Factors affecting macromolecule orientations in thin films formed in cryo-EM

Swati Yadav [a] and Kutti R. Vinothkumar [a] *

[a] National Centre for Biological Sciences, Tata Institute of Fundamental Research

GKVK Post, Bellary Road, Bangalore, 560065.

Correspondence – vkumar@ncbs.res.in

Abstract: The formation of vitrified thin film embedded with randomly oriented macromolecules is an essential prerequisite for cryogenic sample electron microscopy. Most commonly, this is achieved using the plunge freeze method described nearly 40 years ago. Although this is a robust method, the behaviour of different macromolecules shows great variation upon freezing and often needs to be optimized to obtain an isotropic, high-resolution reconstruction. For a macromolecule in such a film, the probability of encountering the air-water interface in the time between blotting and freezing and adopting preferred orientations is very high. 3D reconstruction using preferentially oriented particles often leads to anisotropic and uninterpretable maps. Currently, there are no general solutions to this prevalent issue, but several approaches largely focusing on sample preparation with the use of additives and novel grid modifications have been attempted. In this study, through an analysis of selected well-studied macromolecules, the effect of physical and chemical factors on the orientations of macromolecules was investigated, and important parameters that determine the behaviour of proteins on cryo-EM grids were revealed. These insights highlight the nature of interactions that cause preferred orientations and can be utilized to systematically address orientation bias for any given macromolecule and provide a framework to design small-molecule additives to enhance sample stability and behaviour.

Keywords: cryo-EM; thin films; preferred orientation; surfactants; temperature

1. Introduction

Single particle cryo-EM has become an indispensable technique in structural biology owing to its relative ease of image acquisition, reconstruction, and structure determination. Consequently, there has been a steady increase in the number of structures deposited in the Electron Microscopy Database (EMDB) in the past few years (Patwardhan, 2017). Near atomic resolution maps of biological macromolecules can now be routinely obtained due to improved hardware and specimen preparation methods along with developments in algorithms for image processing (Bai, 2021; Cheng, 2015; Lyumkis, 2019; Nogales, 2015; Vinothkumar & Henderson, 2016; Zheng et al., 2023). Obtaining a thin film of purified macromolecules in a vitreous layer of ice transparent to electrons is an important first step in cryo-EM structure determination. Currently, the plunge freeze method is the most commonly used technique for specimen preparation. This method involves placing a drop of the sample on a plasma-cleaned metal grid (typically coated with carbon and has an array of holes) on which a thin film is formed as the excess fluid is blotted away using blotting paper (Dubochet et al., 1982; Dubochet & McDowall, 1981). The grid is then plunged into a cryogen, causing the vitrification of water and preserving the native structures of macromolecules (Dubochet et al., 1985; Dubochet & McDowall, 1981; Taylor & Glaeser, 1974).

The freezing and blotting conditions must be extensively optimised for temperature and humidity to reproducibly create a single layer of macromolecules on a thin film of ice suitable for imaging. Because of their large size, the first test specimens used to optimise these conditions were liposomes and viruses. They lose their structural integrity if the osmolarity of the buffer changes owing to slow cooling (Adrian et al., 1984; Dubochet et al., 1988). It has been shown through recent studies that during freezing, very often macromolecules can interact with the air-water interface (AWI). This interaction can cause the macromolecules to denature, dissociate, adapt preferential orientations, or avoid the holes completely and stick to the carbon support (D'Imprima et al., 2019; Drulyte et al., 2018; Kampjut et al., 2021; Liu & Wang, 2023; Noble, Dandey, et al., 2018; Noble, Wei, et al., 2018). Along with protein biochemistry, the propensity of macromolecules to adopt preferential orientations is perhaps a common rate-limiting step in high-resolution structure determination.

Depending on the instrumental setup, the time between blotting and freezing can vary, but it is typically around a second. During this period, the macromolecules are suspended in an extremely thin film and constantly tumble in a solution, undergoing Brownian motion. An

intact macromolecule undergoing diffusion can encounter and get trapped in specific orientations at the freshly formed AWI. This effect is protein-dependent and more severe in some cases than in others (Naydenova & Russo, 2017).

When and how does having preferred orientations become a problem? There are cases where the preferred orientation exists but is not a major problem, such as in proteasomes and viruses, because of their highly symmetric architecture (Campbell et al., 2015; Vogel et al., 1986). A characteristic feature of the maps reconstructed from particles with orientation bias is the stretching of density in the cryo-EM map in one direction (anisotropy) (Figure 1), which makes the maps uninterpretable (Sorzano et al., 2022). In such cases, the resolution estimate based on the comparison of the half-maps is unrealistic, and the features in the map do not justify the resolution. Two examples of macromolecules exhibiting orientation bias are shown in Figure 1. It should be noted that 2D projections of different views of a macromolecule contribute differently to the final reconstruction, and this is an important criterion for obtaining a reconstruction with isotropic resolution, as shown by (Tan et al., 2018) and is further demonstrated using catalase as an example in Figure S1.

Several approaches have been proposed to address the preferred orientation problem (Drulyte et al., 2018). One such approach is to change the sample preparation process by skipping the blotting with filter papers altogether and minimising the time between thin film creation and freezing to as little as 100 ms. This method uses pL to nL quantities of the sample, which is sprayed onto a self-wicking grid that contains nanowires to absorb the excess liquid, creating a thin film, followed by rapid freezing of the grid. This protocol has been shown to introduce views in some proteins, however, even in these cases, the timescale of 100 ms has been large enough to allow the formation of denatured protein films on the surface (Dandey et al., 2020; Jain et al., 2012; Wei et al., 2018). Other options include using support films, such as carbon and functionalized graphene, or tilting the stage to obtain alternate views (Aiyer et al., 2023; Lu et al., 2022; Noble, Wei, et al., 2018; Tan et al., 2018; Xu et al., 2021). Tilting the stage or use of carbon support film suffers from a decrease in signal/noise ratio, especially if the protein is small, as the contrast is decreased due to either the support film or thick ice while imaging at higher tilt angles and tilting is less commonly used. While carbon support films are more commonly used in samples such as ribosomes, viruses, and some ion channels (Baker et al., 2021; Hesketh et al., 2018; Tobiasson et al., 2022), graphene/graphene oxide is suitable for low-molecular-weight specimens. These techniques have been successful in alleviating orientation bias in a few cases but are not

generally applicable. Additionally, there is now some image processing software that deals with the orientation bias in the data better than others because of the better weighing of the different projections during reconstruction, but the problem still persists (Ramírez-Aportela et al., 2022; Sorzano et al., 2021).

The most widely used approach in decreasing the preferred orientation of macromolecules is to add surfactants to the sample buffer prior to grid preparation (Chen et al., 2019; B. Li et al., 2021). This method is popular because it is simple and does not require any additional steps, techniques, or instruments. A large number of surfactants can be used, but care must be taken to avoid disrupting the protein's native structure in the process. Often, extensive trial and error is involved when surfactants are screened to overcome the preferential orientation problem. EMDB has at least 83 different soluble protein structures (retrieved in 2022), in which a detergent is added to the buffer. The most popular choice of detergent is non-ionic or zwitterionic. Recently, cationic or anionic detergents have been found to work for a few samples but their use is not generalized yet (B. Li et al., 2021). We also encountered the preferred orientation problem in many of our projects. We asked if an informed decision on the choice of a surfactant or any other small molecule can be made based on the properties of the protein, thus minimising the time spent screening many surfactants and specimen preparation conditions.

To achieve this goal, we tested some commonly used surfactants with different properties on a set of five proteins with varying surface charge distribution, shape, size, oligomeric status, and symmetry. In addition, we explored the effect of the presence of the histidine tag and physical factors, such as the temperature during the sample application step and the hole dimensions of the holey carbon grid, on the orientation distribution of the macromolecule. Through this analysis, we identified factors that determine macromolecule behaviour on grids prior to freezing and studied their effects with a focus on the preferred orientation problem. This account provides valuable information that can assist in achieving the optimal freezing conditions for any given macromolecule.

2. Results

2.1. Analysis of preferred views of selected macromolecules

A few well-characterized proteins with varying molecular weights, shapes, symmetries, surface charge distributions, and known experimental structures were selected, in order to rule out any artefacts that may arise owing to the changing conditions of sample

preparation. The test proteins used for the analysis include C-reactive protein (CRP) pentamer, CRP-decamer, human erythrocyte catalase, SARS-CoV-2 spike protein, and *E. coli.* PaaZ. Micrographs of these datasets, their respective 2D class averages showing the preferred views, anisotropic reconstructions, and orientation plots collected at the start of the project are shown in Figure 2.

CRP exists in a concentration-dependent equilibrium of pentamer and decamer of a 25 kDa polypeptide under physiological conditions (Okemefuna et al., 2010). At the concentration range used for cryo-EM experiments, both pentamer and decamer populations were observed in the same micrograph and analyzed separately (Figure 2A). C5 symmetry was applied to the smaller pentameric molecule, and no symmetry was applied to the decamer during image processing. The preferred view in the CRP-pentamer is the top/bottom view, where the symmetric arrangement of the monomers can be seen (Figure 2B). Both the negatively charged B-face and the neutrally charged A-face are sampled, but the A-face is more abundant (Figure S2B). Some side views are observed, but no tilted side views are sampled. In contrast, side views are predominantly observed in the CRP-decamer. Catalase exists as a dimer of tetramers and possesses D2 symmetry (Ko et al., 2000). It adopts a preferred top view on the grids when the grids are prepared at 4°C using Quantifoil 0.6/1 holey carbon grids (Figure 2C). This enzyme has previously been noted for its peculiar behaviour during cryo-EM grid preparation and has been used as a test specimen to overcome orientation bias, and we used it as a control sample and to probe additional parameters (S.Chen et al., 2022; Fan et al., 2021; Vinothkumar & Henderson, 2016). PaaZ is a bifunctional enzyme and consists of six polypeptides, where a domain-swapped dimer assembles to form a trimeric structure. In an earlier study, when frozen on ice, the enzyme was found to form clumps, and aggregates, which was overcome by the use of graphene oxide and a low concentration of protein (Sathyanarayanan et al., 2019). PaaZ adopts a preferred side view in the absence of an additive when frozen on ice at 4 °C, with a tendency to clump or cluster (Figure 2D). Nevertheless, the isolated particles were sufficient to obtain reasonable maps, and here we were interested in screening parameters to check if some of the aggregation and clumping could be overcome to enhance the quality of the data collected on ice. The SARS-CoV-2 spike protein is now a well-studied trimeric viral membrane protein whose symmetry can vary depending on the conformation of the receptor-binding domain (Huang et al., 2020; Wrapp et al., 2020). Soluble spike protein secreted into the medium by HEK293F cells shows a preference for a bottom view on grids in the absence of an additive

(Figure 2E). Spike trimer has three receptor binding domains (RBD), one per monomer, and they can adopt different conformational states and thus dictate the symmetry of the protein. In this data set, we observed that there was heterogeneity in the RBD domain, and hence C1 symmetry was applied for the rest of the analysis.

