

## Research Article

## How to correct relative voxel scale factors for calculations of vector-difference Fourier maps in cryo-EM

Jimin Wang<sup>a,\*</sup>, Jinchuan Liu<sup>a</sup>, Christopher J. Gisriel<sup>b</sup>, Shenping Wu<sup>c</sup>, Federica Maschietto<sup>b</sup>, David A. Flesher<sup>a</sup>, Elias Lolis<sup>c</sup>, George P. Lisi<sup>d</sup>, Gary W. Brudvig<sup>a,b</sup>, Yong Xiong<sup>a</sup>, Victor S. Batista<sup>b</sup>

<sup>a</sup> Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06536, USA

<sup>b</sup> Department of Chemistry, Yale University, New Haven, CT 06511-8499, USA

<sup>c</sup> Department of Pharmacology, Yale University, New Haven, CT 06520-8066, USA

<sup>d</sup> Department of Molecular and Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA

## ARTICLE INFO

## Keywords:

Cryo-EM maps  
Voxel Scaling  
Absolute EM Magnification  
Volumetric Expansion/Contraction Coefficients  
Center of Mass  
Monomeric Apo-Photosystem II  
Spike Protein  
SARS-CoV-2  
pH-Dependent Structural Transition

## ABSTRACT

The atomic coordinates derived from cryo-electron microscopy (cryo-EM) maps can be inaccurate when the voxel scaling factors are not properly calibrated. Here, we describe a method for correcting relative voxel scaling factors between pairs of cryo-EM maps for the same or similar structures that are expanded or contracted relative to each other. We find that the correction of scaling factors reduces the amplitude differences of Fourier-inverted structure factors from voxel-rescaled maps by up to 20–30%, as shown by two cryo-EM maps of the SARS-CoV-2 spike protein measured at pH 4.0 and pH 8.0. This allows for the calculation of the difference map after properly scaling, revealing differences between the two structures for individual amino acid residues. Unexpectedly, the analysis uncovers two previously overlooked differences of amino acid residues in structures and their local structural changes. Furthermore, we demonstrate the method as applied to two cryo-EM maps of monomeric apo-photosystem II from the cyanobacteria *Synechocystis* sp. PCC 6803 and *Thermosynechococcus elongatus*. The resulting difference maps reveal many changes in the peripheral transmembrane PsbX subunit between the two species.

## 1. Introduction

The comparison of experimentally determined electrostatic potentials (ESP) or electron density (ED) maps of macromolecules can provide valuable biological insights at the molecular level. For X-ray diffraction data, the isomorphous difference method can be applied to reveal subtle structural differences for comparison of pairs of structures with great precision. This is because the X-ray data can be placed onto an absolute coordinate scale with very small errors (Moore, 2012). The energy of X-ray photons as well as the sample-to-detector distances are typically known or can be accurately calibrated. In principle, the absolute magnification for cryo-EM images (i.e., voxel scale factors) can also be accurately calibrated (Frank, 2018; Glaeser, 2019; Glaeser et al., 2021). However, errors in the absolute magnification, which affect the coordinate scale or voxel scale of map reconstruction, have largely been ignored, despite extensive efforts made to correct for anisotropic

magnification (Zhao et al., 2015; Grant and Grigorieff, 2015; Zivanov et al., 2020). As a consequence, two cryo-EM maps being compared could have large coordinate-scale differences that would overwhelmingly obscure biologically important differences between them (Joseph, 2020). Here, we introduce a procedure for removal of artifactual differences due to errors in absolute magnification. This procedure corrects for the relative voxel scale factors of the two cryo-EM maps being compared so that important underlying structural differences can be revealed.

The overall features of macromolecules are relatively resilient to small errors in the voxel scaling factors used for map reconstruction. However, those errors can have a significant impact on the accuracy of inter-atomic distances and hence the interpretation of biological states. Voxel scale factors are especially important for redox active metal ions in structures of metalloenzymes. This is because minor but vital map characteristics must be assessed when comparing two structures to

\* Corresponding author.

E-mail address: [jimin.wang@yale.edu](mailto:jimin.wang@yale.edu) (J. Wang).

<https://doi.org/10.1016/j.jysbi.2022.107902>

Received 16 June 2022; Received in revised form 23 August 2022; Accepted 28 September 2022

Available online 3 October 2022

1047-8477/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

determine changes in coordination bond lengths associated with oxidation state transitions of redox-active cofactors. To demonstrate this point, we compare two cryo-EM maps of photosystem II (PSII), the conserved water-splitting metalloenzyme involved in oxygenic photosynthesis whose mechanism of assembly and repair is central to its function (Vinyard and Brudvig, 2017; Vinyard et al., 2013; Cox et al., 2020; Bao and Burnap, 2016; Heinz et al., 2016; Nixon et al., 2010). We focus on the comparison of the monomeric apo-PSII structures from the organisms *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803) and *Thermosynechococcus elongatus* (Gisriel, 2020; Zabret, 2021). We find that they have a relative voxel scale factor difference of 2.4 %, making the comparison of the two structures by difference Fourier maps nearly impossible.

The functionality of most macromolecules depends upon changes of protonation states in pH-sensing residues, where the addition or removal of a proton can be readily visible in cryo-EM maps but not directly detectable by X-ray crystallography. Cryo-EM can provide insights into the molecular origin of protonation state transitions before the structural changes take place. The comparative analysis of cryo-EM maps requires structures to be on the same relative voxel scale. To demonstrate this point, we compare two cryo-EM maps of the trimeric spike protein of SARS-CoV-2 determined at pH 8.0 and pH 4.0, respectively (Zhou, 2020; Mannar, 2021). We find that these two maps exhibit a voxel scale difference of 2.9 % that would completely obscure any subtle pH-dependent structural changes, or amino acid differences between them. However, it is known that the spike protein structure is sensitive to pH. In fact, the spike protein is responsible for the initiation of the host-viral membrane fusion process that is pH-dependent, critical for viral entry into the host cell (Li, 2016; Harvey, 2021). Therefore, it is imperative to establish methods that could enable comparisons of cryo-EM maps of the spike protein at different values of pH.

Furthermore, we compared two cryo-EM maps of the replication-transcription complex (RTC) of SARS-CoV2 that exhibit a voxel scale factor difference of 2.4 %. This difference is so large that it prevents accurate comparisons of the RTC complex that would be valuable for studies of the translocation process of the RNA duplex product upon RNA synthesis (Chen, 2020; Yin, 2020). In fact, without voxel scale corrections, the artificial expansion/contraction of the two models hinders the detection of any significant displacement of the RNA duplex. These limitations prevent the analysis of the dynamical process of nucleic acid synthesis through a pairwise comparison of accurate structures with substrates and products in different functional states (Wang and Konigsberg, 2022; Park, 2021).

## 2. Materials, analytical methods, and computational procedures

### 2.1. Cryo-EM maps and coordinates of reported structures

Cryo-EM maps and coordinates were retrieved from the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB), as follows: (i) emd12335/7nho, pH 6.5, monomeric apo-PSII from *T. elongatus* (Zabret, 2021), (ii) emd21690/6wj6, pH 7.0, monomeric apo-PSII from *Synechocystis* 6803 (Gisriel, 2020), (iii) emd22251/6xlu, pH 4.0, “2P” mutated SARS-CoV-2 spike protein (Zhou, 2020), (iv) emd25515/7jwy, pH 4.5, “2P” mutated spike protein (Zhou, 2020), (v) emd22505/7sxt, pH 8.0, “6P” mutated spike protein plus D614G and N501Y substitutions (Mannar, 2021), (vi) the pre-translocation RNA duplex within the replication and transcription complex (RTC) of SARS-CoV-2 reported for 7bv2, at pH 7.4, after incorporation of remdesivir monophosphate with pyrophosphate remaining bound (Yin, 2020), and (vii) the post-translocation RNA duplex within the RTC reported for 6xez, at pH 8.0, with a vacant nucleoside triphosphate (NTP) substrate-binding pocket (Chen, 2020).

The “2P” mutations of the spike protein are K987P and V989P. Unexpectedly, two additional substitutions of R1107A/Y904S in these “2P” mutated spike protein structures were identified in this study, which

were neither described in the publication nor present in the deposited coordinate file. The “6P” mutations are F817P, A892P, A899P, A942P, K987P, and V989P, of which the last two substitutions were incorrectly assigned in the publication as K968P and V969P (because they were neither present in the deposited coordinate files nor in the cited paper). For the analysis of the “2P” mutated spike protein, the entire structure was included. For the analysis of differences between the “6P” and “2P” mutated spike proteins, only the S2 fragment trimer was carved out for analysis. Carving was carried out by using the program Chimera (e.g., *color-zone* with *carve* radius of 2.0 Å) (Pettersen, 2004).

### 2.2. Linear volumetric expansion/contraction of Cryo-EM derived coordinates

Structural alignment of the two monomeric apo-PSII structures from *Synechocystis* 6803 and *T. elongatus* was based on Cα superpositions, as carried out using the program Coot (Gisriel, 2020; Zabret, 2021; Emsley and Cowtan, 2004). Equivalent residues were extracted, and Cα coordinate shifts were calculated and plotted as a function of Cα distances to the center-of-mass of the entire structures, using the open-source program X-motif grace (*xmgrace*) (xmgrace, 2021). The coordinate shifts projected on the radial axis (i.e., the vector from the center-of-mass to given Cα coordinates) were calculated and plotted. Other graphical figures were made by using the program PyMol (Delano).

### 2.3. Voxel rescaling, map alignment, and structure-factor amplitude scaling

The initial matrices for the alignment of two cryo-EM maps were generated from the corresponding coordinates, using the program Chimera (e.g., “*mm #2:A #3:A*”, where #2:A is reference coordinate chain A, #3:A is target coordinate chain A, and the initial matrix is saved in the header of model #3), and applied to the target map (e.g., “*matrixcopy #3 #1*”, where the #3 PDB header contains the initial matrix and #1 is the target map to be rotated) (Pettersen, 2004). The voxel scaling factor for the target map was systematically changed with a small increment and the two maps were least-squares fitted for selected target grids (“*fit-in-map*”). The resulting Pearson real-space correlation coefficient (CC) within a selected contouring level of the target map was plotted as a function of the voxel rescaling factor to find the best relative voxel rescaling correction factor with the maximal CC value. The target map with the best voxel rescaling correction was resampled on the same grid of the reference map (e.g., “*vop resample #1 OnGrid #0*”, where #1 is the target map with the rotation matrix in its header and #0 is the reference map) before the resampled map was written out (Pettersen, 2004). Both reference and rescaled target maps were Fourier-inverted using the program Phenix with the same parameter sets (Adams, 2010). The two sets of structure factors were concatenated (*cad*, see below for summary) and their amplitudes were scaled for analysis of differences between them as a function of resolution (*scaleit*) with various cut-offs, using the program suite CCP4 (Winn, 2011). Vector-difference Fourier maps with selected resolution ranges were calculated (*fft*) using the program suite CCP4 (Winn, 2011).

