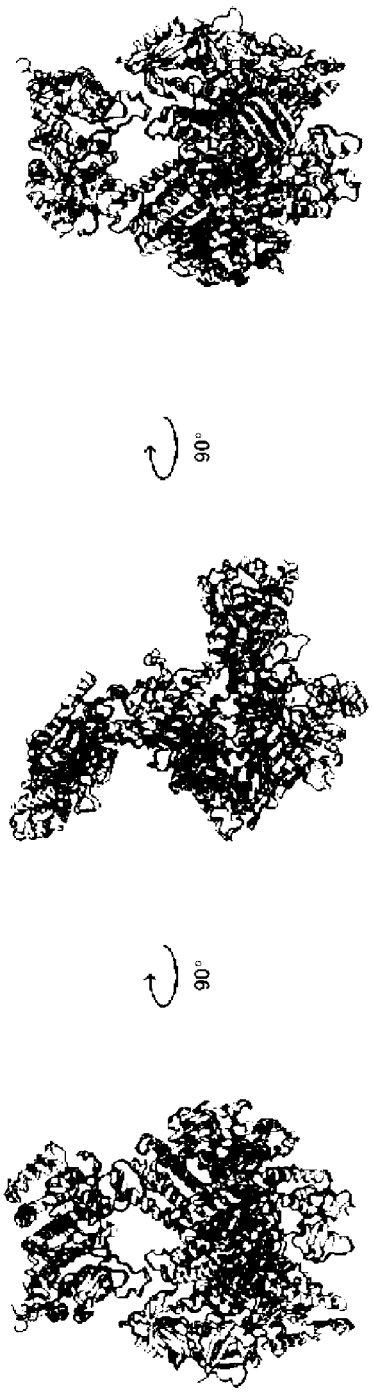

FIG. 13

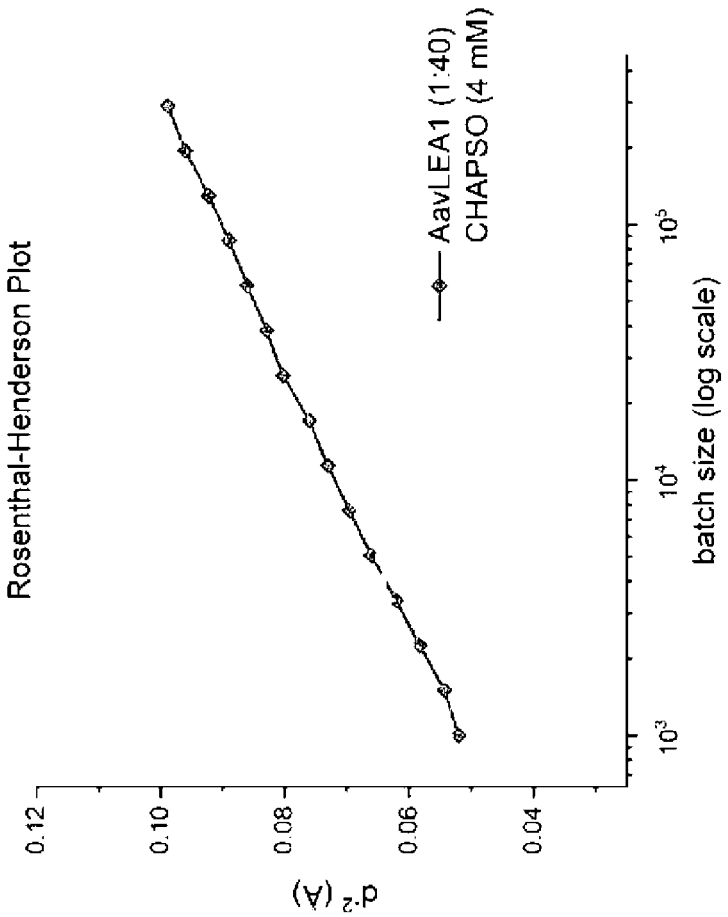
Rosenthal-Henderson Plot
AavLEA1 (1:40)
CHAPSO (4 mM)
d'² (Å)
0.12
0.10
0.08
0.06
0.04
10³
10⁴
10⁵
batch size (log scale)

FIG. 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/040485

A. CLASSIFICATION OF SUBJECT MATTER

IPC: ***G01N 1/42*** (2024.01); ***C07K 14/435*** (2024.01)