2.2. Effect of surfactant headgroup charge on protein orientations

We tested a specific set of surface-active molecules with varying head group charges (cationic, anionic, zwitterionic, and non-ionic), chain length packing (unsaturated vs. saturated alkyl chain), concentration, and critical micelle concentration (CMC) on the set of proteins described above. All the surfactants were used at concentrations lower than their respective CMCs. The properties of the surfactants are listed in Table 1.

Table 1 Properties of surfactants used in this study * - CMC values are from the Anatrace handbook

Additive	Charge	Ionic or non-ionic	CMC*	Concentration used	Aggregation number	Molecular weight	Alkyl chain length	Saturation in alkyl chain
CTAB	Positive ++	Ionic	1mM (0.027%)	0.54mM (0.002%)	75-120	364	15	Saturated
SLS	Negative --	Ionic	14.57mM (0.4%)	1.38mM (0.04%)	2	293	12	Saturated
Tween20	Neutral	Non-ionic	0.05mM (0.06%)	0.004mM (0.005%)	80	1225	11	Saturated
Tween80	Neutral	Non-ionic	0.015mM (0.01 %)	0.003mM (0.005%)	22-350	1310	17	Unsaturated
A8-35	Neutral -+	Zwitterionic	NA	0.01%	NA	NA	NA	NA

2.2.1. Effect of addition of charged surfactants

Upon addition of positive surfactant cetyltrimethylammonium bromide (CTAB), a change in the orientation distribution was observed for all the samples tested. An improvement in coverage of the orientations of CRP-pentamer and decamer was observed, resulting in isotropic 3D reconstructions. The initial no-additive dataset of catalase was collected from a specimen prepared at 4°C on a Quantifoil 0.6/1 grid, which has a severe

preference for the top view (Figure 2C). However, as the grids for all the additive datasets of catalase were made at 20°C, for the current comparison, specimen preparation was carried out at 20°C on a Quantifoil 1.2/1.3 grid. This dataset shows a preference for a tilted side view, annotated as side view A (Figure S4). These changes in orientation distributions of catalase are a result of variations in the incubation temperature and grid hole geometry, which we did not anticipate to have such a large effect on the quality of the reconstruction and were explored further in this study (discussed below). Upon addition of CTAB to catalase grids, the preference for the tilted side view A was lost, instead the particles were oriented towards the other tilted side view B (Figure S4), and a map of similar quality to the no additive dataset was obtained. In the case of PaaZ, the preference for the side view was maintained even in the presence of CTAB, but additional side-tilted views were sampled, which led to a better resolution and higher-quality map. The addition of a negatively charged detergent sodium lauryl sarcosine (SLS), resulted in evenly distributed orientations, with no preference for any particular view for all cases tested (it was also tested for PaaZ but no data was collected as visually it looked similar to CTAB as additive).

To understand the importance of the discrete views that were sampled and their contribution to the quality of the map, the efficiency of the orientation distribution (E_{od}) using CryoEF (Naydenova & Russo, 2017) and the sphericity of the final maps using the 3DFSC (Tan et al., 2018) server were calculated. E_{od} provides an estimate of the coverage of the 3D Fourier space by the 2D projections in the data, whereas sphericity provides a measure of anisotropy in the final maps by estimating the directional resolution. The highest E_{od}, sphericity values, and resolution for the given protein are highlighted in Figures 3&4. Among all the datasets, the CTAB datasets stand out and result in higher-resolution maps. We think this could be a result of optimum ice thickness, along with a sampling of the orientations that contribute more to the reconstruction (Figure S1).

2.2.2. Effect of addition of neutral surfactants

The addition of non-ionic detergents Tween 20 and Tween 80 to CRP-pentamer, decamer, and catalase led to evenly distributed orientations, but PaaZ showed aggregation on grids (data not shown). Tween 80 was tested as it is known to form denser layers at AWI compared to Tween 20 due to differences in the alkyl chain (length and unsaturation) (Szymczyk et al., 2018). Additionally, it has been shown to differ in its ability to separate protein films from the AWI as compared to Tween 20 (Rabe et al., 2020). In the

macromolecules tested in our study, no significant difference was observed between the orientation distributions obtained from either of the Tween surfactant datasets (Figure 4). However, we note that with CRP-pentamer, the orientation distributions look similar with both Tween 20 and 80 but only the addition of Tween 20 led to an isotropic higher-resolution map (R=3.3 Å), while the addition of Tween 80 led to a low-resolution map (R=7.5 Å). Amphipol A8-35, a surface-active zwitterionic polymer that does not form micelles and is thus different from other surfactants was also tested to see its effect on orientations. The orientation distributions obtained in this case were similar to those of the Tween 20 and Tween 80 datasets. The observed results with the given test samples indicate that the charge on the detergent headgroup is an important parameter, whereas the chain length and saturation of the hydrophobic chain have very little effect in modulating the orientation distribution of the macromolecules.

2.3. Effect of poly-histidine tag on protein orientations

Several cryo-EM structures of SARS-CoV-2 spike proteins have been reported in the last three years (Bodakuntla et al., 2023; Hardenbrook & Zhang, 2022; Harvey et al., 2021; Wrobel, 2023). We also purified the spike protein and faced severe orientation bias. Despite several attempts using different surfactants, we were unable to obtain an isotropic map of the spike protein (Figure 1A). The bias was more severe compared to other proteins in this study. One major difference is the presence of a poly-histidine tag at its C-terminus that is used for purification. We reasoned that the tag could be playing an important role in this bias.

Thus, the affinity tag from the spike protein was removed by protease cleavage, which led to a change in orientation distribution. Although the orientation bias still remained, the preference shifted from the bottom view (where the histidine tag is attached) to the top view (Figure 5). Subsequently, CTAB was added to the sample, imaging was performed on thin ice, and side-tilted views were obtained, which resulted in a reasonable isotropic map. The final reconstruction was obtained with 261,703 particles to an overall resolution of 3 Å (Figure 5). At this resolution, amino acid side chains and glycosylation on the protein surface were clearly visible in the ordered regions. In comparison to the other datasets, the increase in E_{od} and sphericity is evident in this case (Figure 5). To verify that the improvement in map quality is a result of new orientations and not an increase in the total number of particles, all the top/bottom views were removed for 3D refinement, and the reconstruction using only the side/tilted views corresponding to ~94,000 particles was performed. This resulted in a map

with a sphericity of 0.88 and a resolution of 3.4 Å (Figure S3) similar to that with all the particles. This analysis illustrates that the top views in this case contribute little information to the final reconstruction.

2.4. Effect of grid freezing temperature on protein orientations

The temperature at which the grids are held and blotted is critical for determining ice thickness, protein stability, and dynamics. Hence, grids were prepared at different temperatures to determine the effect of temperature on the orientation distributions. Typically, lower temperatures are used during sample application to grids to keep the macromolecules stable (except in some cases, such as microtubules, where a higher temperature is a prerequisite), and the temperature as a parameter is not very often screened to overcome the issue of preferred orientations. For catalase, the orientation distributions varied significantly when grids were prepared at different temperatures. The coverage of the 3D Fourier space (E_{od}) changes from 0.64 at 4°C to 0.72 at 20°C to 0.77 at 37°C (Figure 6A). However, the resolution of the final reconstruction is best for the 4°C dataset, which could be due to the higher number of particles used for reconstruction, that is, 478,656 compared to 138,031 for the 20°C dataset and 75,312 for the 37°C dataset. The respective B factors (from Relion post-process) for the reconstructions were -107 $Å^2$, -110 $Å^2$, and -126 $Å^2$. Interestingly, when the grid type was varied from Quantifoil 1.2/1.3 to Quantifoil 0.6/1 with catalase, the orientations changed drastically, as shown in Figure S4.

In the case of PaaZ, when the grids were prepared at 4°C, 20°C and 37°C, increasing higher-order structures were observed, particularly at 37°C. At 4°C, higher-order structures were minimal, but a preference for the side view was observed. Data collected at 4°C and 20°C have similar E_{od}, and the orientation distribution plots show a slight increase in the sampling of additional side-tilted views at 20°C, indicated by the reduction in empty spaces in the orientation plots obtained from Relion (Figure 6B). However, clumping of the particles was still observed at 20 °C. Hence, CTAB was added to the sample and the grids were prepared at 4°C, which led to less clumping and consequently resulted in a high-resolution map at 2.3 Å, described further in the next section.

It is evident from these observations that physical factors, such as the grid preparation temperature and choice of grid type, can affect protein behaviour and should be considered as important screening conditions when dealing with orientation bias along with surfactants.