The map Fourier-inversion was also carried out using the program suite CCP4 using a command script with a fake “measurement error” column added (Adams, 2010; Winn, 2011). In summary, the steps involved in the entire computational procedure are: (i) *mapmask* to switch the map sampling axis (“axis z x y”), (ii) *sfall* to Fourier invert the map (e.g., “mode sfall mapin”, “symm p1”, “resolution 20 2.6” as an example), (iii) *mtz2various* to convert the mtz file to a text file so that a fake measurement error column can be added, (iv) *f2mtz* to convert the resulting text file back to the mtz file format, (v) *cad* to concatenate two mtz files, (vi) *scaleit* to scale their amplitude to the selected resolution range (e.g., “resolution 20 2.6”), and finally (vii) *fft* to calculate the vector-difference Fourier maps with a selected resolution (e.g., “resolution 20 2.6”, “labin F1 = F\_target, PHI = PHI\_target, F2 = F\_reference,



PHI2 = PHI\_reference"). Difference maps were visualized with a minimal attempt made to slightly fit coordinates into the corresponding maps for making proper figure illustrations using the graphics programs Coot and PyMol (Emsley and Cowtan, 2004; Delano).

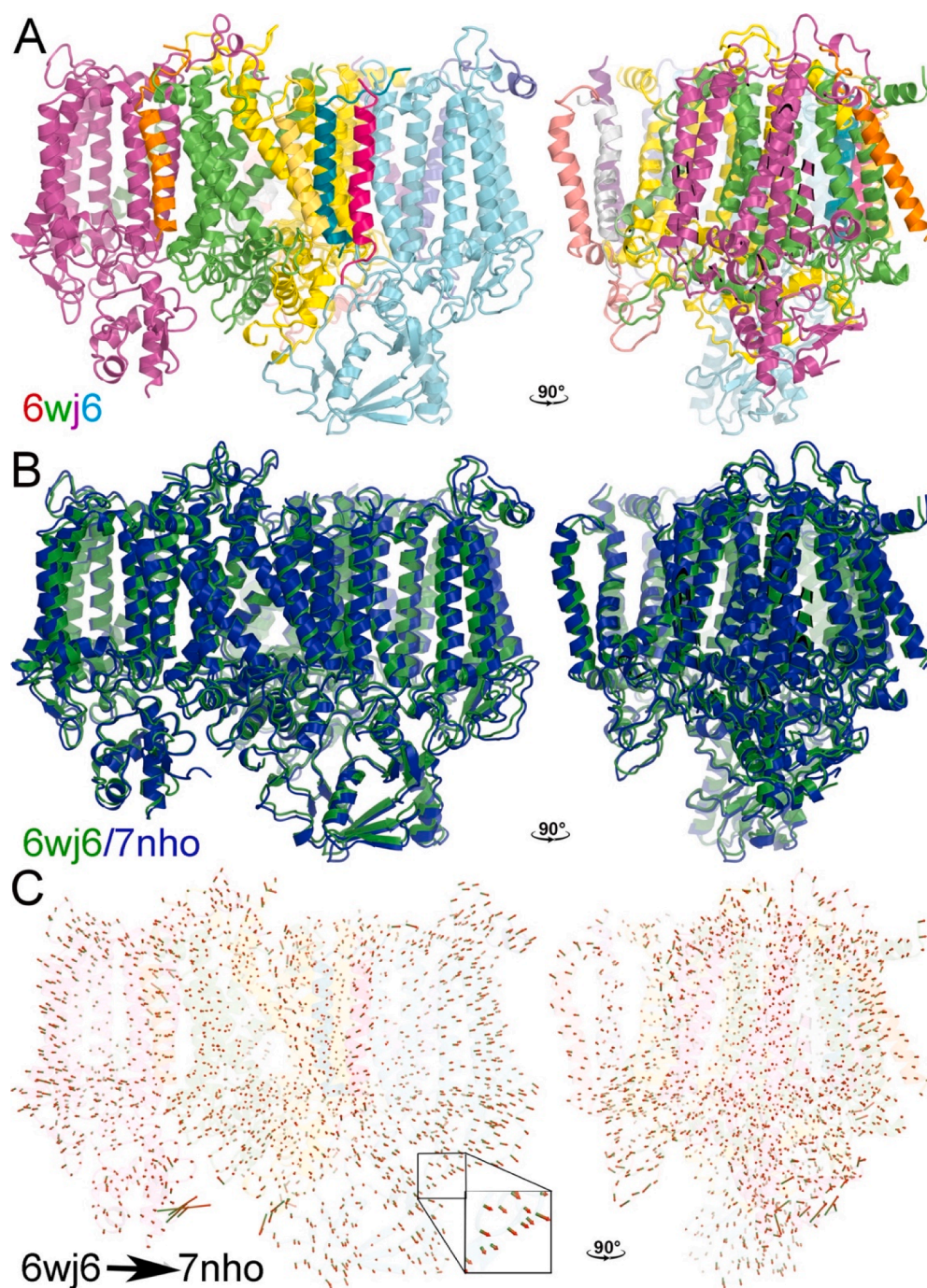
#### 2.4. Automated model refinement against voxel-rescaled experimental maps

The cryo-EM maps of emd22505/7sxt (pH 8.0) and emd22251/6xlu (pH 4.0) were voxel-rescaled by up to  $\pm 10\%$  in steps of 1 % increments (Zhou, 2020; Mannar, 2021; Pettersen, 2004), and the corresponding atomic models for the two structures were used for automated model refinement using Phenix with default settings at 2.3 Å and 2.4 Å resolution, respectively (Adams, 2010). The 6xlu coordinates have also been

modified to include the new R1107A and Y904S substitutions in each subunit (see below). The real-space CC between the experimental maps and the model-calculated maps within masks were calculated, and the model R-factors were plotted as a function of voxel rescaling factors. For map-to-map CC calculations, different sampling rates of target maps were used at different relative contouring levels for inclusion or exclusion of some weak-amplitude voxels.

#### 2.5. Real-space difference maps between density functional theory-derived maps

To demonstrate why removal of resolution differences through Wilson amplitude scaling is important for proper calculation of pairwise vector-difference Fourier maps, we calculated isomorphous difference



**Fig. 1.** Comparison of two monomeric apo-PSII coordinates. (A) Two orthogonal views of the 6wj6 reference structure from *Synechocystis* 6803 in colors: the core subunits D1 (chain A), green, CP47 (chain B), cyan, CP43 (chain C), magenta, D2 (chain D), yellow, and PsbX (chain X), purple. (B) Superposition of the 6wj6 and 7nho structures. (C) Ca shift vectors from the *Synechocystis* 6803 to *T. elongatus* apo-PSII structures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

maps with and without removal of resolution differences in theoretical structures. Isomorphous difference Fourier maps were calculated using the density functional theory (DFT)-derived ED maps at 1.0, 2.0, 3.0, 4.0, and 5.0 Å resolution between cyanide-bound and apo heme molecules after the amplitudes of the two sets of structure factors were properly scaled, using the programs Gaussian 16 and CCP4 (Winn, 2011; Frisch, et al., 2016). Real-space map differences were calculated between the cyanide bound heme at 2.0 Å and the apo heme at 5.0 Å, after the two maps were placed on the same scale and with the same standard deviation of the entire unit cell. Calculations without removal of resolution differences are to demonstrate why amplitude scaling factors are important for a proper analysis of difference maps. Initial coordinates of the cyanide-heme complex (PDB 1shr) and the apo-heme molecule (2dn2) were taken from the corresponding human hemoglobin crystal structures (Sen, 2004; Park et al., 2006). Their geometries were optimized at the UB3LYP/6-31 + G\* level of density functional theory with high-spin Fe(II) of spin 5/2 (Gheidi et al., 2017; Becke, 1988; Hariharan and Pople, 1973). The relationship between the resolution and the overall Wilson B-factor was taken from an empirical equation derived from all crystallographic data deposited in the PDB (Wang, 2017b).

### 3. Results

#### 3.1. Volumetric expansion/contraction in apo-PSII coordinates

The superposition of the two monomeric apo-PSII coordinates shows that the *T. elongatus* apo-PSII, 7nho, is noticeably larger than the *Synechocystis* 6803 apo-PSII, 6wj6 (Fig. 1) (Gisriel, 2020; Zabret, 2021). The increased size of 7nho relative to 6wj6 can be visualized by vectors connecting the equivalent Cα atom pairs in the 6wj6 to 7nho structures (Fig. 1E, 1F). The vectors are largely in radial directions, i.e., from the center-of-mass to given Cα locations. When the lengths of Cα shift vectors are plotted against a function of the radii to the center, the plot can be linearly fitted with a slope of 2.7 % (Fig. 2A). Not all Cα shift vectors are precisely parallel to the radial axes, particularly those of the CP43 subunit (chain C in the PDB coordinate files, cyan in Fig. 1). This is due to different local domain rotations of the CP43 subunit within the complex, associated with different compositions and different pH values (pH 7.0 versus pH 6.5). The plot can also be linearly fitted with a slightly smaller slope of 1.9 % (Fig. 2B) when the Cα shift vectors are projected onto the radial axes, which would largely remove shifts associated with domain rotations since the rotational axes often passed through the center-of-mass. Therefore, an overall volumetric expansion/contraction coefficient of the 7nho structure relative to the 6wj6 structure ranges from 1.9 % to 2.7 %.

The large linear volumetric expansion/contraction coefficients under the experimental conditions of PSII samples are not consistent with the protein compressibility coefficients at high pressure, or contractibility

associated with reduced temperature, as often observed in X-ray crystallography (Roche, 2012; Yamada et al., 2015; Collins et al., 2011). Instead, they result from inaccurate voxel rescaling factors as documented below.