CPC: ***G01N 1/42; C07K 14/43504; C07K 14/43563***

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2018/0335438 A1 (AMPRION INC. et al.) 22 November 2018 (22.11.2018) entire document	1-3
A	WO 2018/073242 A1 (UNIVERSITAT BASEL) 26 April 2018 (26.04.2018) entire document	1-3
A	US 2006/0059583 A1 (KIKAWADA et al.) 16 March 2006 (16.03.2006) entire document	1-3
A	US 2010/0216111 A1 (SCHENK) 26 August 2010 (26.08.2010) entire document	1-3
A	US 2018/0155732 A1 (WISCONSIN ALUMNI RESEARCH FOUNDATION) 07 June 2018 (07.06.2018) entire document	1-3

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 September 2024 (20.09.2024)	**20 November 2024 (20.11.2024)**

Name and mailing address of the ISA/US	Authorized officer
COMMISSIONER FOR PATENTS **MAIL STOP PCT, ATTN: ISA/US** **P.O. Box 1450** **Alexandria, VA 22313-1450** **UNITED STATES OF AMERICA**	**TAINA MATOS**
Facsimile No. **571-273-8300**	Telephone No. **571-272-4300**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/040485

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed.
   b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/040485

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☑ Claims Nos.: **4-9, 13-19, 23-25, 29, 30**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-3, 20-22, and 26-28 are drawn to methods for performing cryogenic electron microscopy (cryo-EM), compositions for preparing a cryogenic electron microscopy (cryo-EM) sample, and constructs.

Group II+: claims 10-12 are drawn to methods for performing cryogenic electron microscopy (cryo-EM) comprising a first and second protectant protein.

The first invention of Group I+ is restricted to a protectant protein selected to be SEQ ID NO: 1 and methods, compositions, and constructs comprising the same. The first named invention has been selected based on the guidance set forth in section 10.54 of the PCT International Search and Preliminary Examination Guidelines. Specifically, the first named invention was selected based on the first listed protectant protein species presented in the claims (see claim 1). It is believed that claims 1-3 read on this first named invention and thus these claims will be searched without fee to the extent that they read on SEQ ID NO: 1.

An exemplary election of Group II+ is a protectant protein selected to be SEQ ID NO: 1, a second protectant protein selected to be SEQ ID NO: 2 and methods comprising the same.

Applicant is invited to elect additional protectant proteins and second protectant proteins and their respective, corresponding, SEQ ID NOs to be searched in a specific combination by paying additional fee for each set of election. An exemplary election would be a protectant protein selected to be SEQ ID NO: 2 and methods, compositions, and constructs comprising the same. Additional protectant proteins and second protectant proteins and their respective, corresponding, SEQ ID NOs will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ and II+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

The Groups I+ and II+ formulas do not share a significant structural element responsible for cryo-EM sample preparation requiring the selection of alternative protectant proteins and second protectant proteins where "adding one or more protectant proteins to a biological sample during preparation of the cryo-EM sample, wherein the one or more protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3" and "adding one or more first protectant proteins to a first portion of a biological sample to provide a first cryo-EM sample, wherein the one or more first protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3; and adding one or more second protectant proteins to a second portion of the biological sample to provide a second cryo-EM sample, wherein the one or more second protectant proteins comprises an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3, and wherein the one or more second protectant proteins are different than the one or more first protectant proteins."

The special technical features of Group I+, methods for performing cryogenic electron microscopy (cryo-EM), compositions for preparing a cryogenic electron microscopy (cryo-EM) sample, and constructs, are not present in Group II+; the special technical features of Group II+, methods for performing cryogenic electron microscopy (cryo-EM) comprising a first and second protectant protein, are not present in Group I+.

Additionally, even if Groups I+ and II+ were considered to share the technical features of a method for preparation of a cryogenic electron microscopy (cryo-EM) sample, comprising: adding one or more protectant proteins to a biological sample during preparation of the cryo-EM sample, wherein the one or more protectant proteins comprises an amino acid sequence; a composition for preparing a cryogenic electron microscopy (cryo-EM) sample, comprising: one or more protectant proteins; a construct comprising a promoter operably connected to a polynucleotide encoding a polypeptide; a method for performing cryogenic electron microscopy (cryo-EM), comprising: adding one or more first protectant proteins to a first portion of a biological sample to provide a first cryo-EM sample, wherein the one or more first protectant proteins comprises an amino acid sequence; and adding one or more second protectant proteins to a second portion of the biological sample to provide a second cryo-EM sample, wherein the one or more second protectant proteins comprises an amino acid sequence, and wherein the one or more second protectant proteins are different than the one or more first protectant proteins; and an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3. However, these shared technical features do not represent a contribution over the prior art.