2.5. High-resolution cryo-EM of PaaZ on ice

To understand the mechanism of action of a macromolecule, high-resolution maps are required to build an accurate model and more importantly, to model the solvent molecules and other ligands of interest. We previously reported the structure of PaaZ, a bifunctional enzyme from *E. coli* that catalyzes two steps of the phenylacetic acid degradation (paa) pathway. In the previous study, the structure of PaaZ was determined at 2.9 Å using graphene-oxide support grids (Sathyanarayanan et al., 2019). During that study, it was observed that the enzyme tended to clump when frozen on ice in the absence of any additive, and the use of a graphene oxide support and a low concentration of enzyme yielded a very good distribution of particles. In the current study, we obtained a high-resolution map using CTAB as an additive during grid preparation, with D3 symmetry and higher-order aberration correction during image processing. The resolution estimated by comparison of the half maps FSC (at 0.143 threshold) is 2.3 Å, and a comparable resolution of 2.4 Å is obtained with the comparison of the map vs. model FSC (at 0.5 threshold). The model-fitted PaaZ structure is shown in Figure 7B. Several water molecules could be modelled at this resolution and an electrostatic potential map of the dimer with water molecules modelled as spheres and coloured in cyan is shown (Figure 7C). To further analyse the quality of the data, the Reslog plot was calculated by obtaining a 3D reconstruction using different numbers of particles and plotting the resolution ($1/d^2$) obtained with the number of particles (lnN) used (Rosenthal & Henderson, 2003). From the experimental data, it can be seen that as the resolution approaches the Nyquist limit of 2.1, the addition of more particles from 40,000 to 80,000 does not lead to an improvement in resolution. The slope of this graph was used to calculate a B-factor of 69.4 $Å^2$, which is slightly lower than the B-factor estimated from the post-processing option in Relion, which was used for map sharpening (-75 $Å^2$). A 200-particle dataset was used as a base, and the particles required to obtain a certain resolution were calculated theoretically. These theoretical estimates are in agreement with the experimental values.

3. Discussion

In this era of artificial intelligence-based protein structure prediction (Baek et al., 2021; Jumper et al., 2021; Schauperl & Denny, 2022), there is still no substitute for the joy one feels when the molecule of interest is imaged and observed for the first time after hard work of expression and purification. However, the first data collection for a new sample often

does not result in a high-resolution structure, as there can be one or more challenges that need to be overcome. Of these, the preferred orientation of macromolecules and the possibility of denaturation due to exposure to AWI is a serious issue and one of the major bottlenecks. It is now known that the tendency of macromolecules to adopt preferential orientation when frozen on ice using cryo-EM grids is a consequence of protein-AWI interaction (J.Chen et al., 2019; Noble, Wei, et al., 2018), but the nature of this interaction remains unclear. On the other hand, amphipathic molecules have been used in cryo-EM grid preparation since the 1980s (Frederik et al., 1989). Their use in solving the preferred orientation of proteins by modulating the air-water interface has recently been appreciated in cryo-EM and is now being used more routinely for high-resolution structure determination (J.Chen et al., 2019; S. Chen et al., 2022; Li et al., 2021). At a fundamental level, several open questions remain: what causes proteins to adapt to certain preferred views? How do surfactants affect this behaviour? What is the role of other physical properties such as ice thickness, surface tension, and temperature in the behaviour of macromolecules during freezing? To address some of these questions, we systematically tested grid freezing conditions on a set of well-characterised proteins that adopt preferred orientations on cryo-EM grids. The selected proteins had molecular weights ranging from 125 kDa to 450 kDa and varying symmetries (Figure 2). The degree of orientation bias also varied significantly. The CRP-decamer and SARS-CoV-2 spike had a more severe orientation bias, with more than 90% abundance for the preferred view, whereas in the case of the CRP-pentamer, the abundance was 50%, and in catalase, it was near 60%. Additionally, both PaaZ and spike that were recombinantly overexpressed had poly-histidine as the affinity tag at the termini of each monomer, whereas CRP and catalase, which were purified from native sources, did not contain any affinity tags.

The air-water interface exhibits an affinity for OH^- ions from water, resulting in a negative charge (Chaplin, 2009; Drzymala et al., 1999). Extensive research has been conducted on the behaviour of proteins and surfactants at the air-water interface and in thin films, spanning the fields of cryo-EM and surface chemistry (Li et al., 2021; Rabe et al., 2020; Samanta & Ghosh, 2011; Szymczyk et al., 2018). Based on these investigations and an examination of the surface charge distributions of macromolecules in our study, we postulated that electrostatic interactions may play a pivotal role in dictating protein orientations. To visualise the surface charge distributions of the proteins, electrostatic potential maps were calculated by assigning charges using the pdb2pqr server (Jurrus et al., 2018) at a given pH and visualised using ChimeraX (Pettersen et al., 2021). It was observed

that the charge on the preferred view was positive in the case of catalase and neutral in CRP-pentamer, CRP-decamer, and PaaZ (Figure S2). The preferred bottom view of the spike protein is negatively charged, but this charge is likely masked by the presence of a 56 amino acid long affinity tag from each monomer. The negatively charged B-face of CRP in the absence of additive is (shown in Figure S2) less sampled in the case of the CRP-pentamer and not sampled at all in CRP-decamer. When the surface charge at the AWI was changed to positive by the addition of cationic surfactant CTAB, the orientations changed in almost all cases, with a slight preference for certain alternate views (Figure 3). Hence, electrostatic interactions at the AWI are an important factor in determining protein orientations in the absence of any additive. When non-ionic and zwitterionic surfactants are added to the sample buffer, the observed orientation distributions are uniform. Among the non-ionic surfactants, the orientation plots are similar, as they all make the AWI charge neutral by occupying the free surface and work by diminishing the charge-based interactions between the protein and the AWI.

Although CTAB addition resulted in isotropic maps, the coverage of the 3D Fourier space was the least when compared with other surfactants (Table in Figure 3). The CTAB datasets showed a preference for one or more views, but in the macromolecules tested, these preferences were desirable. Upon addition of SLS, an even orientation distribution was observed for the CRP-pentamer, CRP-decamer, and catalase. Tween 20 and Tween 80 had similar effects on orientation, except in the case of the CRP-pentamer. A8-35 addition to the sample buffer led to an intermediate orientation distribution, which resulted in isotropic maps (Figure 4). In the order of homogeneity of the orientation distributions, the surfactants can be ranked as SLS > Tween 20 = Tween 80 > A8-35 = CTAB. The map vs. model FSCs were calculated for the datasets to obtain an independent resolution estimate, and they closely match the resolution estimated by the comparison of the half-maps (Figure S5). Further, the use of different surfactants or sample preparation at different temperatures does not affect the structures of macromolecules as shown by the density of heme in catalase (Figure S6).

Poly-histidine tag-dependent changes in orientation have been reported recently by Bromberg et al., 2022, but the mechanism by which this occurs remains unclear. The addition of a poly-histidine tag is a convenient method for expressing and isolating proteins (Hochuli et al., 1988; Merchant et al., 1998). Unlike crystallography, the flexibility of histidine tags does not pose a problem in structure determination by cryo-EM as the molecules are imaged individually, and the tag is averaged out during the reconstruction process if it is present in

random conformations. At pH 8, which is more often used as the sample buffer, histidine should be deprotonated and exist as a hydrophilic polar group, and why this should cause severe preference in orientations is not clear to us. The microenvironment at the AWI is complex and could be interacting with the histidine tag in a favourable manner and stabilizing it in that orientation.

Among the samples studied here, PaaZ and SARS-CoV2 spike had a poly-histidine tag at their N and C-termini, respectively. While the presence of the tag had no detrimental effect on the orientation distribution and reconstruction of PaaZ, it affected the orientations of the spike protein significantly and the addition of surfactants was not sufficient to mitigate the bias. As the spike protein has been extensively studied and many structures have been reported we wondered what could be the reason that we see such severe orientation bias and others do not. From the published structural studies on the SARS-CoV2 spike it was realised that there have been variations in the reported specimens used for cryo-EM in terms of the expression system, detergents added to the sample, cleavage of tags, buffers, etc (Bangaru et al., 2020; Bodakuntla et al., 2023; Hardenbrook & Zhang, 2022; Wrapp et al., 2020; Wrobel, 2023). Other parameters that might affect the orientation distribution include the glow discharge settings for grids that change the hydrophilicity of the grid surface, grid type (for example, Quantifoil with carbon support or Ultrafoil with gold support), the hole diameter (0.6/1, 1.2/1.3, or 2/2) in the grids, freezing temperature, and blotting duration. In our case, the removal of the affinity tag and CTAB addition to the sample buffer was required to sample alternate views of the spike protein, which resulted in a map of reasonable resolution and improved isotropy (Figure 5). Thus, the effect of the poly-histidine tag is a key factor to remember when faced with orientation bias.

As stated above, physical factors, such as the temperature of grid freezing and grid type, can affect the mechanics of thin film formation and evaporation/blotting and, in turn, influence protein behaviour. We show that the temperature of grid freezing can change the orientation distributions of proteins by affecting the ice thickness and protein-protein interactions. Additionally, varying only the hole diameter in a holey grid and keeping other factors constant led to a change in the preference of views in the case of catalase (Figure S4). In Quantifoil 1.2/1.3 grids, the approximate ratio of the area occupied by the hole to carbon is 38:62%, whereas, in 0.6/1 grids, the ratio decreases to 30:70%. This change can affect protein distribution in the holes, carbon, and AWI. It is now evident that this also significantly affects the orientation distribution of macromolecules in the thin films formed in the holes. We note

two caveats in this study: a) all the grids used were from Quantifoil with amorphous carbon as the support film, and it is unclear if similar behaviour will be observed with C-flat grids from Protochips that are manufactured differently or other grids such as UltraFoil or NiTi alloy (Fan et al., 2021; Quispe et al., 2007; Russo & Passmore, 2016; Schürmann et al., 2017). b) only one sample (catalase) was tested and further extensive study with more samples is required to generalise the effect of the holey carbon geometry.