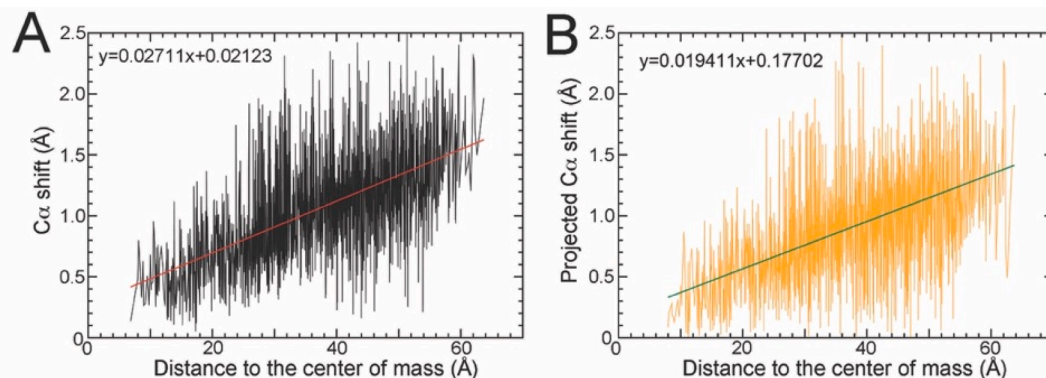
#### 3.2. Determination of relative voxel rescaling factors in cryo-EM maps

The plot of the Pearson real-space CC between the target and reference maps exhibits a bell shape with a well-defined maximum as a function of the correction factor for voxel rescaling of the target map (Fig. 3). When the two maps are on the same scale, the CC between the target and reference map is maximum, enabling us to have both maps on the same scale, as demonstrated for the two apo-PSII structures (Gisriel, 2020; Zabret, 2021). The expansion/contraction coefficient estimated between the two monomeric apo-PSII cryo-EM maps is 2.4 % (Fig. 3A). This value is consistent with the range of 1.9 % and 2.7 % determined from the comparison of Cα coordinates (Fig. 2).

The real-space Pearson CC analysis of two cryo-EM structures of the SARS-CoV-2 spike protein, reported by two independent groups at pH 8.0 and pH 4.0, shows that the voxel rescaling factor difference is 2.9 % (Fig. 3C) (Zhou, 2020; Mannar, 2021). This is typical for cryo-EM maps of the same structure produced by different groups. However, cryo-EM maps obtained by using the same instruments at the same time generally do not exhibit large differences. For example, the two SARS-CoV-2 spike protein structures emd22251/6xlu (pH 4.0) and emd22515/7jwy (pH 4.5) (Fig. 3E) have a relative rescaling correction factor smaller than 0.3 % (Zhou, 2020).

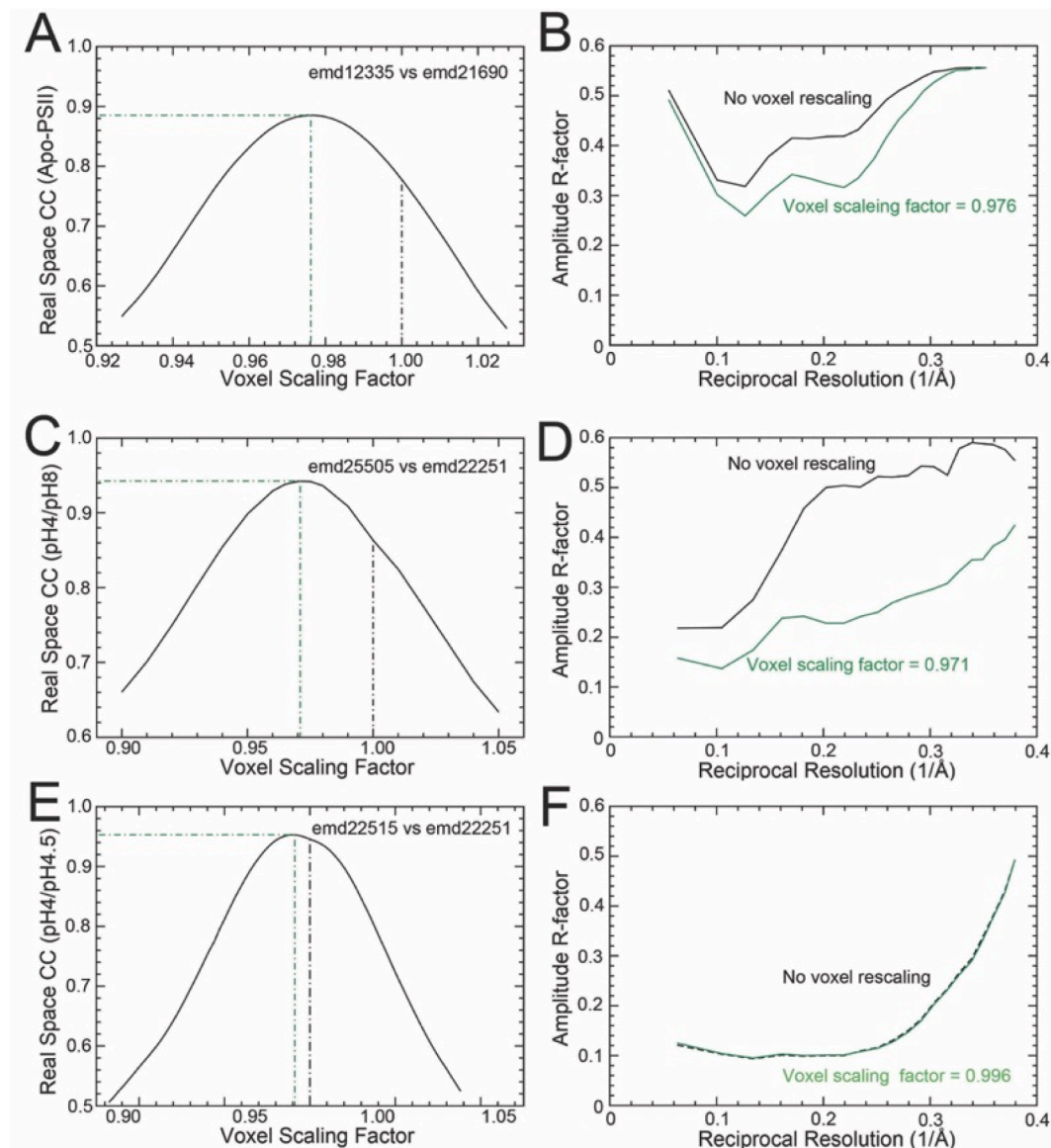
The effect of incorrect voxel scaling factors on cryo-EM maps and the resulting structures can be significant. For example, before applying voxel scaling corrections, the emd25505 and emd22251 maps of the S2 fragments of the spike trimer have an overall amplitude difference of 47.8 % at 2.6 Å, and the amplitude differences at a resolution between 2.6 and 5 Å were > 50 % (Fig. 3B) (Zhou, 2020; Mannar, 2021). After applying the corrections, the amplitude differences are reduced to ~ 20–25 % with an overall value of 25.8 % (Fig. 3B).

X-ray crystallographic structures with an overall amplitude difference > 50 % are considered unrelated since they are close to the Wilson limit of 58.6 % for complete statistical unrelatedness (Wilson, 1949; Wilson, 1950). With such large differences, it would be nearly impossible to determine relative Wilson B-factors to remove resolution differences between them for proper calculations of difference Fourier maps. For the two apo-PSII cryo-EM maps (Gisriel et al., 2020; Zabret et al., 2021), the overall amplitude difference is 44.2 % before voxel scaling correction and is reduced to 38.2 % after it. The remaining large residual difference is mainly due to different rotations of the CP43 subunit in the two structures (see below) as well as different post-refinement correction factors applied to the two maps, which affect data of both high- and low-resolution shells that exhibited very large



**Fig. 2.** Correlation of Cα shift vector length with the Cα distance to the center-of-mass comparing the two apo-PSII structures. (A) The absolute length (fitted slope is 2.711%). (B) Projected Cα shifts onto the radial axes exhibit a smaller slope (fitted slope is 1.9411%).





**Fig. 3.** Numerically varying correction factors for voxel scale between target and reference maps. (A, C, E) Real-space correlation coefficient (CC) after systematically varying the relative correction factors to given voxel scale. (B, D, F) Amplitude differences with and without voxel rescaling after amplitude scaling to remove resolution differences with selected resolution ranges. (A, B) The apo-PSII maps of emd12335/7nho (*T. elongatus*) and emd21690/6wj6 (*Synechocystis* 6803). (C, D) The SARS-CoV-2 spike protein maps of emd22505/7sxt and emd22251/6xlu (pH 4.0). (E, F) The SARS-CoV-2 spike protein maps of emd22515/7jwy (pH 4.5) and emd22251/6xlu (pH 4.0).

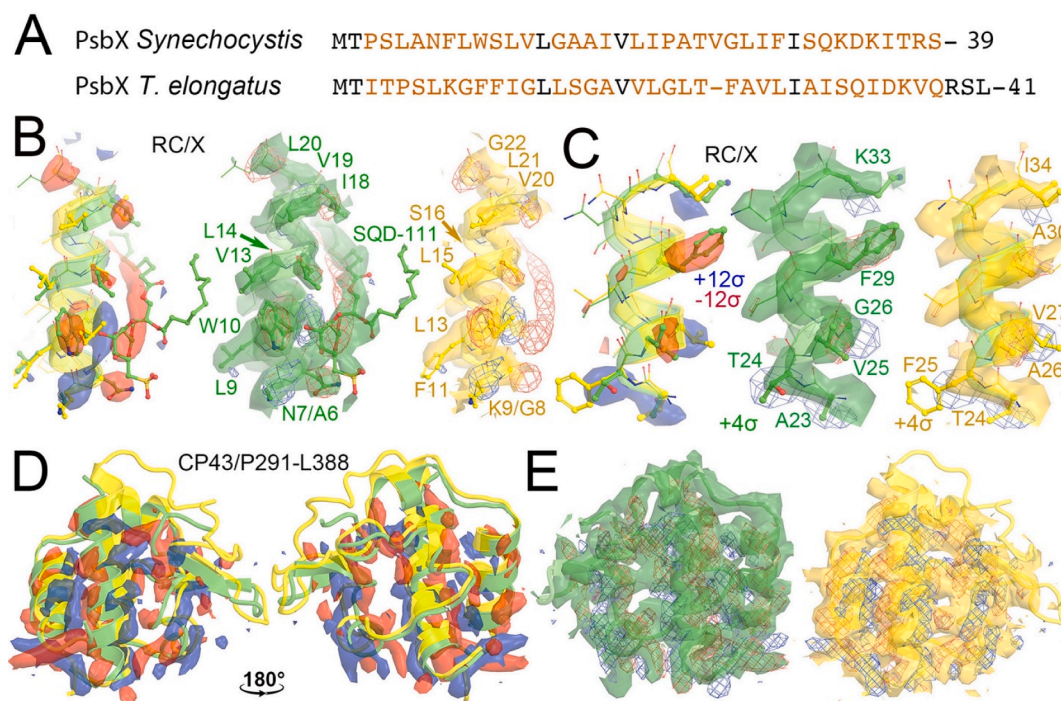
differences even after minimization.