WO 2018/073242 to Universitat Basel (hereinafter, "Basel") teaches a method for preparation of a cryogenic electron microscopy (cryo-EM) sample (a preparation method for preparing a sample for electron microscopy…cryogenic sample-grid "cryo-EM grid" preparation; Pg. 1, 1st – 2nd Paras.; preparation of cryo-grids; Pg. 7, 2nd Para.), comprising: adding one or more protectants to a biological sample during preparation of the cryo-EM sample (both deposition protocols also allow the use of additives such as traces of detergent… which might protect proteins from surface effect; Pg. 31, 2nd Para.), a composition for preparing a cryogenic electron microscopy (cryo-EM) sample, comprising: one or more protectants (a preparation method for preparing a sample for electron microscopy…cryogenic sample-grid "cryo-EM grid" preparation; Pg. 1, 1st – 2nd Paras.; preparation of cryo-grids; Pg. 7, 2nd Para.; both deposition protocols also allow the use of additives such as traces of detergent…which might protect proteins from surface effect; Pg. 31, 2nd Para.), a method for performing cryogenic electron microscopy (cryo-EM) (a preparation method for preparing a sample for electron microscopy…cryogenic sample-grid "cryo-EM grid" preparation; Pg. 1, 1st – 2nd Paras.; preparation of cryo-grids; Pg. 7, 2nd Para.), comprising: adding one or more first protectants to a first portion of a biological sample to provide a first cryo-EM sample (both deposition protocols also allow the use of additives such as traces of detergent…which might protect proteins from surface effect; Pg. 31, 2nd Para.).

US 2010/0216111 to Schenk (hereinafter, "Schenk") teaches one or more protectant proteins (some kinds of protein may also have a cryoprotectant function; Para. [0007]), wherein the one or more protectant proteins comprises an amino acid sequence (some kinds of protein may also have a cryoprotectant function; Para. [0007]; where a protein has an amino acid sequence); one or more second protectants (some kinds of protein may also have a cryoprotectant function; Para. [0007]; cryoprotectants also may by glycerol; Para. [0008]), and wherein the one or more second protectants are different than the one or more first protectant proteins (some kinds of protein may also have a cryoprotectant function; Para. [0007]; cryoprotectants also may by glycerol; Para. [0008]).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

US 2018/0335438 to Amprion, Inc. (hereinafter, "Amprion") teaches adding one or more second proteins to a second portion of the biological sample to provide a second cryo-EM sample (contacting a first portion of the sample with a first substrate protein; Para. [0005]; contacting a second portion of the sample with a second substrate protein; Para. [0090]; the methods include…cryo-electron microscopy; Para. [0106]), wherein the one or more second proteins comprises an amino acid sequence (a second substrate protein; Para. [0090]; where a protein has an amino acid sequence), and wherein the one or more second proteins are different than the one or more first proteins (contacting a first portion of the sample with a first substrate protein; Para. [0005]; contacting a second portion of the sample with a second substrate protein; Para. [0090]; the second substrate protein may be distinct from each other substrate protein; Para. [0094]).

US 2018/0155732 to Wisconsin Alumni Research Foundation (hereinafter, "Wisconsin") teaches a construct comprising a promoter operably connected to a polynucleotide encoding a polypeptide (the construct include a polynucleotide encoding a thymidine kinase operably connected to a promoter; abstract).

US 2006/0059583 to Kikawada et al. (hereinafter, "Kikawada") teaches an amino acid sequence at least 85% identical to SEQ ID NO: 1, SEQ ID NO: 2, or SEQ ID NO: 3 (Arabidopsis thaliana are shown in SEQ ID NOs: 7 to 11; Para. [0025]; wherein SEQ ID NO: 7 is 95% identical to Applicant's SEQ ID NO: 1).

The inventions listed in Groups I+ and II+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☑ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-3**

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.