Previous studies have used surfactants to determine the structures of the proteins examined in this study. A comparison of these results with our observations is presented below. Catalase has been used as a standard test sample in the context of the preferred orientation problem in two independent studies, which take different approaches to solve the problem. In an earlier study, a number of high-CMC detergents were screened, among which CHAPSO resulted in the best orientation distribution and map quality, although a high protein concentration of 30-40 mg/ml was used to obtain a good distribution of particles and a small data set was sufficient to obtain a 2.2 Å reconstruction by averaging 119,000 particles with an E_{od} of 0.76 (S.Chen et al., 2022). Furthermore, the effects of ionic strength and pH were tested and were observed to have little or no effect on the protein orientation (S.Chen et al., 2022). Our results, in extension to this study, provide an understanding of the effects of surfactant headgroup charge and chain length, temperature of grid freezing, and hole carbon geometry on the orientations of catalase (Figure S4). In another study, nickel/titanium grids coated with 2D crystals of the hydrophobic protein HFBI (which shields the protein from its exposure to AWI) and 2.3 mg/ml of catalase were used for preparing specimens and subsequent data collection and structure determination (Fan et al., 2021). The resulting catalase map has a resolution of 2.3 Å from 169,897 particles with an E_{od} of 0.80, which is comparable to our datasets (Figure 6A).

Similarly, multiple structures of CRP-pentamer and decamer were reported in a study by Noone et al. (2021), which focused on the effects of pH and ligand addition on the complement binding properties of CRP. One of the grid freezing conditions reported was CRP apo protein at pH 7.5 with 0.05% Tween 20 added and the grid held at 4 °C and 65% humidity. These freezing conditions are similar to those described here, except for the concentration of Tween 20 used and the temperature and humidity during freezing. The orientations observed in this case look similar, and the resulting reconstructions from 256,289 and 204,354 particles for the pentamer and decamer (C5 and C1 symmetry) resulted in resolutions of 3.2 Å and 2.8 Å, respectively. In comparison, we have obtained a 3.3 Å map

for the pentamer and a 4 Å map for the decamer by averaging 25,000 and 51,000 particles, respectively (Figure 4). In the field of cryo-EM, the biggest variation is observed in sample preparation, sometimes using the same instrument. Therefore, it is reassuring to observe that other independent investigations (along with different grid freezing procedures) of catalase and CRP yield 3D reconstructions of comparable quality.

The air–water interface is a complex environment; the properties of the bulk and the surface vary drastically, and it is an active area of research in surface chemistry and aerosol fields (Martins-Costa & Ruiz-López, 2023; Nguyen et al., 2020; Zhong et al., 2019). Only the effects that occur at the surface can be observed, and trends can be analysed to improve the sample preparation methods. The explanation as to why they occur cannot really be understood using standard cryo-EM experiments and requires complementary approaches such as infrared spectroscopy or dynamic surface tension measurement studies (Carter-Fenk et al., 2021; Tang et al., 2020). Alternative approaches to sample preparation have been proposed in recent years to overcome the preferred orientation issue (Drulyte et al., 2018; Glaeser & Han, 2017; Jain et al., 2012). One such approach was recently reported by Huber et al., where the sample was filled with nanosized capillaries made of silicon-rich nitride membranes embedded on a chip that physically controls the ice thickness. No blotting is involved in this process, and some test samples have demonstrated the potential of this method (Huber et al., 2022). However, the air-water interface is replaced by the solid-water interface, which may contribute to some effects of its own. Further use and standardisation of such methods are required to understand their general applicability.

4. Conclusion and outlook

One of the goals of single particle cryo-EM is to collect a small dataset and average minimum number of particles to obtain a high-resolution reconstruction (Henderson, 1995). However, in reality, many factors affect this including the sample heterogeneity, detectors, beam-induced motion etc., and the recently observed effect of AWI and preferred orientation can be added to this list. We performed a comprehensive examination of how macromolecule orientations respond to alterations in physical factors, such as freezing temperature, and chemical factors, such as the addition of surfactants or the presence of affinity tags, during grid freezing for a standard set of proteins. This analysis provides insights into the behaviour of proteins on grids and can be utilised to address the preferred orientation problem in a systemic manner for any given macromolecule. When using surfactants, it is crucial to

carefully assess protein stability using other techniques such as native gel electrophoresis, differential scanning fluorimetry, or size-exclusion chromatography before data collection to save time and resources. Furthermore, our findings highlight the necessity to innovate and create small-molecule additives that are inert to biological samples and can effectively occupy the AWI and alleviate its deleterious effects on proteins.

5. Materials and Methods

Human C-reactive protein (Cat. No. C4063) and Human erythrocyte catalase (Cat. No. C3556) were obtained from Sigma Aldrich chemicals. The protein samples were either concentrated using an Amicon 100kDa concentrator or diluted in respective buffers for grid freezing. All the detergent stocks were made in ultrapure water and dilutions were made and used on the day of the experiment.

PaaZ was expressed and purified as described in (Sathyanarayanan et al., 2019).

The original plasmid encoding the SARS-CoV-2 spike gene was a kind gift from the Krammer laboratory at Icahn School of Medicine at Mount Sinai. The spike gene was amplified from this plasmid and subcloned in the BacMam vector with a C-terminus HRV-3C cleavage tag followed by a 7x poly-histidine and twin strep tag. Bacmid DNA and virus were prepared as described in the Invitrogen bac-to-bac manual. After two generations of amplification in Sf9 cells, the V2 virus was used to infect HEK293F cells at 2 million/ml density. Sodium butyrate (4 mM) was added to enhance the production of protein 8 hours post-infection. The media supernatant containing the secreted spike protein was harvested on day 3 by centrifuging the cells at 150 g for 10 minutes. The media was incubated with pre-equilibrated Ni-NTA beads at room temperature for 1-2 hours (1 ml beads/200 ml media). Ni-NTA beads were washed with a PBS buffer containing 20 mM imidazole followed by elution with 280 mM imidazole in PBS buffer. The eluted protein was run on SDS-PAGE to assess purity and further concentrated and injected on a 24 ml column S200 size exclusion to exchange the buffer to 50 mM Tris pH 8, 200 mM NaCl,1 mM DTT. To cleave the tag, the eluted fractions from Ni-NTA chromatography were diluted with 50 mM Tris pH 8, 200 mM NaCl, and 1 mM DTT buffer and incubated with HRV3C protease overnight at 4°C, followed by reverse IMAC to obtain the spike protein without tag in the flow through. The flow-through was concentrated using an Amicon 100kDa concentrator and flash-frozen using liquid nitrogen and stored at -80°C until further use.

Grid preparation:

6.3 µl of the protein was thawed on ice and 0.7 µl of 10x additive (surfactant) stock was added to get a final concentration of 1x. This sample was incubated on ice for 2-5 mins and then centrifuged at 21,000g for 20 mins. Meanwhile, Vitrobot Mark IV (ThermoFisher Scientific) chamber was equilibrated at 20 °C (unless stated otherwise) and 100% humidity. Quantifoil 1.2/1.3 or Quantifoil 0.6/1 grids were glow discharged in a reduced air environment using a PELCO EasiGlow chamber using a standard setting of 25 mA current for 1 minute. The grid was mounted on the Vitrobot Mark IV (ThermoFisher Scientific) and 3 µl of the sample was applied to the grid. A blotting time of 3-4 secs, a wait time of 10 secs and a blot force of 0 was used to obtain a thin film of the specimen. For the datasets where grids were prepared at different temperatures, the protein was incubated on a thermal block at the required temperature for 3-7 mins before applying it to the grid. The Vitrobot chamber was maintained at the required temperature and 100% humidity. The blot time, blot force and wait time were kept constant.

Grid screening and data collection:

Grids were screened using standard low-dose settings and the automated data collection was set up either on the Falcon3 or Gatan K2 detectors in counting mode with EPU software (ThermoFisher Scientific). Magnification of 59,000 x was used only for the catalase 37 °C dataset with a pixel size of 1.38 Å. For all the other datasets collected on Falcon3, a magnification of 75,000 x corresponding to a pixel size of 1.07 Å and a dose of approximately 1 $e^-/Å^2$/frame were used and movies of 24-25 frames were collected. The datasets on the K2 detector were collected in EFTEM mode with a slit width of 20 eV and at a nominal magnification of 130,000 x corresponding to a pixel size of 1.05 Å. The dose was 5 e^-/p/s when an exposure of 8 seconds was used. Each movie was fractionated into 32 frames with a total dose of ~36.28 $e^-/Å^2$. The CTAB catalase, Tween 20 catalase, Tween 20 CRP, Tween 80 CRP and SLS CRP datasets were collected with the K2 detector.

Data processing:

RELION 3.1. (Scheres, 2012, Zivanov et al., 2018) was used to process all the data with a standard workflow which is summarized as follows. The multi-frame movies were summed and corrected for beam-induced motion using Relion's in-built algorithm. The resulting summed micrograph images were used to estimate the contrast transfer function (CTF) using GCTF (Zhang, 2016). The micrographs after CTF correction were used to pick particles

using either reference-free methods such as the LoG-based picking in Relion or Gautomatch (https://www2.mrc-lmb.cam.ac.uk/download/gautomatch-053/) or using previously made 2D templates. After picking, the particles were boxed and extracted using the appropriate box size of typically 256 or 320 pixels. These particles were then subjected to multiple rounds of iterative 2D classifications to obtain good-quality, high-resolution classes. A low-resolution initial model was either generated from the same dataset or one from a pre-existing dataset was used as a reference for 3D refinement. The selected good-quality particles were then used for either 3D classification or 3D refinement based on the data (Scheres, 2012). If the resolution at this stage was below 4.5 Å, CTF refinement and Bayesian polishing were performed on the final particles to correct for beam tilt, anisotropic magnification, per-particle defocus and the effects of beam-induced motion respectively (Zivanov et al., 2018, 2019). Another round of 3D refinement was performed after this step. Further, postprocessing was done to sharpen the final map and estimate the final global resolution (Rosenthal & Henderson, 2003).