### 3.3. Visualization of amino acid differences and their local consequences

The PsbX subunit of PSII exhibits only 45 % sequence similarity in PSII between *Synechocystis* 6803 and *T. elongatus* and there are 16 different amino acid residues plus one insertion/deletion involving a helix-breaking proline residue (Fig. 4A) (Gisriel, 2020; Zabret, 2021). This transmembrane subunit is located on the periphery of the PSII core and has a relatively low local resolution so the sidechains cannot be directly determined from the experimental maps. Even a large residue, like Phe (F25) of PsbX in the *Synechocystis* 6803 PSII map, cannot be unambiguously assigned (Fig. 4C). The equivalent position in PsbX from *T. vulcanus* has a Thr residue (T24), which is much more challenging to visualize. Nevertheless, the voxel-scaled difference Fourier map generated from the two structures exhibits an outstanding negative peak at this position, which represents a large size difference between the two

residues. There is also an outstanding negative feature at the position corresponding to V25 from *Synechocystis* 6803 PsbX and the corresponding A26 from *T. vulcanus* PsbX, again, representing their size difference, and so on (Fig. 4C). In fact, size difference features can be unambiguously visualized for 15 of the 16 sequence differences on this helix (Fig. 4B, 4C). Other differences in this region included a lipid molecule that is present only in the PSII structure of *Synechocystis* 6803, but not of *T. elongatus* (Gisriel, 2020; Zabret, 2021). Therefore, when the two experimental maps are properly aligned and scaled, vector-difference Fourier maps can reliably reveal every important subtle difference on the helix. This is because the isomorphous vector-difference Fourier maps for two cryo-EM maps have no model-phase bias issue that is often encountered in X-ray crystallography.

Unlike the spike protein, the largest difference ESP features in the two apo-PSII maps do not correspond to amino acid sequence differences. Instead, they are associated with displacements and local domain rotations of the CP43 subunit within each complex (Fig. 4D, 4E, Fig. S1).



**Fig. 4.** Visualization of known amino acid differences and structural differences between the *Synechocystis* 6803 and *T. elongatus* apo-PSII. (A) Sequence alignment of PsbX for the two species. (B, C) The transmembrane PsbX subunit from A6 to K33 has 16 residue differences described in this figure between the two species including an insertion/deletion involving a Pro residue in the middle (omitted in figures and as a division for panels A and B). First panels, superpositions of two structures with difference maps. Positive differences are in blue. Negative differences are in red. Middle panels, superposition of *Synechocystis* 6803 map with both structures, and differences in meshes. Right panels, superposition of *T. elongatus* map with both structures. Substituted residues are shown in large spheres and sticks. There is a lipid molecule (SQA-111) present in the *Synechocystis* 6803 map, but not in the *T. elongatus* map. (D) Two views of superposition of both models with difference maps show relative domain rotations (i.e., non-random distribution of positive and negative features). (E) Superposition of both models with each map (isosurface) and differences (isomesh). See supporting information Fig. S1 for additional stereodiagram for panels D and E. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

When maps for these domains were locally realigned individually and difference ESP maps were recalculated, all amino acid differences between the two species in these domains could be visualized with outstanding difference peaks.

The descending order of difference peak amplitudes between emd22251 (pH 4.0/the “2P” mutant of the spike protein of SARS-CoV-2) and emd25515 (pH 8.0/the “6P” mutant) shows that all large difference peaks are mainly at the known sites or near vicinities of the proline substitutions (Fig. 5) (Zhou, 2020; Mannar, 2021). Between the two structures, three substitution sites of F817P, A892P, A899P exhibit very large difference features (Fig. 5A–5C). All the three proline sidechains have well defined ESP features, noticeably at the two extra methylene groups in the proline ring relative to the parent alanine or phenylalanine residues. The fourth A942P substitution has a much smaller difference peak because this region has a much lower level of ESP features. At the vicinity of the F817P substitution, the largest differences are associated with the displacements of the Q804 backbone that stacks against F817, but not at the substitution site itself (Fig. 5F).

### 3.4. Two overlooked amino acid substitutions in a SARS-CoV-2 spike protein variant

Systematic examination of difference features led us to identify two unexpected amino acid substitutions in two cryo-EM maps reported for the SARS-CoV-2 spike protein (measured at two different pH values, pH 4.0 and pH 8.0) (Fig. 6) (Zhou, 2020; Mannar, 2021). The difference features revealed that strong negative peaks on the sidechains of both R1107 and Y904, suggesting that these large sidechains are absent in the low pH structure. Associated with the negative peak for the R1107 sidechain, there is a positive peak for the repositioned C $\beta$  atom of the

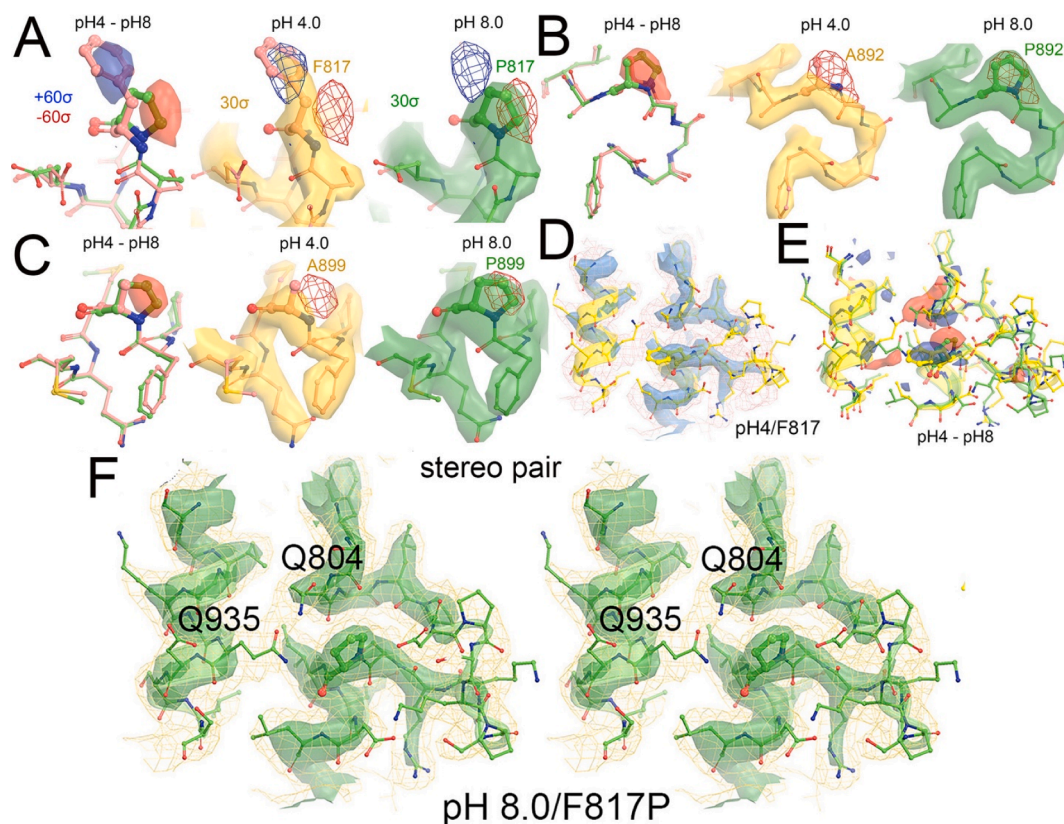
R1107A substitution. The entire R1107 turn is flanked by positive and negative difference features, resulting from a displacement of this loop. This loop has well-defined ESP features in both maps with the highest local resolution (of  $\sim 2.4$  Å) at the center of the protein and can be fitted with great confidence. The fitted loop in the “2P” low-pH structure has C $\alpha$  displacements of 1.0, 1.4, 1.9, 1.6, and 1.4 Å for Q1106, R1107A, N1108, F1109, and Y1110, respectively, relative to the high-pH “6P” structure. These displacements are fully supported by difference ESP features (Fig. 6C).

Upon the displacements of the R1107-containing  $\beta$ -hairpin, the 1.4-Å displacement of R1107A leaves no room for a Tyr sidechain to fit the reported Y904 position in its neighboring subunit because the displaced R1107A (whose identity is unambiguously established in the experimental map) now occupies the Y904 position. Like R1107, the sidechain of Y904 also exhibits a very large negative difference feature, suggesting that a reduced size change of the residue at this position (whose identity is only approximately established in the map). We observed that two alternate conformations of S904 best explain the observed ESP features at this position, both of which are hydrogen-bonded to three ordered water molecules (Fig. 6F). The same substitutions are also clearly present in the emd25515/7jwy structure (pH 4.5) (Zhou, 2020). Without making R1107A and Y904S substitutions in the atomic models of the 6xlu (pH 4.0) and 7jwy coordinates (pH 4.5), the entire R1107-containing loop fitted the ESP features very poorly, as did the Y904 sidechain.

### 3.5. Possible calibration of the absolute voxel scales of cryo-EM maps using model refinement

Voxel scaling errors of 2.7 % in metalloenzymes such as PSII are too





**Fig. 5.** Visualization of known amino acid substitutions in the SARS-CoV-2 spike proteins. (A–C) For F817P, A892P, and A899P substitutions. Left panels, superposition of both structures and their difference maps between pH 4 and pH 8. Middle panels, superposition of pH 4 structure with pH 4 map (isosurface) as well as difference maps (isomesh). Right panels, superposition of pH 8 structure with pH 8 map as well as difference maps. Negative differences are in red, and positive differences in blue. (D) Superposition of pH 4 structure with pH 4 map for F817. (E) Difference maps for local effects of the F817P substitution. (F) Superposition of pH 8 structure with pH 8 map for F817P in stereodiagram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

large to accurately determine oxidation states of metal ions because these errors would far exceed the change of metal–ligand bond lengths associated with different redox states. For the typical single C–C bond length of 1.540 Å, 3 % voxel scaling errors would increase or decrease the ideal C–C bond length by 0.046 Å, and 5 % voxel rescaling errors would alter the ideal bond lengths by 0.077 Å. If one ignores voxel-scaling errors and still applies ideal bond length restraints for model refinement with a permissive standard deviation of 0.010 Å, which many current model refinement programs do, the resulting coordinates could be misinterpreted. Therefore, with incorrect voxel sizes in cryo-EM maps, the coordinates obtained do not bear the physical soundness and they often have incorrectly restrained bond lengths, bond angles, and atomic B-factor distribution between bonded atoms, the latter of which are often used as a fudge factor to artificially improve the CC value between the experimental maps and atomic models by over-fitting the data after masking out voxel scaling errors.

The emd25505/pH 8.0 and emd22251/pH 4.0 cryo-EM maps reported for the SARS-CoV-2 spike protein were manually expanded and contracted through voxel rescaling by up to  $\pm 10$  % in a step of 1 % increments (Zhou, 2020; Mannar, 2021). In each case, the Pearson CC values of the 20 voxel-rescaled maps to the original maps changed rapidly as a function of voxel rescaling factor (Fig. 7A, 7B). The rates of the change depend on the sampling of the target maps. When sampling (before resampling) included all weak-signal voxels, the rates are the fastest. When sampling included only strong-signal voxels, the rates are slower (Fig. 7A, 7B).

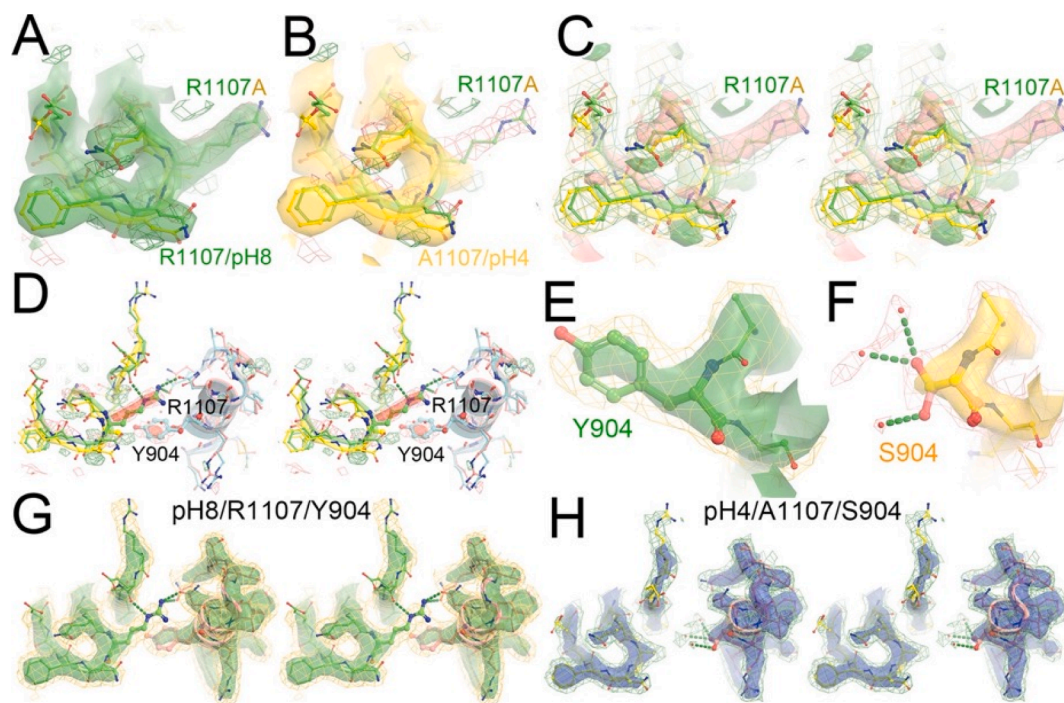
Automated model refinements using Phenix (Adam et al., 2010) against the 20 voxel-rescaled maps also exhibit bell-shaped curves for the CC values between the rescaled experimental maps and model-

calculated maps, with inverted bell-shaped curves for model R-factors (Fig. 7C, 7D). The peak positions (maxima for the CC values and minima for the R-factors) approximately represent the absolute voxel rescaling factors. However, these peaks and troughs are very broad, which imply that the absolute voxel rescaling factors determined this way have large errors. Nevertheless, this model-to-map analysis and the above map-to-map analysis reveal consistent trends of volumetric expansion and contraction observed for the two cryo-EM maps and the coordinates being derived.

All 21 models (the 20 rescaled maps and one original voxel-scaled map) for each structure were refined using the standard bond length restraints because we do not have a prior knowledge about which map is on the absolute voxel scale. The resulting root-mean-squares deviation from the ideal bond lengths varied from 0.006 to 0.007 Å. An analysis of the distribution and mean values of C $\alpha$ -C, C $\alpha$ -C $\beta$ , and C $\alpha$ -N bonds within each residue shows the mean values (as well as median values) are approximately linearly proportional to the voxel rescaling factors and that they are systematically smaller than the corresponding ideal bond lengths used for restraints by significant amounts in both sets of structures (Fig. 8). Without voxel rescaling, the mean differences for these bonds are about  $-0.020$  Å,  $-0.013$  Å,  $-0.027$  Å, respectively (Fig. 8). Even with the maximal expansion ( $+10$  %) studied here, these mean bond lengths remain still smaller than the ideal bond lengths (Fig. 8), which may be due to electron-induced damage to proteins (see discussion). Similar results were obtained when model refinement was carried out using Refmac5 (Fig. 8) (Murshudov et al., 1997).

When the projected C $\alpha$  shift vectors along the radial axes between the original and the expanded/contracted coordinates were plotted as a function of the location of the C $\alpha$  coordinate to the center-of-mass, the





**Fig. 6.** Identification of two overlooked substitutions of amino acid residues in the spike protein of a new SARS-CoV-2 variant. (A) Superposition of R1107/pH 8 structure (green) with pH 8 map (green isosurface) and the A1107/pH 4 structure (gold) with difference maps (isomesh). (B) Superposition of the A1107/pH 4 structure (gold) with pH 4 map (gold isosurface) and the R1107/pH 8 structure (green). (C) Stereodiamgrams of (D and E) but with differences in isosurface and individual maps in isomesh. (D) An extended view of superposition including neighboring subunits where the Y904S substitution is identified. (E) Superposition of the Y904/pH 8 structure with the pH 8 maps at high (green isosurface) and low (gold isomesh) contour level. (F) Superposition of the S904/pH 8.0 structure with pH 4 maps at high (gold isosurface). (G) Superposition of the pH 8 structure with the pH 8 map in stereodiamgram. (H) Superposition of the pH 4 structure with the pH 4 map in stereodiamgram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

resulting coordinate expansion/contraction coefficients were fully consistent with the voxel rescaling factors manually applied to their cryo-EM maps when they were within  $\pm 3\%$  (Fig. 8E). More complex patterns emerged when these coefficients were outside the range of  $\pm 5\%$  (Fig. 8F).

### 3.6. Other methods for determination of the absolute voxel scale of cryo-EM maps

A limitation of model-refinement-based correction for the absolute voxel scaling factor is that protein volumes can sometimes be compressed at high pressures or contracted at reduced temperature without altering local geometry of bond lengths and angles (Roche, 2012; Yamada et al., 2015; Collins et al., 2011). Nevertheless, the coefficients of protein compressibility and contractibility should be much smaller than the large voxel scaling errors observed here. It appears that current model refinement programs cannot effectively distinguish the volume change due to incorrect voxel scaling from that of intrinsic protein compressibility or contractibility. Other more robust procedures need to be developed to address this challenge using an external reference, which could include molecular dynamics simulations and high-resolution crystal structures if they are available.

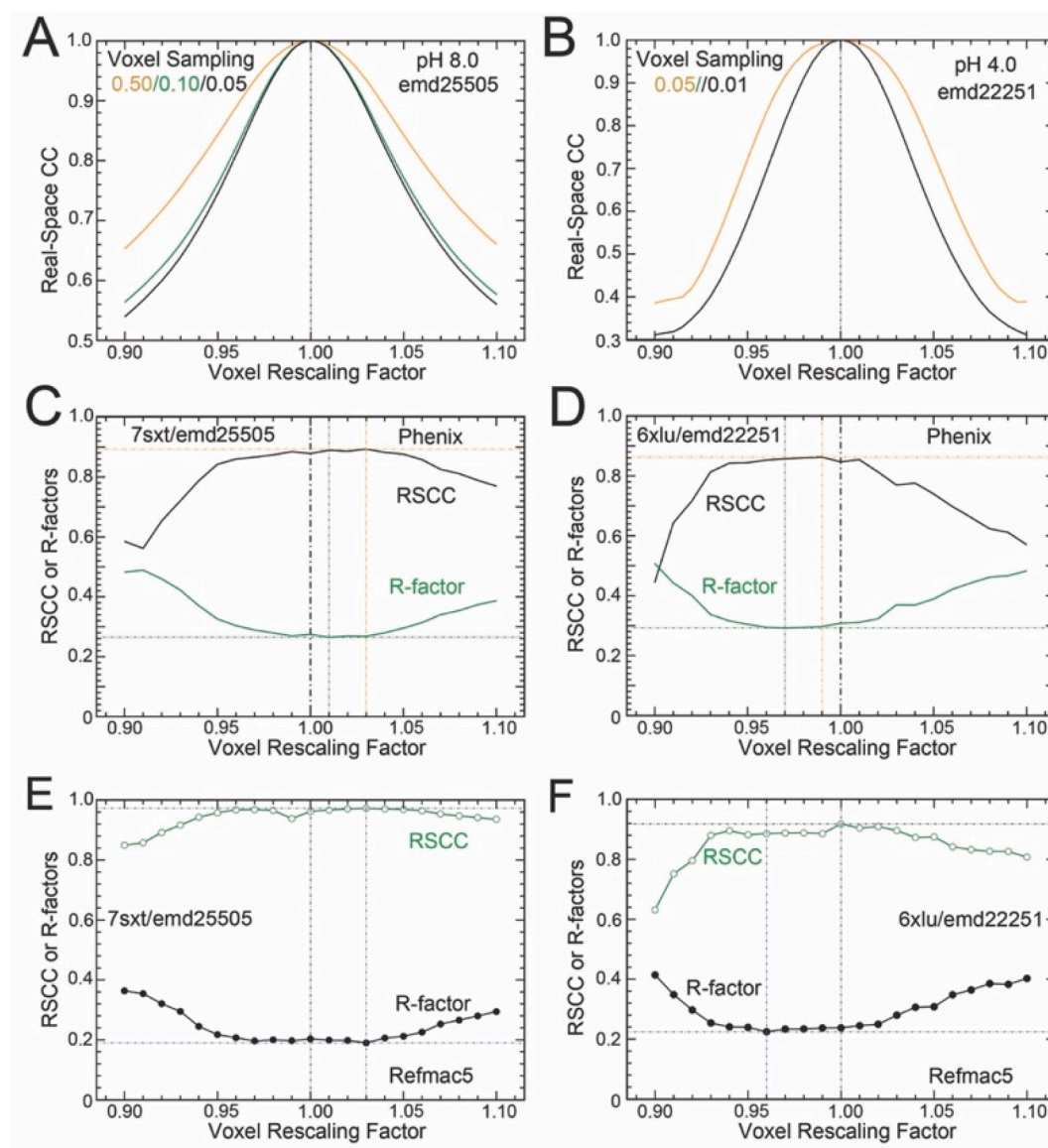
In theory, high-quality crystallographic structures could be used to calibrate the absolute voxel scale factors of a cryo-EM structure. In practice, it is very difficult because often very high-resolution crystal structures are obtained only for small to medium sized protein molecules whereas cryo-EM structures of proteins often do not have accurate crystal structures available for such calibration. The crystal structure reported for dark-adapted PSII (5b5e) represents one of the highest resolution and the most accurate crystal structures for PSII (Tanaka et al., 2017), which in theory could be used for calibration of the absolute voxel scale of PSII cryo-EM maps in theory. However, cryo-EM

and X-ray crystallography may have captured different functional states that could involve some local domain rotations and other structural rearrangements, making the calibration of this kind difficult. Moreover, the metal ion bond lengths and coordination geometry of the oxygen-evolving complex in the crystal structures of PSII intermediates remain highly contested (Wang et al., 2021; Ibrahim, et al., 2021), and they may not be suitable for calibration of cryo-EM maps.