PaaZ-CTAB and spike-CTAB datasets were subjected to multiple rounds of 3D classification. For most datasets (all except CRP-pentamer), as a part of standard workflow the resulting map after reconstruction had the wrong hand and the maps were flipped to obtain the correct hand before the model was fit and refinement was performed. The 3D FSC server was used to calculate the sphericity of the final maps (Tan et al., 2018). The E_{od} was calculated using the cryoEF by using a subset of 1000 angles and the appropriate symmetries for respective molecules (Naydenova & Russo, 2017). ChimeraX (Pettersen et al., 2004) was used for generating the electrostatic potential maps of the structures. The Pymol APBS plugin was used to generate the PaaZ electrostatic potential map (Schrödinger, LLC, 2015). EMAN2 (Tang et al., 2007) and Phenix (Liebschner et al., 2019) were used to generate the map vs. model FSCs. The final model refinement was done in real space for the PaaZ and CRP datasets using Phenix (Afonine et al., 2018) and in reciprocal space using Refmac Servalcat (Murshudov et al., 2011, Yamashita et al., 2021) for catalase and spike datasets. Initial water picking in the PaaZ dataset was performed with Coot (Emsley & Cowtan, 2004) and manually inspected to remove those that were not hydrogen bonded to amino acid residues, also using the fofc omit map from servalcat (Yamashita et al., 2021) and a conservative approach was used so that noise was not modelled. Figures were made with Chimera and Pymol (Pettersen et al., 2004; Schrödinger, LLC, 2015).

Acknowledgements: We acknowledge the National Cryo-EM facility, Bangalore, for data collection, which was supported by the Department of Biotechnology, DBT/PR12422/MED/31/287/2014, and the computing facility in the Bangalore Life Science Cluster. We thank Prof. Ramaswamy S and all the lab members for critical reading of the manuscript. SY acknowledges the graduate fellowship from TIFR/NCBS. K.R.V. acknowledges the support of the Department of Atomic Energy, Government of India, Government of India, under Project Identification No. RTI4006. KRV is part of the EMBO Global Investigator Network.

Conflict of interest: The authors declare no conflict of interest.

Data Availability: The cryo-EM maps and the coordinates have been deposited in the EMDB and PDB respectively with the following accession codes.
CRP-pentamer with CTAB – EMD-37864 and PDB 8WV4
CRP-decamer with CTAB – EMD-37865 and PDB ID 8WV5
PaaZ with CTAB at 4 °C – EMD-37866 and PDB ID 8WV6
Spike with CTAB – EMD-37953 and PDB ID 8WZI
Catalase with CTAB at 20 °C – EMD-37956 and PDB ID 8WZM
Catalase with SLS at 20 °C – EMD-37952 and PDB ID 8WZH
Catalase at 4 °C – EMD-37955 and PDB ID 8WZK
Catalase at 20 °C – EMD-37954 and PDB ID 8WZJ

References

Adrian, M., Dubochet, J., Lepault, J., & McDowall, A. W. (1984). Cryo-electron microscopy of viruses. *Nature*, *308*(5954), 32–36. https://doi.org/10.1038/308032a0

Aiyer, S., Baldwin, P. R., Tan, S. M., Shan, Z., Oh, J., Mehrani, A., Bowman, M. E., Louie, G., Passos, D. O., Selena, Đ., Hull, J. A., Hoshika, S., Barad, B. A., Grotjahn, D. A., Agbandje-mckenna, M., Benner, S. A., & Noel, J. A. P. (2023). *Overcoming Resolution Attenuation During Tilted Cryo-EM Data Collection*.

Bai, X. (2021). Seeing Atoms by Single-Particle Cryo-EM. *Trends in Biochemical Sciences*, *46*(4), 253–254. https://doi.org/https://doi.org/10.1016/j.tibs.2021.01.001

Bangaru, S., Ozorowski, G., Turner, H. L., Antanasijevic, A., Huang, D., Wang, X., Torres, J. L., Diedrich, J. K., Tian, J. H., Portnoff, A. D., Patel, N., Massare, M. J., Yates, J. R., Nemazee, D., Paulson, J. C., Glenn, G., Smith, G., & Ward, A. B. (2020). Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science*, *370*(6520), 1089–1094. https://doi.org/10.1126/science.abe1502

Bodakuntla, S., Kuhn, C. C., Biertümpfel, C., & Mizuno, N. (2023). Cryo-electron microscopy in the fight against COVID-19—mechanism of virus entry. *Frontiers in Molecular Biosciences*, *10*(October), 1–11. https://doi.org/10.3389/fmolb.2023.1252529

Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S., & Carragher, B. (2015). 2.8 Å Resolution Reconstruction of the Thermoplasma Acidophilum 20 S Proteasome Using Cryo-Electron Microscopy. *ELife*, *2015*(4), 1–22. https://doi.org/10.7554/eLife.06380

Carter-Fenk, K. A., Carter-Fenk, K., Fiamingo, M. E., Allen, H. C., & Herbert, J. M. (2021). Vibrational exciton delocalization precludes the use of infrared intensities as proxies for surfactant accumulation on aqueous surfaces. *Chem. Sci.*, *12*(24), 8320–8332. https://doi.org/10.1039/D1SC01276B

Chen, J., Noble, A. J., Kang, J. Y., & Darst, S. A. (2019). Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *Journal of Structural Biology: X*, *1*, 100005. https://doi.org/https://doi.org/10.1016/j.yjsbx.2019.100005

Cheng, Y. (2015). Single-Particle Cryo-EM at Crystallographic Resolution. *Cell*, *161*(3), 450–457. https://doi.org/10.1016/j.cell.2015.03.049

Dandey, V. P., Budell, W. C., Wei, H., Bobe, D., Maruthi, K., Kopylov, M., Eng, E. T., Kahn, P. A., Hinshaw, J. E., Kundu, N., Nimigean, C. M., Fan, C., Sukomon, N., Darst, S. A., Saecker, R. M., Chen, J., Malone, B., Potter, C. S., & Carragher, B. (2020). Time-resolved cryo-EM using Spotiton. *Nature Methods*, *17*(9), 897–900. https://doi.org/10.1038/s41592-020-0925-6

D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M., & Kühlbrandt, W. (2019). Protein denaturation at the air-water interface and how to prevent it. *ELife*, *8*, 1–18. https://doi.org/10.7554/eLife.42747

Drulyte, I., Johnson, R. M., Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P., & Thompson, R. F. (2018). Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallographica Section D: Structural Biology*, *74*(6), 560–571. https://doi.org/10.1107/S2059798318006496

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A. W., & Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, *21*(2), 129–228. https://doi.org/DOI: 10.1017/S0033583500004297

Dubochet, J., Adrian, M., Lepault, J., & McDowall, A. W. (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends in Biochemical Sciences*, *10*(4), 143–146. https://doi.org/10.1016/0968-0004(85)90150-1

Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A., & Homo, J. -C. (1982). Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy*, *128*(3), 219–237. https://doi.org/10.1111/j.1365-2818.1982.tb04625.x

Dubochet, J., & McDowall, A. W. (1981). Vitrification of Pure Water for Electron Microscopy. *Journal of Microscopy*, *124*(3), 3–4. https://doi.org/10.1111/j.1365-2818.1981.tb02483.x

Emsley, P., & Cowtan, K. (2004). {\it Coot}: model-building tools for molecular graphics. *Acta Crystallographica Section D*, *60*(12 Part 1), 2126–2132. https://doi.org/10.1107/S0907444904019158

Fan, H., Wang, B., Zhang, Y., Zhu, Y., Song, B., Xu, H., Zhai, Y., Qiao, M., & Sun, F. (2021). A cryo-electron microscopy support film formed by 2D crystals of hydrophobin HFBI. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-27596-8

Hardenbrook, N. J., & Zhang, P. (2022). A structural view of the SARS-CoV-2 virus and its assembly. *Current Opinion in Virology*, *52*, 123–134. https://doi.org/https://doi.org/10.1016/j.coviro.2021.11.011

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., Robertson, D. L., & Consortium, C.-19 G. U. K. (COG-U. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*(7), 409–424. https://doi.org/10.1038/s41579-021-00573-0

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics*, *28*(2), 171–193. https://doi.org/10.1017/S003358350000305X

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., & Stüber, D. (1988). Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio/Technology*, *6*(11), 1321–1325. https://doi.org/10.1038/nbt1188-1321

Jain, T., Sheehan, P., Crum, J., Carragher, B., & Potter, C. S. (2012). Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *Journal of Structural Biology*, *179*(1), 68–75. https://doi.org/10.1016/j.jsb.2012.04.020

Kampjut, D., Steiner, J., & Sazanov, L. A. (2021). Cryo-EM grid optimization for membrane proteins. *IScience*, *24*(3), 102139. https://doi.org/10.1016/j.isci.2021.102139

Ko, T.-P., Safo, M. K., Musayev, F. N., Di Salvo, M. L., Wang, C., Wu, S.-H., & Abraham, D. J. (2000). Structure of human erythrocyte catalase. *Acta Crystallographica Section D*, *56*(2), 241–245. https://doi.org/10.1107/S0907444999015930

Li, B., Zhu, D., Shi, H., & Zhang, X. (2021). Effect of charge on protein preferred orientation at the air–water interface in cryo-electron microscopy. *Journal of Structural Biology*, *213*(4), 107783. https://doi.org/10.1016/j.jsb.2021.107783

Liu, N., & Wang, H. W. (2023). Better Cryo-EM Specimen Preparation: How to Deal with the Air–Water Interface? *Journal of Molecular Biology*, *435*(9), 167926. https://doi.org/10.1016/j.jmb.2022.167926

Lu, Y., Liu, N., Liu, Y., Zheng, L., Yang, J., Wang, J., Jia, X., Zi, Q., Peng, H., Rao, Y., & Wang, H.-W. (2022). Functionalized graphene grids with various charges for single-particle cryo-EM. *Nature Communications*, *13*(1), 1–8. https://doi.org/10.1038/s41467-022-34579-w

Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. *Journal of Biological Chemistry*, *294*(13), 5181–5197. https://doi.org/10.1074/jbc.REV118.005602