In fact, the voxel scaling issue we initially encountered was during our comparison of MD-derived ESP maps with cryo-EM maps involved in the translocation process of the RNA duplex product after RNA synthesis and other enzyme systems (Wang et al., 2022a; Wang, et al., 2022c). Comparison of atomic models reported for 7bv2 (pre-translocated product complex) with 6xez (post-translocated product complex) shows that these two structures have a relative expansion/contraction factor of 2.4 % associated with uncorrected voxel scale factors of their corresponding cryo-EM maps (Fig. 9) (Chen, 2020; Yin, 2020). A detailed comparison of MD-derived ESP maps with the experimental maps is beyond the scope of this study (Wang et al., 2022a,b,c). Although only three pairs of cryo-EM maps were selected in this study to demonstrate the effect of the voxel scaling issue, this problem exists for nearly all cryo-EM maps and have not yet effectively been addressed.

### 3.7. Real space difference maps without removing resolution differences

Real-space differences can be directly calculated between two cryo-EM maps without going through reciprocal space using Fourier-inverted structure factors, after equalizing the standard deviations of the entire maps while ignoring their resolution differences. The real-space difference maps are challenging to discern if the two maps are in different resolutions. To explain how resolution differences could affect the resulting real-space difference maps, we calculated difference maps between the DFT-derived cyanide-bound heme map at 2.0 Å resolution



**Fig. 7.** Artificially voxel-rescaled maps for model refinement in attempt for determination of the absolute voxel scale of cryo-EM maps using model refinement. (A, B) Pearson real-space correlation coefficients (CC) between voxel-rescaled maps and the original emd25505/7sxt map at the recommended contour level (0.10 unit), half the value, and five times the value, which alter the number of voxels used for calculation. (B) The same plot for emd22251/6xlu. (C) The CC between the voxel-rescaled maps and model-calculated maps (black) as well as model R-factors (green) after standard model refinement using Phenix. Maxima and minima are indicated, which are on the right side of unit rescaling factor. (D) The same plot for emd22251 in which maxima and minima on the left side of the unity rescaling factor after standard model refinement using Phenix. (E, F) The same as (C, D) but with model refinement using Refmac5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and an apo heme map at 5.0 Å resolution (Fig. 10). In the high-resolution map, all ED features are distributed near the atoms whereas in the low-resolution map, the features spread away from the atoms. Differences between the two maps resulted in many negative features above and below the heme plane (small positive features on the heme plane) that could completely obscure the true difference associated with binding of cyanide to the heme (Fig. 10). The resolution-difference problem (as well as the overall B-factor difference problem) in real-space cryo-EM map differences has also been recently addressed elsewhere (Gisriel et al., 2020).

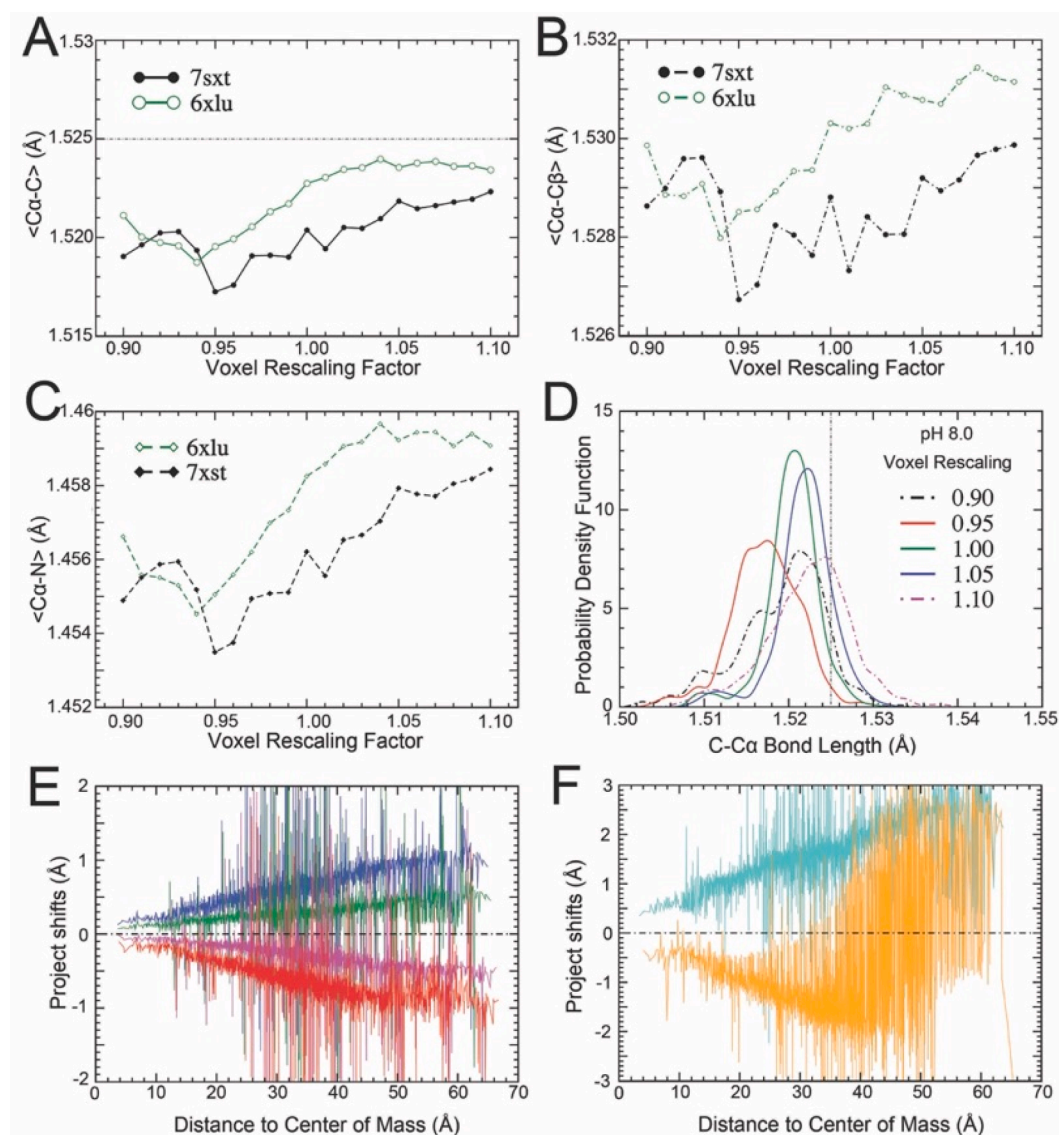
When resolution differences are removed through a reciprocal space amplitude scaling procedure, the feature associated with binding of cyanide to the heme can be unambiguously seen in the difference Fourier maps even at 5.0 Å resolution (Fig. 10). In general, it appears that isomorphous difference Fourier methods are more accurate, partly because it removes B-factor and resolution differences. In fact,

isomorphous difference Fourier methods have been used to reveal azide binding to the heme in heme-containing proteins as well as subtle movements of protein side chains associated with ligand binding even at a very low resolution of 9 Å (Stryer et al., 1964). The isomorphous difference method is the most sensitive method if a high-degree of isomorphism can be achieved (Kraut, 1965).

## 4. Discussion

### 4.1. Isomorphous vector-difference Fourier maps for cryo-EM

In X-ray crystallography, isomorphous difference Fourier maps can reveal binding of small ligands and subtle structural changes between pairs of structures (Kraut, 1965). The method ensures that all differences are stemmed from experimentally observed amplitude differences, not from model phase differences. In cryo-EM maps, however, both



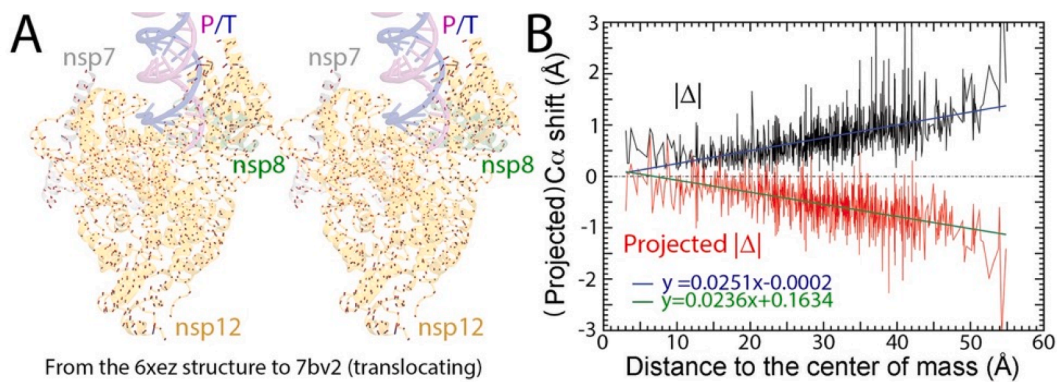
**Fig. 8.** Proportionality of volumetric expansion or contraction coefficients in the resulting coordinates obtained from automated model refinement against voxel-rescaled cryo-EM maps. (A) The mean  $C\alpha-C$  bond lengths as a function of voxel-rescaling factor for the 7sxt (green) and 6xlu (black) coordinates after standard model refinement using Phenix. The ideal  $C\alpha-C$  bond length of 1.525 Å is indicated in dotted line. (B) The  $C\alpha-C\beta$  bond length. (C) The  $C\alpha-N$  bond length. (D) Examples of  $C\alpha-C$  bond length distributions in the entire 7sxt structure for selected voxel rescaling factors of 0.90 and 1.00 in dashed curves and 0.95, 1.00, 1.05 in solid curves. (E) Projected  $C\alpha$  shifts as a function of the  $C\alpha$  distance to the center of the mass for voxel-rescaling factors of 1.01 (green), 1.02 (blue), 0.99 (magenta) and 0.98 (red) relative to the unit. (F) Those of voxel-rescaling factors 1.05 (cyan) and 0.95 (gold) relative to the unit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

amplitudes and phases are experimentally measured. So, vector-difference Fourier maps could be even more sensitive if the amplitudes could be placed on the same absolute scale after the two maps are aligned. It could detect binding of a single proton to an ionizable residue, which would have an enhanced signal in cryo-EM maps. This comparison requires that the cryo-EM maps should be properly scaled to have the same overall B-factors and the same resolution using Wilson scaling methods because the structure factor amplitudes of any molecule should always obey the Wilson statistics (Moore, 2012; Wilson, 1949). This scaling should be applied after correcting for the relative voxel scaling issue and before post-refinement procedures applied to two half maps (Wang et al., 2018a). Afterward, isomorphous vector-difference Fourier maps can reveal subtle differences of maps associated with structural differences that are important for biological information (Wang et al., 2018b).

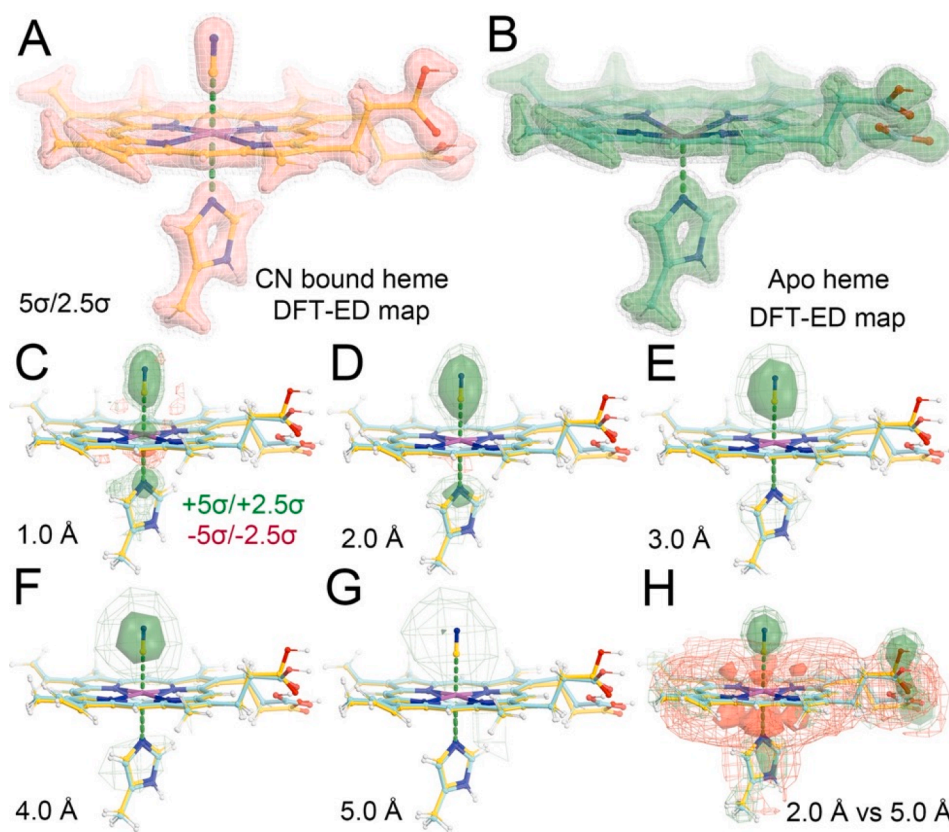
When voxel scale factors are not corrected, the amplitude differences of the two maps being compared are so large that Wilson scaling is no

longer applicable. The resulting vector-difference map between them is then dominated by positive and negative features from misaligned structural features associated with artificially expanded and contracted structures. When voxel scale factors are corrected, these large artificial difference features are eliminated so that the smaller, biologically relevant difference features can be revealed. If local domain rotations can be detected and excluded, even smaller difference features can be identified. For example, after voxel scaling correction, we have revealed two overlooked amino acid substitutions (R1107A/Y904S) in one of the spike proteins of SARS-CoV-2 used for cryo-EM structure determination (Fig. 6). The finding is not completely unexpected because mutations may arise in the plasmids of spike protein after many runs of plasmid amplification and may have escaped from detection if they were not frequently re-sequenced for chemical validation.





**Fig. 9.** Volumetric expansion/contraction in atomic coordinates of the translocating pair of the RTC complexes of SARS-CoV-2. (A) Stereodiagram showing the Cα shift vectors from the 6xez to 7bv2 structures. (B) Total vector lengths (black) and projected vector lengths (red) as a function of distance to the center-of-mass of the complexes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 10.** The DFT derived-ED maps and difference maps. (A) The DFT-derived ED maps for the cyanide bound to the heme molecule contoured at  $2.5\sigma$  (isomesh) and  $5.0\sigma$  (isosurface). (B) For the apo heme. (C-G) Isomorphous difference Fourier maps between the two structures at 1.0, 2.0, 3.0, 4.0, and 5.0 Å resolution, respectively. The maps are contoured at  $\pm 2.5\sigma$  (green/red isomesh), and  $\pm 5.0\sigma$  (green/red isosurface). (H) Real-space difference maps without removing resolution difference (at 2.0 and 5.0 Å resolution). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

#### 4.2. X-ray and electron radiation-induced damage

Crystallographic ED maps can also suffer from the coordinate scaling ambiguity problem when a high-dose X-ray radiation is used for data collection. The unit cell volume is gradually expanded with accumulated dose of X-ray and data collection time (Ravelli et al., 2002; Sliz et al., 2003). A fraction of this expansion is transiently associated with local heating whereas much of the expansion is permanent due to radiation-induced oxidation (Wang, 2016a,b). The averaged unit cell dimensions would represent diffraction data of the structures with different volume-scaling factors. This volume-scaling problem is minimized when a low-dose of X-ray radiation is used for data collection (Tanaka et al., 2017). Moreover, crystallographers routinely exclude severely damaged images during data processing (e.g., when the overall intensity-scale factors have been reduced by  $>30\%$ , or the overall B-

factor increment  $\Delta B$  for intensity-scaling factors are above  $+5$  to  $+10\%$   $\text{\AA}^2$ ). However, very few cryo-EM studies have attempted to exclude severely damaged frames in movie stacks of individual images during image reconstruction (Gisriel, et al., 2022; Kato, 2021). Therefore, the effect of electron-induced damage on the resulting cryo-EM maps remains poorly characterized.

Recent analysis of high-dose high-resolution cryo-EM maps shows that Cα atoms often exhibit very small chiral volumes (Wang, 2017a). Difference maps between late and early time frames of movie stacks of images from the *T. acidophilum* proteasome shows that the protein backbone and sidechains exhibited dehydrogenation reactions and that its β-ring expanded slightly and rotated relative to the α-ring, as a function of the electron dose (Wang et al., 2018b). Dose-dependent image reconstruction for the two PSII cryo-EM maps with higher-resolution showed that high-resolution maps could be obtained using

only the first few frames from the stack of images corresponding to the lowest cumulative electron dose (Gisriel, et al., 2022; Kato, 2021). Addition of the remaining frames increased the total electron dose (by as high as 26-fold) but does not substantially improve the final resolution of the resulting maps. Instead, the added frames primarily corresponded to damaged proteins. The presence of large fractions of damaged proteins in cryo-EM maps makes it very difficult to interpret map features, refine atomic models, and correct for voxel rescaling factors. Electron radiation-induced damage may not affect the interpretation of some important features (Hattné, 2018; Zhang, 2021), but it can affect the absolute voxel scaling factor. Efforts are needed to reliably estimate the extent of damage in image movie stacks and to properly remove corrupted data due to radiation damage during image reconstruction since high-dose electron exposure is often recommended for data collection (Grant, 2015; Naydenova et al., 2020).

A recent analysis of the C $\alpha$ -C bond lengths in cryo-EM structures at 4.0 and 2.8 Å resolution has shown that the bond-lengths were systematically shorter than those in the corresponding crystallographic structures (Sauer, 2020). Correction of the models by refining against + 10 % volume expansion of the cryo-EM maps above still shows that mean bond lengths that are systematically shorter than the expected values. There are two possibilities for these observations: (i) anisotropically corrected voxels may have systematically reduced the magnification, and (ii) bond lengths are shortened by the effect of continuous electron radiation (i.e., increased bond order). In fact, it is known that electron damage introduces structural heterogeneity to the samples and limits the resolution of the resulting cryo-EM maps (Wang et al., 2018). Previous systematic analyses have also shown that none of the C $\alpha$  atoms exhibited a chiral volume in high-dose high-resolution cryo-EM maps as reported for *E. coli*  $\beta$ -galactosidase and the *Thermoplasma acidophilum* proteasome (Wang, 2017a; Wang et al., 2018b). This could be interpreted as electron-induced dehydrogenation (or hydrogen abstraction) inducing  $sp^3$  to  $sp^2$  conversion of C $\alpha$  atoms and systematically increasing the bond order of C $\alpha$  atoms (Wang et al., 2018b).

The high energy of electron radiation always generates both highly oxidizing hydroxyl radicals and highly reducing hydrogen radicals, as well as hydrated electrons. These all damage the protein samples. Through free radical chain reactions, X-ray radiation results in large scale oxygen-atom additions to residues that increase the protein volume (Wang, 2016a,b) whereas electron radiation results in dehydrogenation to increase the rigidity of proteins (Wang et al., 2018b). Therefore, both X-ray and electron radiation produce H<sub>2</sub> although through distinct chemical mechanisms of hydrogen abstraction (Leapman and Sun, 1995; Meents et al., 2009). Radiation induced modifications of metalloenzymes to redox active metal centers can be even more complicated because these enzymes have evolved to funnel either highly reducing species (i.e., electron equivalents) or highly oxidizing species (i.e., electron “hole” equivalents) to metal ion clusters through specific pathways. Thus, the resulting structures may not be accurate in determining the oxidation state of metalloenzymes, especially when the metal ions are continuously oxidized or reduced in complex patterns. The current model refinement process is very powerful and can fit atomic models into cryo-EM maps on any voxel scale (Adams, 2010; Murshudov et al., 1997; Yamashita et al., 2021). Therefore, model refinement alone cannot assess the absolute voxel scaling errors or ensure the correctness of metal ion-ligand bond lengths.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

Authors thank Drs. Frederick J. Sigworth and Charles Sindelar for

insightful discussion during this study. This study has been funded in part by grants from the Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences (DE-SC0001423 to V. S. B. and DE-FG02-05ER15646 to G.W.B.), by NIH grant R01AI163395 to Y.X., by NIH K99GM140174 to C.J.G. D.A.F. was supported by the Predoctoral Program in Biophysics NIH T32 GM008283. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the DOE.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsb.2022.107902>.