Merchant, a M., Zhu, Z., Yuan, J. Q., Goddard, a, Adams, C. W., Presta, L. G., & Carter, P. (1998). 1998 Nature Publishing Group http://www.nature.com/naturebiotechnology. *Group*, *16*(10), 291–294. https://doi.org/10.1038/nbt0898-773

Naydenova, K., & Russo, C. J. (2017). Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature Communications*, *8*(1), 8–12. https://doi.org/10.1038/s41467-017-00782-3

Noble, A. J., Dandey, V. P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y. Z., Zhang, Z., Kim, L. Y., Scapin, G., Rapp, M., Eng, E. T., Rice, W. J., Cheng, A., Negro, C. J., Shapiro, L., Kwong, P. D., Jeruzalmi, D., Georges, A. des, … Carragher, B. (2018). Routine single particle CryoEM sample and grid characterization by tomography. *ELife*, *7*, 1–42. https://doi.org/10.7554/eLife.34257

Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S., & Carragher, B. (2018). Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nature Methods*, *15*(10), 793–795. https://doi.org/10.1038/s41592-018-0139-3

Nogales, E. (2015). The development of cryo-EM into a mainstream structural biology technique. *Nature Methods*, *13*(1), 24–27. https://doi.org/10.1038/nmeth.3694

Patwardhan, A. (2017). Trends in the Electron Microscopy Data Bank (EMDB). *Acta Crystallographica Section D: Structural Biology*, *73*(6), 503–508. https://doi.org/10.1107/S2059798317004181

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/https://doi.org/10.1002/pro.3943

Quispe, J., Damiano, J., Mick, S. E., Nackashi, D. P., Fellmann, D., Ajero, T. G., Carragher, B., & Potter, C. S. (2007). An Improved Holey Carbon Film for Cryo-Electron Microscopy. *Microscopy and Microanalysis*, *13*(5), 365–371. https://doi.org/10.1017/S1431927607070791

Ramírez-Aportela, E., Carazo, J. M., Sorzano, C. O. S., & Sun, F. (2022). Higher resolution in cryo-EM by the combination of macromolecular prior knowledge and image-processing tools. *IUCrJ*, *9*(Pt 5), 632–638. https://doi.org/10.1107/S2052252522006959

Rosenthal, P. B., & Henderson, R. (2003). Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology*, *333*(4), 721–745. https://doi.org/https://doi.org/10.1016/j.jmb.2003.07.013

Russo, C. J., & Passmore, L. A. (2016). Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. *Journal of Structural Biology*, *193*(1), 33–44. https://doi.org/https://doi.org/10.1016/j.jsb.2015.11.006

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Schrödinger, LLC. (2015). *The {AxPyMOL} Molecular Graphics Plugin for {Microsoft PowerPoint}, Version~1.8*.

Schürmann, U., Chluba, C., Wolff, N., Smazna, D., Lima de Miranda, R., Junker, P., Adelung, R., Quandt, E., & Kienle, L. (2017). Functional NiTi grids for in situ straining in the TEM. *Ultramicroscopy*, *182*, 10–16. https://doi.org/https://doi.org/10.1016/j.ultramic.2017.06.003

Sorzano, C. O. S., Jimenez-Moreno, A., Maluenda, D., Martinez, M., Ramirez-Aportela, E., Krieger, J., Melero, R., Cuervo, A., Conesa, J., Filipovic, J., Conesa, P., Del Cano, L., Fonseca, Y. C., Jiménez-De La Morena, J., Losana, P., Sanchez-Garcia, R., Strelak, D., Fernandez-Gimenez, E., De Isidro-Gómez, F. P., … Carazo, J. M. (2022). On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. *Acta Crystallographica Section D: Structural Biology*, *78*, 410–423. https://doi.org/10.1107/S2059798322001978

Sorzano, C. O. S., Semchonok, D., Lin, S. C., Lo, Y. C., Vilas, J. L., Jiménez-Moreno, A., Gragera, M., Vacca, S., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Melero, R., Cuervo, A., Conesa, J. J., Conesa, P., Losana, P., Caño, L. del, de la Morena, J. J., Fonseca, Y. C., … Carazo, J. M. (2021). Algorithmic robustness to preferred orientations in single particle analysis by CryoEM. *Journal of Structural Biology*, *213*(1). https://doi.org/10.1016/j.jsb.2020.107695

Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Clinton, S., Carragher, B., Lyumkis, D., Microscopy, S. E., York, N., Biology, S., Biophysics, M., Biology, C., Jolla, L., Studies, B., & Jolla, L. (2018). *Cryo-EM through Tilting*. *14*(8), 793–796. https://doi.org/10.1038/nmeth.4347.Addressing

Tang, F., Ohto, T., Sun, S., Rouxel, J. R., Imoto, S., Backus, E. H. G., Mukamel, S., Bonn, M., & Nagata, Y. (2020). Molecular Structure and Modeling of Water-Air and Ice-Air Interfaces Monitored by Sum-Frequency Generation. *Chemical Reviews*, *120*(8), 3633–3667. https://doi.org/10.1021/acs.chemrev.9b00512

Taylor, K. A., & Glaeser, R. M. (1974). Electron Diffraction of Frozen, Hydrated Protein Crystals. *Science*, *186*(4168), 1036–1037. https://doi.org/10.1126/science.186.4168.1036

Vinothkumar, K. R., & Henderson, R. (2016). Single particle electron cryomicroscopy: trends, issues and future perspective. *Quarterly Reviews of Biophysics*, *49*. https://doi.org/10.1017/s0033583516000068

Vogel, R. H., Provencher, S. W., Von Bonsdorff, C. H., Adrian, M., & Dubochet, J. (1986). Envelope structure of semliki forest virus reconstructed from cryo-electron micrographs. *Nature*, *320*(6062), 533–535. https://doi.org/10.1038/320533a0

Wei, H., Dandey, V. P., Zhang, Z., Raczkowski, A., Rice, W. J., Carragher, B., & Potter, C. S. (2018). Optimizing "self-wicking" nanowire grids. *Journal of Structural Biology*, *202*(2), 170–174. https://doi.org/10.1016/j.jsb.2018.01.001

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260–1263. https://doi.org/10.1126/science.aax0902

Wrobel, A. G. (2023). Mechanism and evolution of human ACE2 binding by SARS-CoV-2 spike. *Current Opinion in Structural Biology*, *81*, 102619. https://doi.org/https://doi.org/10.1016/j.sbi.2023.102619

Xu, J., Cui, X., Liu, N., Chen, Y., & Wang, H. (2021). Structural engineering of graphene for high-resolution cryo-electron microscopy. *SmartMat*, *2*(2), 202–212. https://doi.org/10.1002/smm2.1045

Yamashita, K., Palmer, C. M., Burnley, T., & Murshudov, G. N. (2021). Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta*

Crystallographica Section D: Structural Biology, *77*, 1282–1291. https://doi.org/10.1107/S2059798321009475

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, *193*(1), 1–12. https://doi.org/10.1016/j.jsb.2015.11.003

Zheng, L., Liu, N., Gao, X., Zhu, W., Liu, K., Wu, C., Yan, R., Zhang, J., Gao, X., Yao, Y., Deng, B., Xu, J., Lu, Y., Liu, Z., Li, M., Wei, X., Wang, H.-W., & Peng, H. (2023). Uniform thin ice on ultraflat graphene for high-resolution cryo-EM. *Nature Methods*, *20*(1), 123–130. https://doi.org/10.1038/s41592-022-01693-y

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., & Scheres, S. H. W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *ELife*, *7*, e42166. https://doi.org/10.7554/eLife.42166

Zivanov, J., Nakane, T., & Scheres, S. H. W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*, *6*(1), 5–17. https://doi.org/10.1107/S205225251801463X

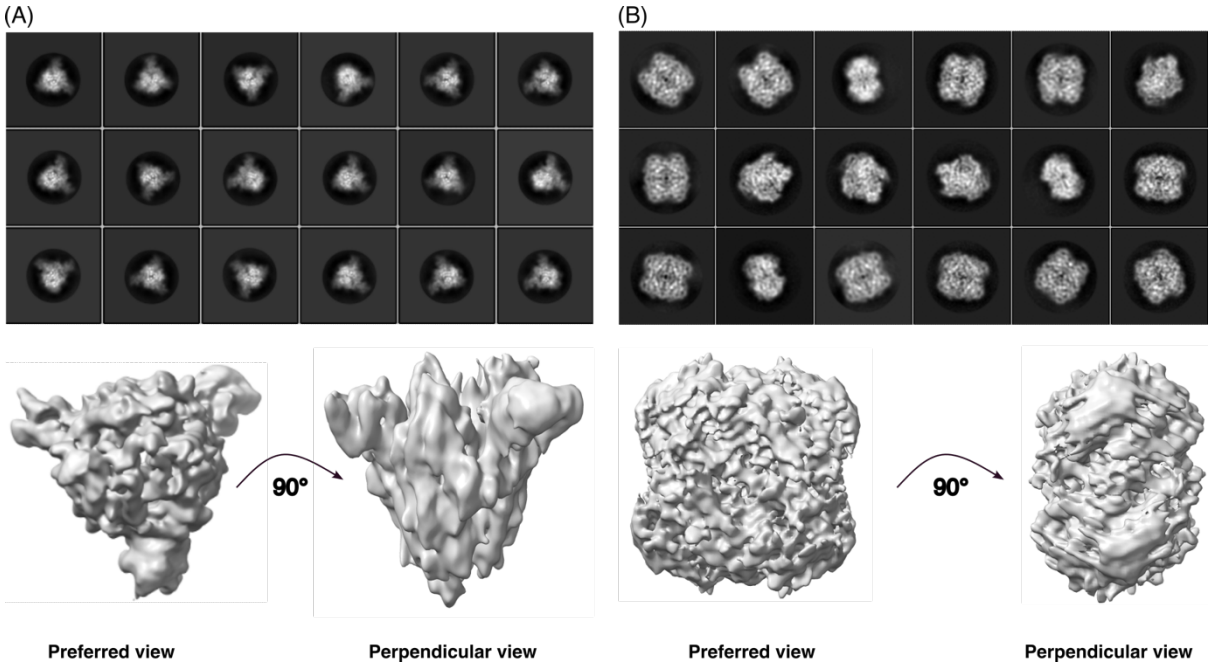


Figure 1: Examples of anisotropic cryo-EM maps resulting from orientation bias. Reference-free 2D class averages of SARS-CoV-2 spike protein (A) and human erythrocyte catalase (B) are shown in the upper panel. For the spike protein, the preferred bottom views are clearly seen. In the case of catalase, preference for the top/bottom view is evident. In the lower panel, 3D maps with anisotropic features are shown with the preferred and perpendicular views as labelled. The symmetries applied during reconstruction were C1 and D2 for the spike protein and catalase, respectively.