## References

- Adams, P.D., et al., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Bao, H., Burnap, R.L., 2016. Photoactivation: The Light-Driven Assembly of the Water Oxidation Complex of Photosystem II. *Front. Plant Sci.* 7, 578.
- Becke, A.D., 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A Gen Phys* 38, 3098–3100.
- Chen, J., et al., 2020. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* 182, 1560–1573 e1513.
- Collins, M.D., Kim, C.U., Gruner, S.M., 2011. High-pressure protein crystallography and NMR to explore protein conformations. *Annu. Rev. Biophys.* 40, 81–98.
- Cox, N., Pantazis, D.A., Lubitz, W., 2020. Current understanding of the mechanism of water oxidation in photosystem II and its relation to XFEL data. *Annu. Rev. Biochem.* 89, 795–820.
- W. L. Delano, Pymol. Schrodinger, Inc., <http://pymol.org/>.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Frank, J., 2018. Single-particle cryo-electron microscopy: the path toward atomic resolution (selected papers of Joachim Frank with commentaries). Series in structural biology. World Scientific, New Jersey pp. xii, 565 pages.
- M. J. Frisch et al. (2016) Gaussian 16, Revision C.01. (Wallingford, CT).
- Gheidi, M., Safari, N., Zahedi, M., 2017. Density functional theory studies on the conversion of hydroxyheme to iron-verdoheme in the presence of dioxygen. *Dalton Trans.* 46, 2146–2158.
- Gisriel, C.J., et al., 2020. Cryo-EM structure of monomeric photosystem II from *Synechocystis* sp. PCC 6803 lacking the water-oxidation complex. *Joule* 4, 2131–2148.
- Gisriel, C.J., et al., 2022. High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, *Synechocystis* sp. PCC 6803. *PNAS* 119.
- Gisriel, C.J., Wang, J., Brudvig, G.W., Bryant, D.A., 2020. Opportunities and challenges for assigning cofactors in cryo-EM density maps of chlorophyll-containing proteins. *Commun Biol* 3, 408.
- Glaeser, R.M., 2019. How good can single-particle Cryo-EM become? What remains before it approaches its physical limits? *Annu. Rev. Biophys.* 48, 45–61.
- Glaeser, R.M., Chiu, W., Nogales, E., 2021. *Single-particle Cryo-EM of biological macromolecules* (IOP. Publishing Ltd 120).
- Grant, E.J., et al., 2015. A report from the 2013 international symposium: the evaluation of the effects of low-dose radiation exposure in the life span study of atomic bomb survivors and other similar studies. *Health Phys.* 108, 551–556.
- Grant, T., Grigorieff, N., 2015. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.* 192, 204–208.
- Hariharan, P.C., Pople, J.A., 1973. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* 28, 213–222.
- Harvey, W.T., et al., 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424.
- Hattné, J., et al., 2018. Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* 26, 759–766 e754.
- Heinz, S., Liauw, P., Nickelsen, J., Nowaczyk, M., 2016. Analysis of photosystem II biogenesis in cyanobacteria. *Biochim. Biophys. Acta, Mol. Cell. Biol. Lipids* 1857, 274–287.
- Ibrahim, M., et al., 2021. Reply to Wang et al.: Clear evidence of binding of O<sub>2</sub> to the oxygen-evolving complex of photosystem II is best observed in the omit map. *PNAS* 118.
- Joseph, A.P., et al., 2020. Comparing Cryo-EM Reconstructions and Validating Atomic Model Fit Using Difference Maps. *J. Chem. Inf. Model.* 60, 2552–2560.
- Kato, K., et al., 2021. High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams. *Commun Biol* 4, 382.
- Kraut, J., 1965. Structural studies with X-Rays. *Annu. Rev. Biochem.* 34, 247–268.
- Leapman, R.D., Sun, S., 1995. Cryo-electron energy loss spectroscopy: observations on vitrified hydrated specimens and radiation damage. *Ultramicroscopy* 59, 71–79.
- Li, F., 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 3, 237–261.

- Mannar, D., et al., 2021. Structural analysis of receptor binding domain mutations in SARS-CoV-2 variants of concern that modulate ACE2 and antibody binding. *Cell Rep* 37, 110156.
- Meents, A., Ditttrich, B., Gutmann, S., 2009. A new aspect of specific radiation damage: hydrogen abstraction from organic molecules. *J. Synchrotron Radiat.* 16, 183–190.
- Moore, P.B., 2012. Visualizing the invisible: imaging techniques for the structural biologist. Oxford University Press, Oxford, U.K.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.
- Naydenova, K., Jia, P., Russo, C.J., 2020. Cryo-EM with sub-1 Å specimen movement. *Science* 370, 223–226.
- Nixon, P.J., Michoux, F., Yu, J., Boehm, M., Komenda, J., 2010. Recent advances in understanding the assembly and repair of photosystem II. *Ann. Bot.* 106, 1–16.
- Park, J., et al., 2021. Structure of New Binary and Ternary DNA Polymerase Complexes From Bacteriophage RB69. *Front Mol Biosci* 8, 704813.
- Park, S.Y., Yokoyama, T., Shibayama, N., Shiro, Y., Tame, J.R., 2006. 1.25 Å resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms. *J. Mol. Biol.* 360, 690–701.
- Pettersen, E.F., et al., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Ravelli, R.B., Theveneau, P., McSweeney, S., Caffrey, M., 2002. Unit-cell volume change as a metric of radiation damage in crystals of macromolecules. *J. Synchrotron Radiat.* 9, 355–360.
- Roche, J., et al., 2012. Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci U S A* 109, 6945–6950.
- Sauer, D.B., et al., 2020. Structural basis for the reaction cycle of DASS dicarboxylate transporters. *Elife* 9, (61350).
- Sen, U., et al., 2004. Crystal structures of HbA2 and HbE and modeling of hemoglobin delta 4: interpretation of the thermal stability and the antisickling effect of HbA2 and identification of the ferrocyanide binding site in Hb. *Biochemistry* 43, 12477–12488.
- Sliz, P., Harrison, S.C., Rosenbaum, G., 2003. How does radiation damage in protein crystals depend on X-ray dose? *Structure* 11, 13–19.
- Stryer, L., Kendrew, J.C., Watson, H.C., 1964. The mode of attachment of the azide ion to sperm whale metmyoglobin. *J. Mol. Biol.* 8, 96–104.
- Tanaka, A., Fukushima, Y., Kamiya, N., 2017. Two Different Structures of the Oxygen-Evolving Complex in the Same Polypeptide Frameworks of Photosystem II. *J. Am. Chem. Soc.* 139, 1718–1721.
- Vinyard, D.J., Brudvig, G.W., 2017. Progress toward a molecular mechanism of water oxidation in photosystem II. *Annu. Rev. Phys. Chem.* 68, 101–116.
- Vinyard, D.J., Ananyev, G.M., Dismukes, G.C., 2013. Photosystem II: the reaction center of oxygenic photosynthesis. *Annu. Rev. Biochem.* 82, 577–606.
- Wang, J., 2016a. X-ray radiation-induced addition of oxygen atoms to protein residues. *Protein Sci.* 25, 1407–1419.
- Wang, J., 2016b. Oxygen additions in serial femtosecond crystallographic protein structures. *Protein Sci.* 25, 1797–1802.
- Wang, J., 2017a. Experimental charge density from electron microscopic maps. *Protein Sci.* 26, 1619–1626.
- Wang, J., 2017b. On contribution of known atomic partial charges of protein backbone in electrostatic potential density maps. *Protein Sci.* 26, 1098–1104.
- Wang, J., et al., 2022a. Insights into Binding of Single-Stranded Viral RNA Template to the Replication-Transcription Complex of SARS-CoV-2 for the Priming Reaction from Molecular Dynamics Simulations. *Biochemistry* 61, 424–432.
- Wang, J., et al., 2022b. Structural Basis for Reduced Dynamics of Three Engineered HNH Endonuclease Lys-to-Ala Mutants for the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Associated 9 (CRISPR/Cas9) Enzyme. *Biochemistry* 61, 785–794.
- Wang, J., Konigsberg, W.H., 2022. Two-Metal-Ion Catalysis: Inhibition of DNA Polymerase Activity by a Third Divalent Metal Ion. *Front Mol Biosci* 9, 824794.
- Wang, J., Liu, Z., Crabtree, R.H., Frank, J., Moore, P.B., 2018a. On the damage done to the structure of the Thermoplasma acidophilum proteasome by electron radiation. *Protein Sci.* 27, 2051–2061.
- Wang, J., Liu, Z., Frank, J., Moore, P.B., 2018b. Identification of ions in experimental electrostatic potential maps. *IUCr* 5, 375–381.
- Wang, J., Armstrong, W.H., Batista, V.S., 2021. Do crystallographic XFEL data support binding of a water molecule to the oxygen-evolving complex of photosystem II exposed to two flashes of light? *Proc Natl Acad Sci U S A* 118, e2023982118.
- J. Wang et al., Structural insights into binding of remdesivir triphosphate within the replication-transcription complex of SARS-CoV-2. *Biochemistry In Press* (2022).
- Wilson, A.J.C., 1949. The probability distribution of X-ray intensities. *Acta Crystallogr* 2, 318–321.
- Wilson, A.J.C., 1950. Largest likely values for the reliability index. *Acta Crystallogr* 3, 397–398.
- Winn, M.D., et al., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242.
- xmgrace, (<https://plasma-gate.weizmann.ac.il/Grace/>). (2021).
- Yamada, H., Nagae, T., Watanabe, N., 2015. High-pressure protein crystallography of hen egg-white lysozyme. *Acta Crystallogr. D Biol. Crystallogr.* 71, 742–753.
- Yamashita, K., Palmer, C.M., Burnley, T., Murshudov, G.N., 2021. Cryo-EM single-particle structure refinement and map calculation using Servcat. *Acta Crystallogr D Struct Biol* 77, 1282–1291.
- Yin, W., et al., 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368, 1499–1504.
- Zabret, J., et al., 2021. Structural insights into photosystem II assembly. *Nat. Plants* 7, 524–538.
- Zhang, Y., et al., 2021. Single-particle cryo-EM: alternative schemes to improve dose efficiency. *J. Synchrotron Radiat.* 28, 1343–1356.
- Zhao, J., Brubaker, M.A., Benlekbi, S., Rubinstein, J.L., 2015. Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images. *J. Struct. Biol.* 192, 209–215.
- Zhou, T., et al., 2020. Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains. *Cell Host Microbe* 28, 867–879 e865.
- Zivanov, J., Nakane, T., Scheres, S.H.W., 2020. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCr* 7, 253–267.