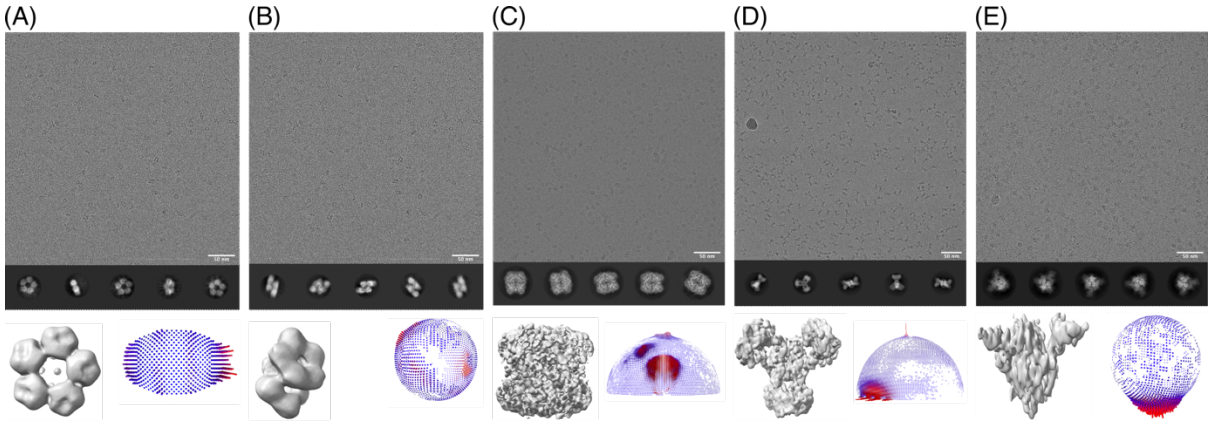
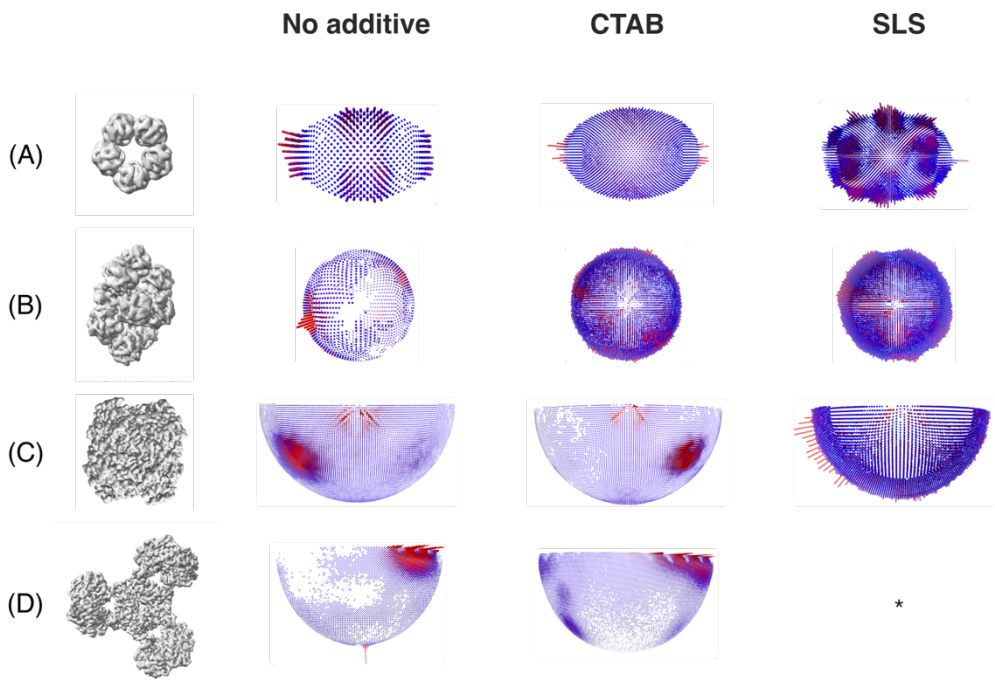
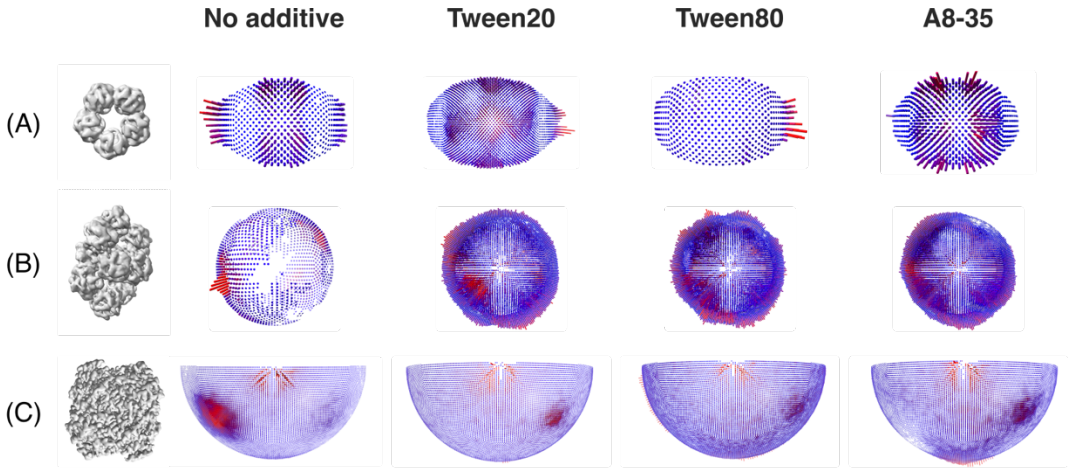


Figure 2: Representative micrographs, 2D class averages, 3D reconstructions, and orientation distribution plots of test proteins used in this study. (A) CRP-pentamer (B) CRP-decamer (C) catalase 4°C (D) PaaZ (E) spike. The orientation distribution plots obtained from Relion exhibit one preferred view, as indicated by the clustering of the red bars in the plot. Empty spaces indicate regions with missing projections.



Protein	Condition	No. of particles	Resolution(Å)	Efficiency	Sphericity
CRP pentamer	No additive	14,601	18	0.78	NA
	CTAB	36,353	3.3	0.80	0.98
	SLS	31,699	4.2	0.80	0.97
CRP decamer	No additive	9,419	20	0.52	NA
	CTAB	25,992	3.5	0.85	0.98
	SLS	59,211	3.7	0.79	0.97
Catalase	No additive	1,38,000	2.7	0.72	0.96
	CTAB	1,53,336	2.8	0.76	0.97
	SLS	33,241	3.7	0.80	0.98
PaaZ	No additive	51,393	4.0	0.76	0.8
	CTAB	89,454	2.3	0.75	0.98

Figure 3: Orientation distribution plots of different proteins upon addition of charged surfactants. (A) CRP-pentamer (B) CRP-decamer (C) catalase (D) PaaZ. The table indicates the different parameters calculated to interpret the quality of the datasets. NA = not available because the maps were of low resolution and sphericity values could not be calculated reliably. The effect of SLS on PaaZ was also tested, but visual inspection of the micrographs showed little difference and no data was collected and thus not included in the table.



Protein	Condition	No. of particles	Resolution(Å)	Efficiency	Sphericity
CRP pentamer	No additive	14,601	18	0.78	NA
	Tween 20	25,674	3.3	0.80	0.86
	Tween 80	32,330	7.5	0.69	NA
	A8-35	26,737	10	0.78	NA
CRP decamer	No additive	9,419	20	0.52	NA
	Tween 20	51,784	4.0	0.78	0.92
	Tween 80	36,870	4.2	0.79	0.76
	A8-35	104,369	3.5	0.78	0.98
Catalase	No additive	1,10,000	2.7	0.72	0.96
	Tween 20	88,395	2.9	0.78	0.98
	Tween 80	92,163	2.9	0.80	0.96
	A8-35	1,22,000	3.1	0.77	0.98

Figure 4: Orientation distribution plots of different proteins upon addition of non-ionic and zwitterionic surfactants (A) CRP-pentamer (B) CRP-decamer (C) catalase. The table summarizes the different parameters calculated to interpret and compare the quality of the datasets.

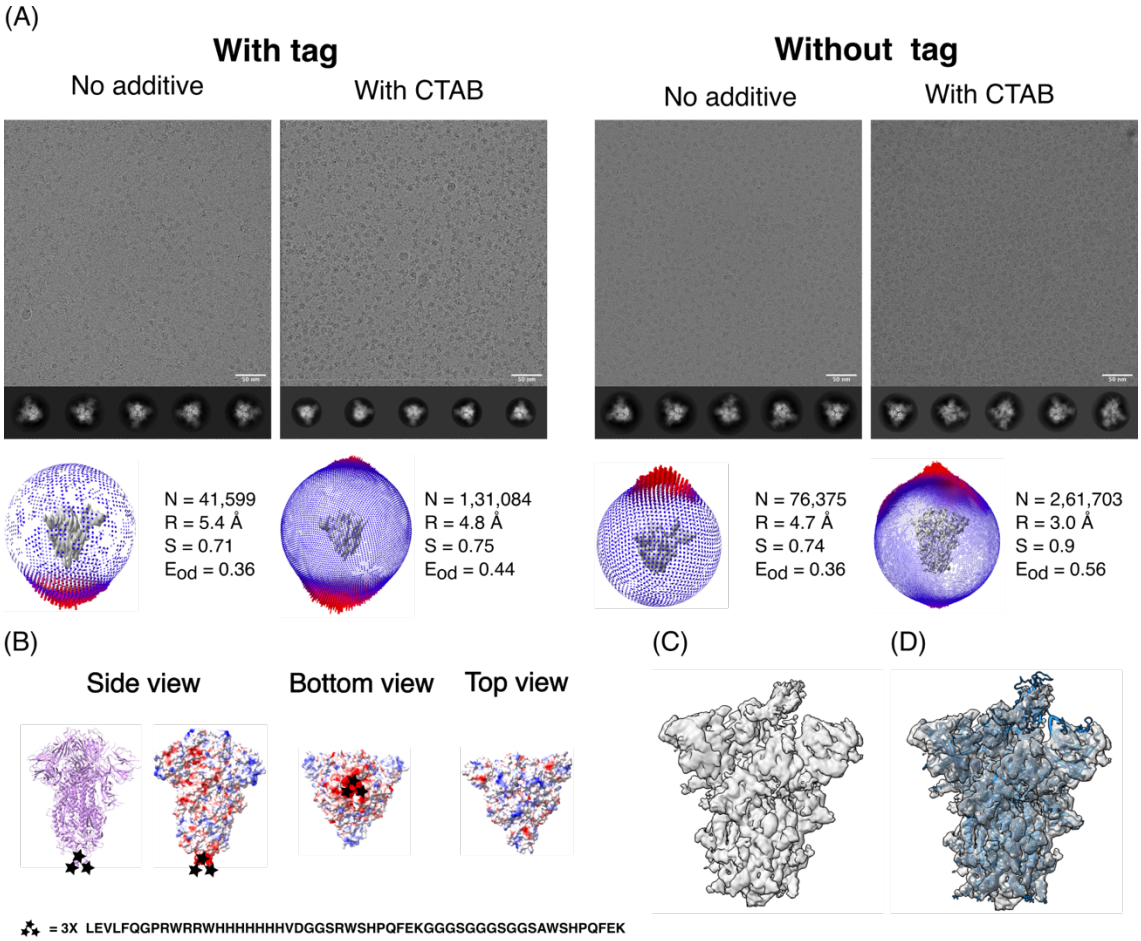


Figure 5: Effect of tag on the SARS-CoV2 spike protein orientation distribution. (A) Representative micrographs, 2D class averages, and orientation distribution plots of spike proteins under different conditions as indicated. Different parameters that are used to analyse the quality of the maps are shown next to the orientation plots. N = no. of particles used for reconstruction, R = resolution, S = sphericity, E_{od} = efficiency. (B) The electrostatic surface representation of the spike protein and the model where the location of the affinity tag is marked by an asterisk and the surface representation is shown in different views. (C) Cryo-EM map of spike protein at an overall resolution of 3 Å. Refinement and reconstruction were performed using C1 symmetry. (D) SARS-CoV-2 spike model with PDB ID-8d56 fitted into the EM map after the deletion of the C-terminal membrane helix region.

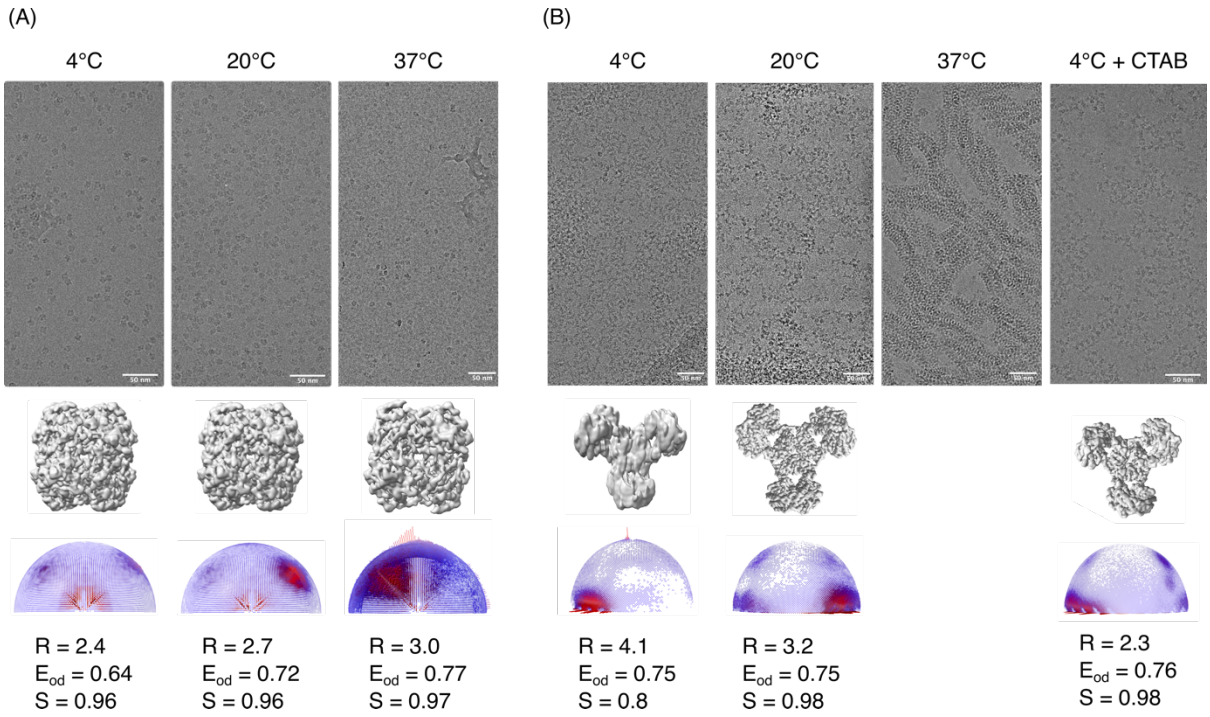
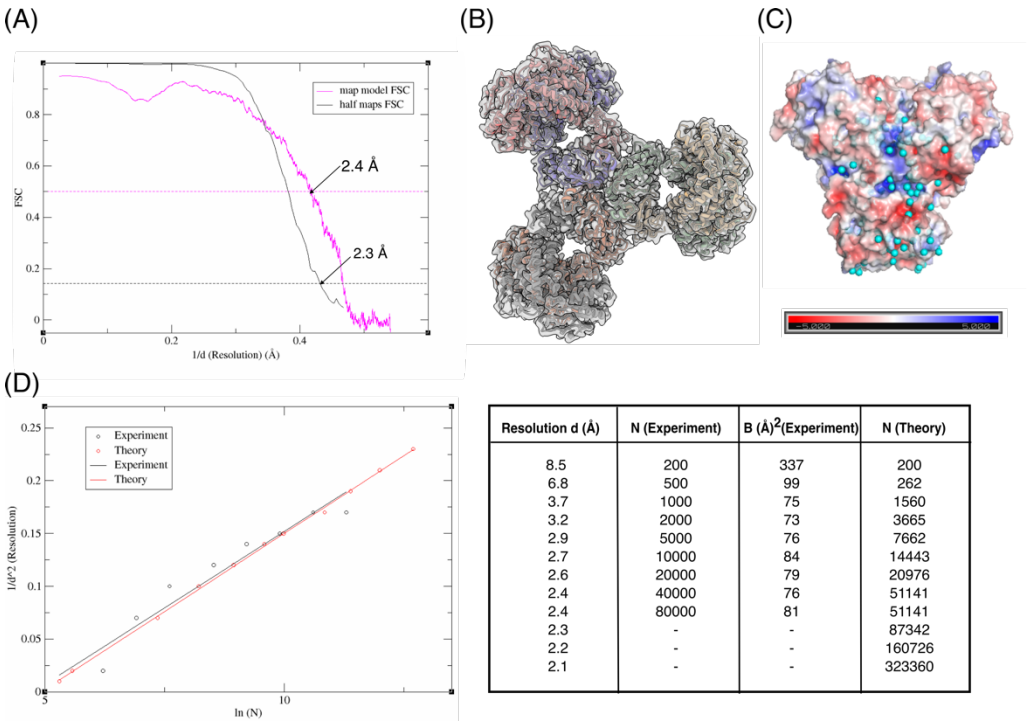


Figure 6: Effect of temperature during cryo-EM sample preparation of catalase and PaaZ. (A) Micrographs, maps, and orientation distribution plots of catalase grids held and blotted at different temperatures in the absence of additive. (B) Micrographs, maps, and orientation distribution plots of PaaZ grids held and blotted at different temperatures along with 4°C + CTAB condition included for comparison. R = resolution, S = sphericity, E_{od} = efficiency



Resolution d (Å)	N (Experiment)	B (Å)^2(Experiment)	N (Theory)
8.5	200	337	200
6.8	500	99	262
3.7	1000	75	1560
3.2	2000	73	3665
2.9	5000	76	7662
2.7	10000	84	14443
2.6	20000	79	20976
2.4	40000	76	51141
2.4	80000	81	51141
2.3	-	-	87342
2.2	-	-	160726
2.1	-	-	323360

Figure 7: High-resolution Cryo-EM map of PaaZ (A) Comparison of the half map and map versus model FSCs of the PaaZ dataset. (B) The six polypeptides coloured individually and in cartoon representation fit into the cryo-EM map (transparent grey surface) of PaaZ. (C) Electrostatic potential surface representation of the domain-swapped PaaZ dimer with modelled waters shown as cyan spheres. (D) Reslog plot of PaaZ with the experimental and theoretical numbers of particles required to reach a particular resolution. N indicates the number of particles used for reconstruction, d is the resolution, and B indicates the B factor, as estimated by the Relion post-processing. D3 symmetry was applied for the reconstruction and the Reslog plot indicates the number of particles used, not the asymmetric units averaged.